Journal of Positive Psychology & Wellbeing http://journalppw.com
2023, Vol. 7, No. 1, 1005 – 1018 ISSN 2587-0130

Initial Conceptions in the Statistical Reasoning and Modeling Processes in Engineering Students

Rosa Virginia Hernández[1], Elisabeth Ramos Rodríguez[2], Luis Fernando Mariño[3]

[1]Departamento de Matemáticas y Estadística, Universidad Francisco de Paula Santander, Colombia; rosavirginia@ufps.edu.co; Orci: https://orcid.org/0000-0002-2638-671X,
[2]Instituto de Matemáticas, Pontificia Universidad Católica de Valparaíso, Chile; elisabeth.ramos@pucv.cl, Orci: https://orcid.org/0000-0002-8409-4125,
[3]Departamento de Matemáticas y Estadística, Universidad Francisco de Paula Santander, Colombia; fernandoml@ufps.edu.co, https://orcid.org/0000-0002-3438-6963

Abstract

The cognitive view in educational psychology is considered one of the traditional perspectives for psychologists, who recognize that students may have different but identifiable conceptions of learning. The work aimed to compare initial conceptions of statistical modeling processes and levels of statistical reasoning in engineering students. The research had a qualitative approach with a multiple-case design. As a source of data collection, a questionnaire was designed and applied to 38 students taking the subject Probability and Statistics. The results and data analysis showed that none of the participants went through all the phases of statistical modeling, evidencing a low ability to establish relationships in the data and interconnect statistical concepts. Consequently, the students failed to reach the maximum level of statistical reasoning. The participants also identified difficulties in establishing coherent interpretations when making representations of statistical models such as frequency tables, graphs and measures of central tendency. Therefore, there is a need to continue developing research that provides theoretical foundations for the characterization of the modeling and statistical reasoning processes to improve the learning of statistics and probability in university students.

Keywords: Educational Psychology; Initial Conceptions; Statistical Modeling; Statistical Reasoning; Higher Education.

Resumen

La visión cognitiva en la psicología educativa se considera una de las perspectivas tradicionales para los psicólogos, quienes reconocen que los estudiantes pueden tener concepciones de aprendizaje diferentes, pero identificables. El objetivo del trabajo estuvo centrado en comparar las concepciones iniciales hacia los procesos de modelación estadística y los niveles de razonamiento estadísticos en estudiantes de ingeniería. La investigación tuvo un enfoque cualitativo con un diseño de casos múltiples. Como fuente de recolección de información se diseñó y aplicó un cuestionario a 38 estudiantes que cursaron la materia Probabilidad y Estadística. Los resultados y el análisis de datos mostraron que ninguno de los participantes logró transitar por todas las fases de modelación estadística, evidenciando una baja capacidad para establecer relaciones en los datos e interconectar conceptos estadísticos. En consecuencia, los estudiantes no lograron alcanzar el nivel máximo de razonamiento estadístico. Se identificaron también dificultades en los participantes para establecer interpretaciones coherentes al realizar representaciones de modelos estadísticos como tablas de frecuencia, gráficos y medidas de tendencia central. Por tanto, se evidencian la necesidad de continuar desarrollando investigaciones que aporten fundamentos teóricos hacia la caracterización de los procesos de

modelación y razonamiento estadístico para mejorar el aprendizaje de la estadística y probabilidad en los estudiantes universitarios.

Palabras clave: Psicología Educativa, Concepciones Iniciales; Modelación Estadística; Razonamiento Estadístico; Educación Superior.

1. INTRODUCTION

Cognitive psychology focuses on people's mental processes to understand how the mind works and to use this knowledge to conduct empirical research about how humans perceive, remember, think, and reason (Woolfolk, 2020). Initial conceptions or preconceived ideas that people have about something are not superficial beliefs. These initial conceptions represent a framework that can be used to seek explanations about people's mental representation models at a given time and, at the same time, be useful to design and implement didactic strategies that contribute to the student's conceptual development. Therefore, it is necessary to identify what the student knows and build on this knowledge (Sawyer, 2006; Bransford et al., 2000).

In the specific case of learning statistics and probability concepts, it is an act of interpretation that is affected by the prior knowledge that the learner has (Vosniadou, 2012). Teachers who guide the subject of statistics and probability in university programs have a great commitment to promoting academic quality (Kazak & Pratt, 2017). One possible way to achieve that quality is to adopt and use modeling constructs, modeling and statistical reasoning that support classroom practices according to students' context and specific needs (Cazco et al., 2018).

Although modeling processes have been implemented in mathematical research in recent decades, statistical modeling constructs and reasoning have been little addressed in higher education research (Radke et al. 2020; Dvir and Ben-Zvi, 2021; Lehrer and Schauble, 2019; Biehler et al., 2018; Pfannkuch et al. 2018; Wild et al. 2018; Tacoma et al. 2018; and Doerr et al. 2017).

