Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2025-23.1.0059

RESEARCH ARTICLE

Correlation and Regression in Secondary Education: Textbook Analysis

Audy Salcedo*[1], Jaime I. García-García[2], Danilo Díaz-Levicoy[3], Yudi Caterine Diaz-Perdomo[4]

[1] Interdisciplinary Research Group on Early Childhood Education, Autonomous University of Chile, Talca, Chile.
[2] Department of Mathematics, Metropolitan University of Education Science, Santiago, Chile.
[3] Centre for Research in Mathematics and Statistics Education, Catholic University of Maule, Talca, Chile.
[4] Doctoral Programme in Mathematics Education, Catholic University of Maule, Talca, Chile.

ARTICLE INFO

Received: Sep 18, 2024

Accepted: Dec 10, 2024

Keywords

Association

Correlation

Regression

Textbooks

Secondary education

***Corresponding Authors:**

audy.salcedo@gmx.com

ABSTRACT

This study compares the situations proposed to address the topic of correlation and regression in mathematics textbooks from Colombia, Chile, Ecuador and Venezuela. For this purpose, we analyze how four Secondary Education textbooks from Latin American countries present situations related to association; specifically, we study the situations by identifying the problem field to which they belong, the intensity and the direction of the relationship. Under a qualitative methodology, a content analysis of the situations was carried out, considering the examples and problems proposed to the student. The results indicate that the books emphasize the analysis of the existence of a relationship between variables, while the estimation of one variable as a function of another is infrequent. In addition, situations with high relationship intensity and direct direction prevail. It is concluded that there is a need to strengthen the teaching of these contents in the textbooks analyzed.

INTRODUCTION

Most countries have incorporated statistics as part of students' comprehensive education, aiming to develop competencies that enable them to understand and interpret statistical information presented through various media, as well as to communicate and discuss such information effectively (Gal, 2004; Salcedo et al., 2021; García-García, 2021; Conejero et al., 2023; Esquivel, 2016). The goal is for students to achieve proper statistical literacy, which, according to Gal et al. (2020), refers to how individuals handle the mathematical, quantitative, and statistical demands of adult life; it is a product of schooling and a fundamental skill for all adults. Engel (2019) adds that statistical education can support civic participation in modern societies and help combat misinformation that threatens democracy.

In this regard, focusing teaching on fundamental statistical concepts (Watson et al., 2018; Batanero et al., 2013) and addressing them throughout students' education, in various contexts and with multiple representations, could help students recognize the underlying structure of statistical knowledge (Garfield & Ben-Zvi, 2008). One such concept is bivariate distributions, the study of which is recommended starting in secondary education. This is because they help describe the relationships

between statistical variables in various contexts, although their interpretation can be complex and often poses a teaching challenge (Batanero & Borovcnik, 2016; Engel & Sedlmeier, 2011). Additionally, the study of correlation and regression takes advantage of the potential of mathematical modeling, bridging the gap between mathematics and statistics (Burrill & Biehler, 2011; Fergusson & Pfannkuch, 2022).

Furthermore, textbooks are highly significant pedagogical resources (Herbel, 2007; Cordero & Flores, 2007) used to develop students' statistical competence. While it is clear that these resources alone do not determine what will be taught or learned, there is no doubt about their influence on curriculum implementation and the selection of learning activities (Fan et al., 2013; Rezat et al., 2021; Qi et al., 2018; Díaz-Levicoy et al., 2018; Olsher & Even, 2014; Valverde et al., 2002). The space and depth textbooks dedicate to a particular curriculum topic often reflect the amount of time students will spend working on it. Moreover, textbook analysis is considered a research line in the Didactics of Statistics (Díaz-Levicoy et al., 2017), with a promising outlook and a solid foundation (Rodríguez-Muñiz and Díaz, 2018). Consequently, it directly relates to students' learning opportunities and their achievements in mathematics (Hadar, 2017; Stylianides, 2009; Wijaya et al., 2015).

Background

Research focused on analyzing activities related to correlation and regression proposed in textbooks is limited. Below, we briefly present some studies to contextualize our research.

Estepa and Sánchez-Cobo (1998) examined the theoretical presentation and activities on correlation and regression in Spanish secondary education textbooks. The authors reported a predominance of the formal classical methodological approach of theory-practice, as well as the lack of or inadequate treatment of fundamental concepts. They also identified biases toward high and direct correlations in the activities. Similarly, Lavalle et al. (2006) conducted a didactic analysis of Argentine high school textbooks. The authors noted that most activities described a high and direct association without incorporating technological resources.