The word model has several meanings. For some authors, a model means a symbolic representation of certain aspects of the real world and its forms of representation can be through different systems of signs, images, patterns, and languages, inscribed in different registers of representations (Henry, 1997). Models can be understood as basic units from which one can seek explanations and make inferences about something (Hestenes, 2010).

A model is considered statistical if it has two distinguishing characteristics: 1) the phenomenon it attempts to explain has an aspect of variability, and 2) using it includes use of probabilistic considerations (Brown & Kass, 2009). Statistical reasoning, on the other hand, can be defined as how people reason with statistical ideas, making interpretations based on data sets, graphical representations and statistical summaries to interpret results and make inferences (Garfield, 2002).

Statistical modeling is considered a theoretical approach based on a complex structure or process derived from initial models, making it possible to explore and learn about students' reasoning processes (Lehrer & Schauble, 2019; Divir & Ben-Zvi, 2021 and Pfannkuch et al., (2018). Biehler et al. (2018) assert that statistical modeling processes are methodologically interconnected with the randomness of data and make it possible to answer research questions in real contexts. These authors recognize the importance of creating models through technological tools simulating random behaviors as an alternative to enhance and improve the student's statistical reasoning capacity.

For their part, Alston-Knox et al. (2019) state that understanding the cognitive challenges of statistical modeling requires understanding the student's reasoning roots and progress. Furthermore, model-based statistical reasoning allows for the development of a wide range of representations, comprising a language of expression through drawings, diagrams, maps, mathematical functions, and simulations, providing an instructional sequence of activities (Doerr et al., 2017)

Some researchers such as Lehrer and Schauble, (2019), state that statistical reasoning based on models means having the ability to implement teaching and learning activities in statistics to build a representation through real data. This implies leaving aside traditional teaching, where the subject continues to be oriented through mechanical processes, substituting formulas, where the student must resort to tools and techniques that are soon forgotten (Biehler et al., 2018).

Statistical modeling processes have strong theoretical, didactic, and computational support as an area of development in applied mathematics (Tacoma et al., 2018). However, its implementation in the classroom has been little exploited. This, together with the excessive content load in the curriculum, prevents future professionals from achieving the integration of statistical knowledge with technological tools in the analysis of real data (Wild et al., 2018)

Cazco et al. (2018) state that it is required to implement the constructs of modeling, modeling and statistical reasoning with the use of technology to improve student learning. In addition, it is of utmost importance to identify the difficulties and initial conceptions about modeling processes and statistical reasoning levels of the student when entering higher education and thus be able to obtain a timely diagnosis about their ability to reason statistically (Pfannkuch et al., 2016 and Verdejo et al., 2021).

Inquiring about the initial conceptions (ideas or ways of understanding something) of university students in Modeling and Statistical Reasoning processes will encourage the development of research that contributes theories towards the search for didactic strategies through organized and sequenced activities, improving the instructional design (Garfield and Ben-Zvi, 2008 and Porras Lizano and Castro-Rodríguez, 2021).

The paper presented here aims to answer the research question: What are the initial conceptions of students training to become engineers in the processes of modeling and statistical reasoning at the time of starting a Probability and Statistics course at the Universidad Francisco de Paula Santander?

By answering this question, the initial conceptions of the statistical modeling process and the level of reasoning of university students are compared without didactic intervention, without having mediated any learning or review proposal that allows the development of such constructs.

2. THEORETICAL FRAMEWORK

The constructs, statistical model, statistical modeling and reasoning have been characterized differently. This section presents some characterizations to provide a general framework while specifying which of these constructs was a reference in the study.

2.1 Statistical model

From formal theory, a model is a symbolic representation of certain aspects of an object or phenomenon of the real world; that is, a model is an expression or formula written following the rules of a specific symbolic system. For Henry (1997), a model can be represented in different systems of signs, images, patterns, and languages inscribed in different registers of representations. Models are basic units of coherently structured knowledge from which logical inferences, predictions, explanations, plans and designs can be made (Hestenes, 2010). For its part, a model is considered statistical if it is constructed for a statistical purpose, which generally means that it must have two distinguishing characteristics: 1) the phenomenon it attempts to explain has an aspect of variability, and 2) using it includes employment of probabilistic considerations (Brown and Kass, 2009).

The statistical model provides a common referent and language for discussing the system and its structure; it includes both the objectives in the system and their relationships, incorporating the method, meaning-making and cultural activities surrounding it (Henry, 1997; Hestenes, 2010 and McLean, 2001). Models are intentional representations, i.e., a representation is considered a statistical model if used for a descriptive, explanatory, or predictive statistical purpose, which generally means explaining patterns that can be observed in real data or predicting other possible values (Garfield et al., 2008).

From the various constructs of the statistical model proposed by various authors, it can be deduced that there is still a consensus to define this type of model. However, this paper presents the position proposed by McLean (2001, p. 91), who conceives the statistical model as:

a) Predictive: related to the ability to predict what will happen under certain conditions.

b) Deterministic: says what will happen under certain circumstances.

c) Probabilistic: predicts by specifying what can happen, and assigns a probability to each possible case.

d) Causal: provides predictive capability through a theoretical framework that relates the variables involved so that the results can be predicted if one or more variables are changed.

2.2 Statistical Modeling (SM)

The work presented here referred to the contributions of (Lehrer and English, 2018, p. 235), who define EM in an educational context as a particular case of mathematical modeling. Considering that ME has its methodological underpinnings that give it great viability as an area of development in the learning of applied statistics in which the student's ability to:

a) Pose statistical questions within meaningful contexts that highlight variability.

b) Generate, select and measure attributes that vary in light of the questions posed.

c) Collect first-hand data to make research design decisions.

d) Represent, structure and interpret sample and sampling variability.

e) Make informal inferences from all these processes by recognizing uncertainty, detecting variations and making predictions.

2.3 Statistical Reasoning (SR)

SR is how people reason with statistical ideas, making interpretations based on data sets, graphical representations and statistical summaries, making inferences and interpreting results (Garfield, 2002). Therefore, for a student to fully understand and reason, he/she needs to experience various activities, such as verbal explanations and activities that involve deepening statistical concepts in real contexts using technological tools (Garfield and Ben-Zvi, 2008).

Jones et al. (2000, 2001) and Mooney (2002) have developed specific frameworks that characterize students' ER development based on two theories. First, for students to develop ER, they need to understand multifaceted data management concepts that develop over time. For these authors, students' reasoning can be characterized through levels that reflect changes in the complexity of their reasoning.

From this theoretical perspective, Jones et al. (2000) and Mooney (2002) describe students' ER at four levels: idiosyncratic, transitional, quantitative and analytical. Garfield (2002) extends this conceptual framework on the developmental model of students' ER towards learning sampling distributions in five levels: idiosyncratic, verbal, transitional, procedural and integrated processes. These levels proposed by Garfield (2002) allow for assessing students' ER levels from elementary to higher education.

This study considered the general model of SR defined by Garfield (2002), who describes the process of gradual cognitive integration of statistical concepts through five levels:

Idiosyncratic reasoning: The student knows some statistical words and symbols, uses them without fully understanding them, often incorrectly, and may mix them with unrelated information. For example, they calculate the concepts of arithmetic mean, median and mode but do not interpret or relate them to the symmetry of the data.

Level 2. Verbal reasoning. The student has a verbal understanding of some statistical concepts or processes, but cannot relate them in a real context. For example, what happens to the variable under study when the mean is greater than the median and the median is greater than the mode.

Transient reasoning: The student can correctly identify one or two dimensions of a statistical concept or process without fully integrating these dimensions. For example, a large sample size requires a normal distribution and the normal distribution must be symmetric.

Procedural reasoning: The student can correctly identify the dimensions of a statistical concept or process but does not fully integrate them or understand the process. For example, the student knows how to calculate measures of central tendency and perform their interpretation but does not integrate this concept with the normal distribution.

Level 5. Integrated process reasoning: The student thoroughly understands a statistical process, coordinates the rules and behavior, and can explain the process in his/her own words (p. 8). For example, the student can calculate and explain in his or her own words the concepts and the result of obtaining a 95% confidence interval for a large sample size $(n > 30)$.

3. METHODOLOGY

3.1 Research Approach and Design

The research was conducted under a qualitative approach with a multiple case study design. Qualitative because the work is characterized by: a) the participants were students who are training to be engineers and took a course in probability and statistics, b) the instrument for collecting the information was designed specifically for this work based on the constructs of statistical modeling and statistical reasoning, and c) the data analysis was focused on interpreting and making sense of the written statements of the participants when solving problems involving statistical data.

A multiple case study was carried out as a technique and method for the research development. The case study is a research strategy that contributes to the knowledge of social groups. In this case, the groups were formed according to the program they studied: Civil Engineering, Mining Engineering, Electronic Engineering and Civil Works Technology at the Universidad Francisco de Paula Santander. This way of organizing the work focuses on answering questions related to the how? (Yin, 2003). In addition, more than one unit of analysis is involved (Yin, 2003; Kröll, 2004; Stake, 2005) composed by the processes of Statistical Modeling and the Statistical Reasoning Level collected through an instrument called an initial questionnaire.