Gea et al. (2013, 2017) studied the situations used to contextualize correlation and regression in Spanish high school textbooks, observing a predominance of tasks where students analyze the existence of relationships between variables, often accompanied by activities for data organization/reduction. The authors reported a wide variety of contexts used to present the topic but highlighted a significant number of decontextualized exercises. Additionally, they observed limited use of technology, despite it being included in the recommendations of curricular guidelines. Batanero et al. (2017) analyzed concepts associated with correlation and regression in Spanish high school textbooks, identifying instances where the presentation of these concepts was partially correct or incomplete, or where non-equivalent concepts were equated.

This literature review identified studies comparing textbooks from the same country on the topic of statistical association between variables but found no studies contrasting textbooks from different countries.

MATERIALS AND METHODS

In this study, the activities on correlation and regression proposed in secondary education textbooks from four Latin American countries—Colombia, Chile, Ecuador, and Venezuela—were analyzed. A qualitative methodology was employed (Fraenkel & Wallen, 2006), using content analysis (Stempler, 2001) of the examples and exercises included in the textbooks. These resources were selected based on their current use and authorization by each country's Ministry of Education, making the sample intentional and non-probabilistic (Hernández et al., 2014). Additionally, it is worth noting that the selection of textbooks corresponds to the school year in which the topic is first introduced (Table 1).

Table 1. Textbooks Analyzed

Country	Authors (year)	Title	Editorial	Age
Chile (T1)	Fresno, Torres y Ávila (2020)	Matemática. Texto del estudiante. 1ro. Medio [Mathematics. Student text. 1st. Middle]	Santillana	15 years
Colombia (T2)	Ministry of National Education [MEN] (2017)	Vamos a aprender matemáticas. Libro del estudiante 9 [We are going to learn mathematics. Student's Book 9]	SN	15 years
Ecuador (T3)	Ministry of Education of Ecuador [MEE] (2016)	Matemática. 3ro Bachillerato General Unificado. Serie Ingenios Mathematics. 3rd Unified General Baccalaureate. Ingenios Series]	Don Bosco	17 years
Venezuela (T4)	Ministry of People's Power for Education [MPPPE] (2014)	La Matemática y el Vivir Bien. Matemática 5to año. Colección Bicentenario [Mathematics and Living Well. Mathematics 5th year. Bicentenary Collection]	MPPPE	17 years

It should be highlighted that the textbooks from Chile, Colombia, and Ecuador are produced by private publishers, contracted, and later distributed to schools. In contrast, the Venezuelan textbook was developed and produced directly by the Ministry. In all the analyzed textbooks, the topic of correlation and regression is introduced during the final years of secondary education, though there are some differences in the students' estimated age. When studying this topic, students in Colombia and Chile are expected to be 15 years old, while those in Ecuador and Venezuela are generally 17 years old.

In each textbook, all activities on correlation and regression were identified, and the problem domain, intensity, and direction of the relationship were analyzed. For classifying the problem domains, the framework proposed by Gea (2014) was used:

- P0. *Organization/representation of bivariate data*. This involves summarizing or representing bivariate data sets using graphs or statistical tables and includes interpreting the representations.
- P1. *Analyzing the existence of a relationship between variables*. These tasks allow students to examine the potential association between variables.
- P2. *Predicting one variable based on another*. These activities involve finding a function that allows one variable to be derived from the other, including calculating coefficients for regression lines, writing regression equations, or making predictions using those equations.

The intensity and direction of the correlation were additional units of analysis used to study the activities. For exercises, calculations were performed to determine the correlation coefficient and establish the intensity and direction of the relationship. For examples or solved exercises, the information provided in the textbook was used. The intensity of the relationship was classified as follows:

- *Independence*. The linear correlation coefficient is zero.
- *Low*. The correlation coefficient falls within the interval [0.1, 0.5).
- *Medium*. The correlation coefficient falls within the interval [0.5, 0.8).
- *High*. The correlation coefficient falls within the interval [0.8, 1).
- *Functional*. When the relationship is perfect, the correlation coefficient is +1 or -1. Observed data exactly matches theoretical data, and all points in the scatter plot fall on a single regression line.

- *Verbal description*. Situations where calculating the correlation coefficient is not required or feasible, but the nature of the relationship must be described.