3.2 Participants

The participants were 38 students, 68% male and 32% female, aged between 17 and 23 years old, who took a Probability and Statistics course during the first semester of the year 2022 and are training to become engineers at the Universidad Francisco de Paula Santander. The access to the participants was done naturally since the researcher is in charge of guiding this subject. Sixty-seven percent of the participants are studying Civil Engineering, 17% Mining Engineering, 8% Electronic Engineering and 8% Civil Works Technology.

3.3 Sources and data collection

With the consent of the participants, an initial questionnaire was applied to the students before orienting the content of the Statistics and Probability subject during the first semester of the year 2022 to identify the ability of ME and analyze the level of RE of the university students.

The initial questionnaire was pilot-tested, adjustments were made, and experts validated it. The students used technological tools such as the Excel spreadsheet to solve the questionnaire.

The instrument is composed of two questions. The first question (Figure 2) is a problem situation corresponding to phases a), b) and c) of the ME: a) posing the statistical question, b) generation, selection and measurement of attributes that vary in light of the question posed, and c) data collection for the student to find decisions. It was also expected that with the information provided, the participants would make the transition to phases d) and e) of the EM characterized by: a) representing, structuring and interpreting the data through sample variability and b) making informal inferences through the development of statistical-descriptive processes that allow them to answer the question posed.

Figure 2.

Question 1: Initial Questionnaire

Universidad Francisco de Paula Santander

INITIAL QUESTIONNAIRE
STATISTICS AND PROBABILITY

1. The director of the Department of Traffic and Transportation of the City of Cúcuta is concerned about the speed of private vehicles that travel through a section of the road ring on Saturdays after 7 pm. Therefore, it sent personnel to take the random records of the speed in Km/h of 45 cars, obtaining the following data:

58	32	45	46	42	39
31	48	49	56	52	39
44	42	38	52	55	58
56	58	48	47	52	37
38	29	62	49	69	18
18	61	48	55	49	29
47	69	15	68	48	62
61	55	64			

The safest speed margins for this road are more than 36 km / h and less than 59 km / h for private vehicles How can the Traffic Director know if the behavior of citizens is within these speed ranges? With the information given, identify the study population, the sample, the variable of the data and realize a complete statistical study for pooled data (you can consider the use of frequency table, graphs, and basic and robust statistical concepts) that allows you to help the Director of Transit and Transportation to present a complete report of the problem that is evident and answer the question posed. Justify your response.

Source: Own elaboration.

While with the second question (Figure 3), students were expected to evidence capabilities in developing the EM phases.

Figure 3.

Question 2: Initial Questionnaire

Universidad Francisco de Paula Santander

INITIAL QUESTIONNAIRE
STATISTICS AND PROBABILITY

2. Student, identify a problem situation related to your profession, to that problem situation pose a statistical question and collect data. With the data collected, identify the variable, population, sample, perform a statistical study (table, graphs), interpreting the results and conclude by answering the question. Justify every step you take.

Thank you very much for your participation.

4. DATA ANALYSIS AND RESULTS

From a didactic point of view, the initial questionnaire was developed in groups of three students from the same program; as Morgan (1996) reiterates, one of the best ways to collect qualitative data is by involving small numbers of people through focus groups.

As a reference for the analysis of the results of the first question of the initial questionnaire, the last two levels of ME were taken into account: representing, structuring and interpreting sample variability and making informal inferences in the light of all these processes, recognizing uncertainty, detecting variations and making predictions. Table 1 shows the results after having interpreted and made sense of the work done by the students regarding the last two levels of statistical modeling and statistical reasoning processes.

Table 1.

Results of the First Question of the Initial Questionnaire

Group	**ME**		**RE**
	Phase d) representing, structuring and interpreting sample variability	**Phase e) making informal inferences**	
4, 5 y 11	Correct process: 1. Identified the type of variable the data represent 2. Identified that they are sample data 3. Data sorted 4. They made a table of grouped data with their respective interpretation. They plotted the histogram with its respective interpretation. Correctly calculated and interpreted the Measures of Central Tendency (Mean, Median and Mode).	By correctly developing phase d) of the ME, only group 11 could infer the data and answer the question.	Groups 4 and 5 reached the verbal SR level, because they correctly understood some statistical concepts but did not relate it to answer the question posed. While group 11 reached the level of Transitory ER, because they developed correctly the EM processes and with the analysis of the results they were able to infer and answer the question.
2 y 7	The process was incorrect: 1. Presented grouped data table 2. The interpretations of the graphs are not related to the question. 3. They calculated the average, but the interpretation is unrelated to the question.	They did not respond to the questionnaire question; that is, the students did not reach this process.	The level of SR was verbal, since: they made interpretations to the data table, diagrams and the average, but these results were not integrated to infer about the data and answer the question.
1, 3, 6, 8, 9, 10 y 12	The process was incomplete and incorrect: 1. They presented a table of	They did not respond to the questionnaire question; therefore, this	The level of SR was idiosyncratic, since: they performed procedures

non-clustered data. 2. Bar chart without interpretation 3. Groups 8 and 12 performed circular diagrams of the continuous quantitative variable.	process was not reached by the students.	incompletely and incorrectly; consequently, they failed to infer or answer the question posed.

The Figure shows the work done by students in groups 4, 5 and 11, who could perform phase d) of the statistical modeling by correctly developing the table of grouped data with its respective interpretation.

Figure 1.

Example: Frequency table for grouped data

REPRESENT, STRUCTURE AND INTERPRET DATA

Groups 4, 5 and 11

BUILDING INTERVAL TABLES

	Class intervals		Marca de Clase				
N° de Class	Lower Lim	Lim Superior	Xi	fi	Fi	hi	Hi
0	7	15	11	0	0	0,0000	0,00
1	15	25	20	3	3	0,0667	0,07
2	25	35	30	4	7	0,0889	0,16
3	35	45	40	8	15	0,1778	0,33
4	45	55	50	14	29	0,3111	0,64
5	55	65	60	13	42	0,2889	0,93
6	65	75	70	3	45	0,0667	1,00
7	75	80	77,5	0	45	0,0000	1,00

Table 1. Speed of private drivers driving along a stretch of the ring road on Saturdays after 7 pm

Interpretation

1) It is observed in table 1 of grouped data that the minimum speed in the road ring at 7 pm is 15 km / h and the maximum is 69 km \ h where the largest number of cars go between 45 and 55 km / h

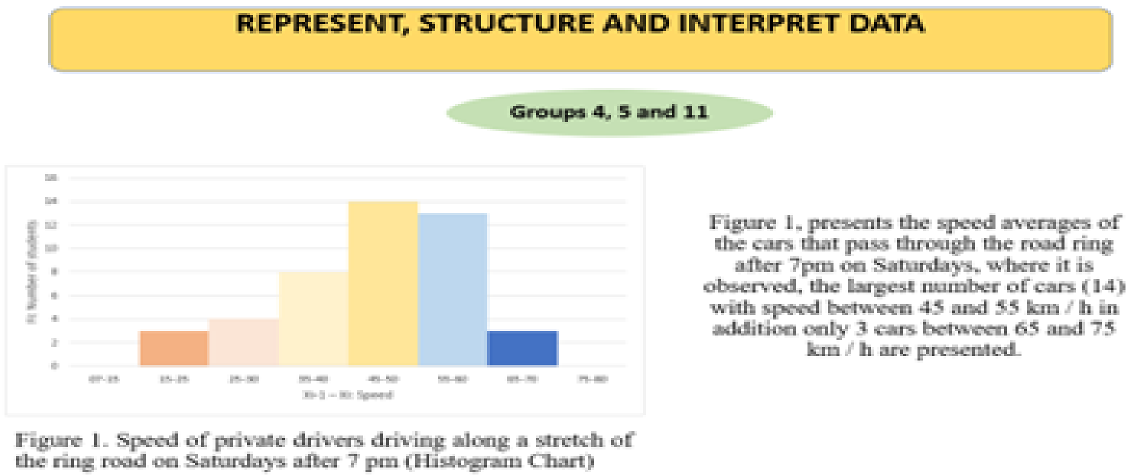
2) The table presents 29 cars or less with an average equal to or less than 55 km / h being the average 54km /

3) 33% of cars have an average equal to or less than 40 km/h with a minimum average of 15 km/h

Figure 2 shows the capacity of groups 4, 5 and 11 to develop phase d correctly), of the ME by representing with the histogram the data of the variable speed of the first point, where the RE capacity of groups 4, 5 and 11, evidenced when interpreting the statistical models of Figures 5 and 6 made in the Excel spreadsheet.

Figure 2.

Example: Result of the first question of the Initial Questionnaire - Histogram



The analysis of the second question is shown in Table 2, where each group of students will identify a problem situation related to their professional profile to develop the ME phases where the RE Level will be verified with the interpretations made.

Table 2.