To classify the direction of the relationship, the following categories were considered:

- *Independence*. There is no relationship between the variables, making it impossible to assign a direction.
- *Direct*. Increases (or decreases) in one variable are accompanied by increases (or decreases) in the other.
- *Inverse*. Increases in one variable are accompanied by decreases in the other.
- *Other*. Nonlinear relationships, such as exponential or parabolic, are included.
- *Verbal description*. Situations where the direction of the relationship cannot be determined but must still be described.

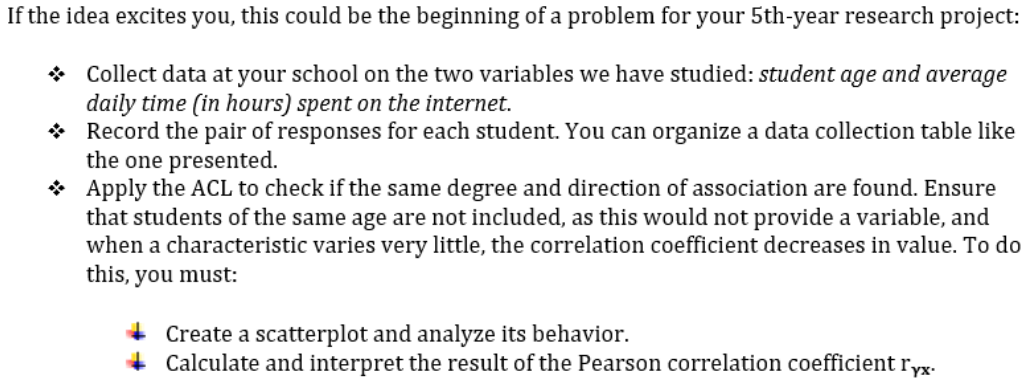
Each researcher independently analyzed and classified all activities on correlation and regression from each textbook. These classifications were then discussed until consensus was reached, resulting in a unified classification. It should also be noted that some activities contained two or more tasks related to different problem domains, so the counting and analysis were conducted exclusively. The results of the analysis of the selected textbooks are presented in the following section.

RESULTS

The results of the analysis of situations involving correlation and regression proposed in the analyzed textbooks are presented below, organized by data treatment categories.

Problem field

First, the situations are analyzed based on the problem field to which they belong. In Figure 1, a situation is shown with several tasks. Among them, students are asked to collect data about their classmates' ages and the hours they spend using the Internet, then construct a scatter plot. This latter task is considered part of problem field P0 (organization/representation of bivariate data) since a graph is used to represent the set of bivariate data.



If the idea excites you, this could be the beginning of a problem for your 5th-year research project:

- Collect data at your school on the two variables we have studied: *student age and average daily time (in hours) spent on the internet*.
- Record the pair of responses for each student. You can organize a data collection table like the one presented.
- Apply the ACL to check if the same degree and direction of association are found. Ensure that students of the same age are not included, as this would not provide a variable, and when a characteristic varies very little, the correlation coefficient decreases in value. To do this, you must:
  - Create a scatterplot and analyze its behavior.
  - Calculate and interpret the result of the Pearson correlation coefficient r_{yx}.

Figure 1. Situation of field P0 and P1 (T4, p. 16)

Subsequently, students are asked to analyze the graph, aiming to decide whether or not there is a potential statistical relationship between the variables. This task belongs to problem field P1 (analyzing the existence of a relationship between variables). Following this, students are instructed to calculate Pearson's correlation coefficient, which involves determining the magnitude of the linear relationship between the variables. Finally, they must interpret the coefficient, considering its magnitude, direction, and context.

In Figure 2, the first question corresponds to problem field P2 (predicting one variable based on another) since students are required to find the regression line that allows them to estimate the effect time of an antipyretic medication based on the dosage provided. In a second question, students are asked to evaluate a new value to estimate the time the medication will take to take effect.

1. In a hospital, a medication that regulates body temperature is being tested. To do this, different doses of the product are administered to 10 patients with high fever, and the time it takes for their temperature to fully normalize is observed. The following results are obtained:

Answer:
How long would it take for a patient's temperature to normalize if they are given 11.5 mg of the medication? And what if they take a dose of 25 mg?

Dose (mg)	Time (min)	Dose (mg)	Time (min)
2	136	12	60
4	126	14	55
6	115	16	42
8	98	18	38
10	75	20	31

Figure 2. Situation of field P2 (T3, p. 208)

Table 2 summarizes the problem fields identified in the analysis. According to the results, the most common situations belong to field P1 (56.7%), which are more than double those in P0 (26.6%) and slightly more than triple those in P2 (16.7%). Within each country, the predominance of P1-type situations is also evident. For example, in the Chilean textbook, P1 situations (53 out of 83) are more than double the number of P0 (24 out of 83), while the latter are four times the number of P2 (6 out of 83). In this regard, Gea et al. (2017) also found a higher presence of P1 situations in Spanish textbooks, although in their case, the second most common were P2 situations.

Table 2. Frequency (percentage) of situations by problem field

Problem field	Chile (T1)	Colombia (T2)	Ecuador (T3)	Venezuela (T4)	Total
P0	24 (28,9)	19 (34,5)	6 (16,7)	5 (17,2)	54 (26,6)
P1	53 (63,9)	27 (49,1)	22 (61,1)	13 (44,8)	115 (56,7)
P2	6 (7,2)	9 (16,4)	8 (22,2)	11 (37,9)	34 (16,7)
Total	83 (100)	55 (100)	36 (100)	29 (100)	203 (100)

There is a clear inclination in this group of textbooks toward analyzing the existence of relationships between variables, suggesting a greater emphasis on correlation, which seems logical given it is the first encounter with bivariate data analysis. However, one would expect a higher number of P0 situations, as they are essential for developing P1 and P2, among other reasons because students need to interpret data representations and identify potential relationships between variables. In this regard, the Colombian textbook shows a better balance among the three problem fields. The Chilean textbook is the only one that studies two data sets on the same scatter plot (P1).

Venezuela's textbook has the highest percentage of P2 situations (11 out of 29), which are important for understanding the statistical modeling of bivariate data. However, it only considers one regression line, estimating y based on x, which suggests that the relationship between the variables is mathematical dependence, potentially leading to a semiotic conflict (Gea et al., 2015). The Chilean textbook has the lowest percentage in this field (6 out of 83), as it focuses on developing the intuitive notion of correlation.

Research frequently reports that students face confusion between predictor and explanatory variables, difficulties in interpreting the parameters of regression lines, and challenges in

understanding their relationship with the slope of the line and the type of correlation (Batanero et al., 2015; Estepa & Sánchez, 2003). Therefore, it seems appropriate to pay more attention to P2-type situations and their inclusion in secondary education textbooks, given the opportunities they provide for using regression in various modeling and prediction scenarios (Engel & Sedlmeier, 2011).

The textbook with the highest number of situations across all three problem fields is the Chilean textbook, while the Venezuelan textbook contains the fewest. However, the total number is far below those identified by Gea et al. (2013) in a similar analysis of Spanish textbooks, where a minimum of 176 tasks was found in one of the analyzed books.

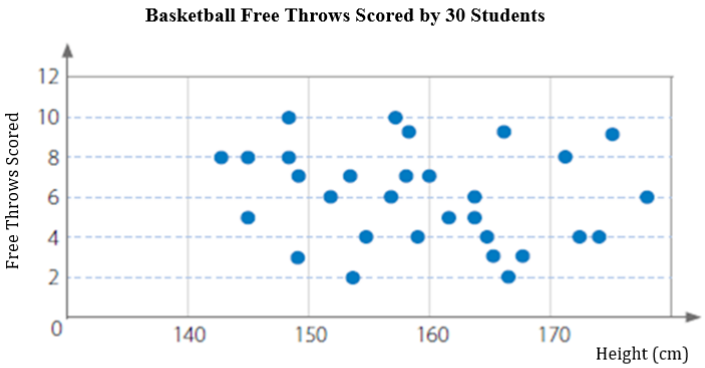
Intensity of the relationship

Figure 3 presents an example of a situation with independence in the intensity of the relationship, which is found in the Chilean textbook. In this example, two variables are presented that could be considered a case of spurious correlation, as it attempts to link height with free-throw performance in basketball. After displaying the graph, it is noted that there is neither a positive nor negative correlation, and the student is asked for their opinion on that statement.

Example 4

A group of students claims that greater height does not guarantee better performance in executing a free throw in basketball. To demonstrate this, they suggested conducting a study with a sample of 30 students of different heights, asking them to shoot ten free throws each.

They recorded the results in the following graph:



The students conclude that since there is neither a positive nor a negative correlation, it confirms that greater height does not guarantee better performance in executing a free throw.

What do you think about the students' conclusion?

Discuss it with a classmate.

Figure 3. Situation with independence in the intensity of the relationship (T1, p. 150)

An example of a situation with low relationship intensity is shown in Figure 4. In this case, the task involves determining the type of relationship among three groups of bivariate variables. For all three relationships, the correlation coefficient is less than 0.5.