Results of the second question of the Initial Questionnaire

Group	ME		RE
	Description	ME Phases	
4,7 y 9	They presented the problem situation related to their profile. For example, group 9 posed the following question, "What is the average square footage of the housing plot of civil engineering students?" The statistical concepts developed were: population, sample, variable, frequency table, bar chart and frequency polygon.	They posed the statistical question. Data were collected. They represented, structured and interpreted sample variability. They made informal inferences	For the ER level, it was identified as follows: • Group 3 and 10: Idiosyncratic. They know some statistical concepts without relating them to the context. • Groups 1, 5, 8: Verbal. An understanding of statistical-descriptive concepts not related to their professional profile was evidenced.
1,3,5,8 y 10	They presented problem situations not related to their	They posed the statistical question.	• Groups 4, 7 and 9: Transitory.

	professional profile. For example, from an academic field, group 5 posed the following statistical question: "How did the students of Civil Engineering at the Universidad Francisco de Paula Santander do in the second semester of statistics and probability? The statistical concepts used to answer the exercise were: population, sample, variable, pie chart and bar chart.	Data were collected. They represented, structured and interpreted sample variability. Only group 8 could conclude and make informal inferences, answering the question.	The results show mastery of statistical concepts with interpretations that allowed him to answer the statistical question posed.
2 y 12	They did not answer item 2. Instead, they argued that they lacked time.		
6 y 11	They did not answer item 2 due to the lack of ability to contextualize a problem situation related to their profile.		

To better illustrate this analysis, Figure 6 shows the production of group 9, who collected information from a statistical question related to their professional profile; this question was sent to their classmates through the WhatsApp group previously created, and they collected 29 data. This information allowed them to use two statistical models: the frequency table for grouped data and the arithmetic mean, achieving an answer to the question posed with its respective interpretation; from which the students' RE capacity is evident.

Figure 3.

Example Result second item Initial Questionnaire

The average in square meters of the housing land of the students of the Civil Engineering career of the UFPS is?

Parameter	Values
n	30
Maximum value	200
Minimum value	25
Range	175
m # intervals	6
Amplitude	30

BUILDING INTERVAL TABLES

	Class intervals							
N° de Class	Lower Lim	Lim Superior	Xi	fi	xi*fi	Fi	hi	Hi
0	0	25	12,5	0	0	0	0,0000	0,00
1	25	55	40,0	16	640	3	0,3556	0,36
2	55	85	70,0	7	490	7	0,1556	0,51
3	85	115	100,0	4	400	15	0,0889	0,60
4	115	144	129,5	0	0	29	0,0000	0,60
5	144	174	159,0	1	159	42	0,0222	0,62
6	174	204	189,0	0	0	45	0,0000	0,62
7	204	234	219,0	2	438	45	0,0444	0,67
				n= 30	2127			

Arithmetic average $\bar{x} = \frac{\sum(xi*fi)}{n} = 70{,}9$

1. The average in square meters of student housing of 70.9
2. The table presents 29 cars or less with an average equal to or less than 55 km / h being the average 54 km / h
3. The 33% of cars have an average equal to or less than 40 km/h with a minimum average of 15 km/h

5. DISCUSSION

The results obtained in the initial questionnaire, applied as a group to the Faculty of Engineering enrolled in the subject of statistics and probability without any conceptual orientation, allowed comparing the initial conceptions of the EM processes and the levels of SR. According to Alston-Knox et al. (2019), it is necessary to identify what kind of skills and knowledge in statistics and probability students have when entering the various university programs to have a conceptual formation referent.

Therefore, it was evidenced in the results in each group the lack of ability to interconnect data in real contexts; that is, students lack conceptual skills to represent data with statistical models, including a lack of knowledge in the use of technological tools that allow them to develop ME processes and improve RE capacity (Biehler et al., 2018 and Verdejo et al., 2021).

This confirms the statements of Lehrer and Schauble (2019); Biehler et al. (2018); Pfannkuch et al. (2018); Doerr et al. (2017); Radke et al. (2020) and Divir and Ben-Zvi (2021) on the need to do research that generates a paradigm shift where didactic strategies are implemented in ME so that the student can identify a problem situation and from there construct a statistical question that leads him/her to collect data in a real context, incorporating processes such as decision making, prediction and inference when collecting, exploring data and beginning to analyze the existence of variation, data reduction, population parameters considering the samples, the logic of sampling, processes and causal factors, improving the ability of statistical reasoning in the university student.

6. Conclusions

In this article, a comparison was made between the initial conceptions of the EM process, and the level of SR in students of the Faculty of Engineering enrolled in the statistics and probability course. The instrument was applied to 38 students distributed in 12 groups; of the 12 groups 8 belonged to the Civil Engineering program representing 67%; the other programs were Mining Engineering, Electronic Engineering and Civil Works Technology.

25% of the groups (4, 5 and 11) developed the processes correctly when solving the first item of the initial questionnaire. Group 11 is related to a transient SR level because they developed the EM processes correctly, and with the analysis of the results, they answered the question posed, while groups 4 and 5 reached a verbal SR level because they did not integrate the EM processes to answer the question posed.

The other 75% of the groups that answered the first item failed to develop the EM processes correctly. Groups 2 and 7 are related to a verbal SR level by presenting some statistical concepts correctly but were not integrated to answer the statistical question and 7 groups (1, 3, 6, 8, 9, 10 and 12) are related to an idiosyncratic SR level by not having the ability to perform the EM processes correctly; consequently, they failed to infer or answer the question posed.

For developing the second item of the initial questionnaire, 67% (8 groups) managed to address it. Of these 8 groups, 25% (groups 4, 7 and 9) presented a problem situation and a statistical question related to their professional profile, reaching a transient SR level. The remaining 4 groups did not answer the second point, 2 groups (2 and 12) argued that they lacked the time, and the other two groups (6 and 11) stated that they could not pose a question and collect data related to their professional profile.

The relationship between the initial conceptions of EM and SR allowed identifying three levels of SR in the groups of students, 33% corresponding to the idiosyncratic level, 42% to the verbal level, and 5% to the transitional level. In addition, of the 12 groups of students, none of them managed to go through all the phases of EM successfully, nor did they reach the highest ER level 5 called integrated reasoning of the process, in which the student is expected to fully understand a process starting from a statistical question, collect data and identify their behavior, apply statistical concepts with their respective interpretations, and draw conclusions in response to the problem situation posed.

The analysis of the results of the present study raises many questions; among them, the following stand out: What are the students' difficulties in not managing to go through the EM phases? Why was the maximum ER level reached by one group in the first question and three groups in the second question of the initial questionnaire, the transitory one? What effects can developing all the EM phases have on university students? What kind of relationship can exist between having the ability to go through the EM phases and the ER levels?

The results of this research evidence and agree with the literature consulted and cited, about the need to move forward with the implementation of the processes of ME and the levels of RE through didactic strategies with technological tools, providing an instructional sequence of activities so that the university student achieves to build, modify and apply representations with dynamic conceptual models related to their context and professional profile, improving the capacity of RE.

References

[1] Alston-Knox, C., Stricklan, C., Gazos, T., & Mengersen, K. (2019). Teaching and Learning in Statistics: Harnessing the power of modern statistical software to improve students statistical reasoning and thinking. *In HEAD'19. 5th International Conference on Higher Education Advances* (págs. 1171-1178). València: Editorial Universitat Politècnica de València. https://doi.org/10.4995/HEAD19.2019.9239

[2] Biehler, R., Frischemeier, D., Reading, C., & Shaughnessy, J. (2018). Reasoning about data. En D. Ben-Zvi, K. Makar, & J. Garfield, *International handbook of research in statistics education* (págs. 139-192). International handbook of research in statistics education.

[3] Bransford, J., Brown, A., & Cocking, R. (2000). *How people learn* (Vol. 11). Washington, DC: National academy press.

[4] Brown, E., & Kass, R. (2009). What is statistics? *The American Statistician, 63*(2), 105-110. https://doi.org/10.1198/tast.2009.0019

[5] Cazco, G., Olalla, M., & Abad, F. (2018). Modelos didácticos en la educación superior: una realidad que se puede cambiar. *Profesorado, Revista de Currículum y Formación del Profesorado, 22(2)*, 447-469. https://doi.org/10.30827/profesorado.v22i2.7732

[6] Doerr, H., Delmas, R., & Makar, K. (2017). Modeling approach to the development of students'informal inferential reasoning. *Statistics Education Research Journal, 16*(2), 86-115. https://doi.org/10.52041/serj.v16i2.186

[7] Dvir, M., & Ben-Zvi, D. (2021). Informal statistical models and modeling. *Mathematical Thinking and Learning, 23*(1), 1-21. https://doi.org/10.1080/10986065.2021.1925842

[8] Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of statistics education, 10*(3), 1-12. https://doi.org/10.1080/10691898.2002.11910676

[9] Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice.* San Luis Obispo, USA: Springer Science & Business Media.

[10] Garfield, J., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., & Zieffler, A. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice.* Springer: New York. https://doi.org/10.1007/978-1-4020-8383-9

[11] Henry, M. (1997). Notion de modéle et modélization en l'enseignement. *Enseigner les probabilités au lycée*, 77-84. https://publimath.univ-irem.fr/numerisation/WV/IWV01006/IWV01006.pdf

[12] Hestenes, D. (2010). Modeling theory for math and science education. *In Modeling students' mathematical modeling competencies (pp. 13-41)*, Springer, Boston, MA.

[13] Jones, G., Langrall, C., Thornton, C., Mooney, E., Wares, A., Jones, M., & Nisbet, S. (2001). Using students' statistical thinking to inform instruction. *The Journal of Mathematical Behavior, 20*(1), 109-144.

https://doi.org/10.1016/S0732-3123(01)00064-5

[14] Jones, G., Thornton, C., Langrall, C., Mooney, E., Perry, B., & Putt, I. (2000). A framework for characterizing children's statistical thinking. *Mathematical thinking and learning, 2*(4), 269-307. https://doi.org/10.1207/S15327833MTL0204_3

[15] Kazak, S., & Pratt, D. (2017). Pre-service mathematics teachers' use of probability models in making informal inferences about a chance game. *Statistics Education Research Journal, 16*(2), 1 - 18.

[16] Kröll, H. (2004). El método de los estudios de caso. Observar, escuchar y comprender. En M. Tarrés, *Observar, escuchar y comprender sobre la tradición cualitatitva en la investigación* (págs. 231-264). El Colegio de México-FLACSO México.

[17] Lehrer, R., & English, L. (2018). Introducing children to modeling variability. En D. Ben-Zvi, K. Makar, & J. Garfield, *In International handbook of research in statistics education* (págs. 229-260). Springer, Cham.

[18] Lehrer, R., & Schauble, L. (2019). Learning to play the modeling game. En A. zu Belzen, D. Krüger, & J. van Driel, *Towards a competence-based view on models and modeling in science education* (págs. 221-236). Cham: Springer.

[19] McLean, A. (2001). Statistics in the catwalk. The importance of models in training researchers in statistics. *Training researchers in the use of statistics*, 87-102.

[20] Miller Jr, H. (2016). *he Sage encyclopedia of theory in psychology.* SAGE Publications.

[21] Mooney, E. (2002). A framework for characterizing middle school students' statistical thinking. *Mathematical Thinking and Learning, 4*(1), 23-63. https://doi.org/10.1207/S15327833MTL0401_2

[22] Morgan, D. (1996). *Focus groups as qualitative research* (Vol. 16). Sage publications.

[23] Pfannkuch, M., Ben-Zvi, D., & Budgett, S. (2018). Innovations in statistical modeling to connect data, chance and context. *ZDM, 50*(7), 1113-1123. https://doi.org/https://doi.org/10.1007/s11858-018-0989-2

[24] Pfannkuch, M., Budgett, S., Fewster, R., Fitch, M., Pattenwise, S., Wild, C., & Ziedins, I. (2016). Probability modeling and thinking: What can we learn from practice? *Statistics Education Research Journal, 15*(2), 11-37.

[25] Porras Lizano, K., & Castro-Rodríguez, E. (2021). Errors Middle School Students Make When Performing Mathematical Modelling Tasks. *Acta Scientiae, 23*(2), 29-57. http://hdl.handle.net/11056/21451

[26] Radke, S., Vogel, S., & Hoadley, C. (2020). Representing percents and personas: Designing syncretic curricula for modeling and statistical reasoning. *International Society of the Learning Sciences (ISLS)*, 1365-1372. https://doi.org/https://doi.dx.org/10.22318/icls2020.1365

[27] Sawyer, R. (2006). Educating for innovation. *Thinking skills and creativity, 1*(1), 41-48. https://doi.org/10.1016/j.tsc.2005.08.001

[28] Stake, R. (2005). Qualitative case studies. En N. Denzin, & Y. Lincoln, *The SAGE Handbook of Qualitative Research* (págs. 443-446). Thousand Okas: SAGE Publications.

[29] Tacoma, S., Sosnovsky, S., Boon, P., Jeuring, J., & Drijvers, P. (2018). The interplay between inspectable student models and didactics of statistics. *Digital Experiences in Mathematics Education, 4*(2), 139-162. https://doi.org/https://doi.org/10.1007/s40751-018-0040-9

[30] Verdejo, A., Arenas, M., & Uclés, R. (2021). Errores de profesores de Matemáticas en formación inicial al resolver una tarea de modelización. *PNA. Revista de Investigación en Didáctica de la Matemática, 15*(2), 109-136. https://doi.org/ https://doi.org/10.30827/pna.v15i2.20746

[31] Vosniadou, S. (2012). Reframing the classical approach to conceptual change: Preconceptions, misconceptions and synthetic models. En B. Fraser, K. Tobin, & C. McRobbie, *Second international handbook of science education* (págs. 119-130). Springer, Dordrecht.

[32] Wild, C., Utts, J., & Horton, N. (2018). What Is Statistics? En D. Ben-Zvi, K. Makar, & J. Garfield, *International handbook of research in statistics education.* (págs. 5-36). Cham: Springer.

[33] Woolfolk, A. (2020). *Educational Psychology* (Vol. 20). Global Edition. Pearson Education, Limited.

[34] Yin, R. (2003). *Case study research: Design and method.* Thousand Oaks, CA: SagePublications.