Indicate what type of relationship the two-dimensional variables (x, y_1), (x, y_2) and (x, y_3) in Table 4.35 have.

x	23	21	2	4	5	7
y_1	9	5	21	25	27	211
y_2	4	3	2	1	0	21
y_3	22	2	21	1	5	3

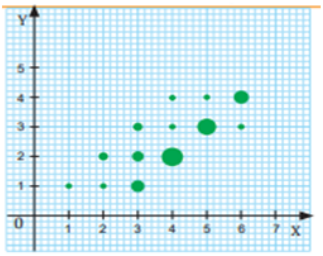
Table 4.35

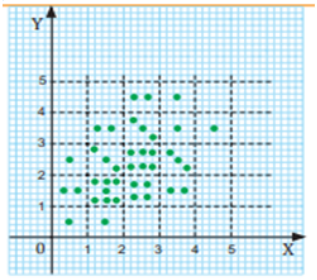
Figure 4: Situation with low relationship intensity (T2, p. 119)

In Figure 5, two scatter plots are shown. The plot on the left provides an example of a situation with medium relationship intensity ($0.5 \leq r < 0.8$), while the plot on the right represents an example of low relationship intensity ($0.1 \leq r < 0.5$).

Scatterplot or Point Cloud Diagram

The data from a bidimensional distribution can be represented on Cartesian axes in the plane.





In the case of ungrouped data, points are typically drawn with an area proportional to the absolute frequency of each pair of values that the bidimensional variable (X, Y) can take.

The figure reflects the scatterplot or point cloud diagram corresponding to the data from Table 11.

In the case of ungrouped data, points are placed in each cell, determined by a pair of class intervals (one for X and one for Y), in quantities proportional to the absolute frequency of that pair of intervals.

The figure shows the scatterplot or point cloud diagram corresponding to the data from Table 12.

Figure 5. Situation with medium relationship intensity (left diagram) and with low relationship intensity (right diagram) (T3, p. 201)

Figure 6 illustrates the potential relationship between height and shoe size for ten individuals. The graph reveals that the scatter plot follows a positive linear trend, and it is explained that the bivariate variable represents a positive correlation. Although the magnitude of the correlation is not specified in the example, calculations conducted by the authors indicate that the relationship intensity is high ($0.8 \leq r < 1$).

Example 3

In Table 4.32, the height in centimeters of ten people and the shoe size they use are shown.

Height (cm)	154	156	157	158	160	163	165	170	171	172
Shoe Size	35	36	36	36	37	38	38	40	41	41

Table 4.32

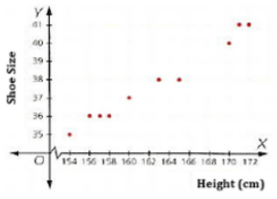


Figure 4.25

When representing the pairs of values of the bidimensional statistical variable (Height, Shoe Size), a set of points is obtained that represents a positive correlation, as shown in Figure 4.25.

Figure 6. Situation with high relationship intensity (T2, p. 117)

Figure 7 presents an example of a situation involving functional relationship intensity. Two scatter plots are provided, and students are asked to analyze their bivariate behavior and then suggest examples of variables that could follow these patterns. For the plot on the left (c), students must

propose two variables that could exhibit low relationship intensity, while for the plot on the right (d), the relationship should be functional.

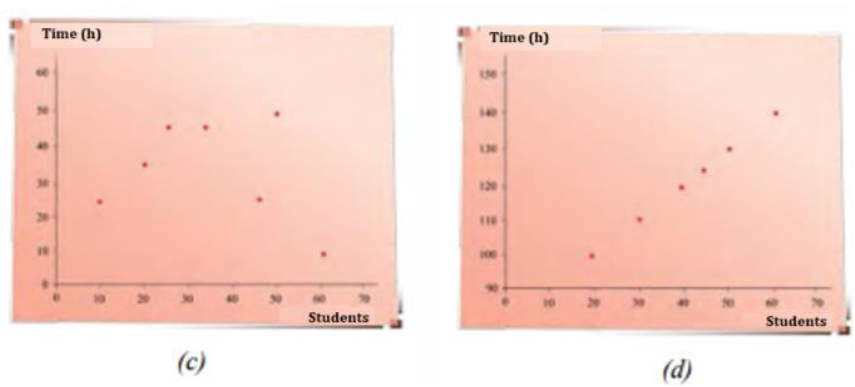
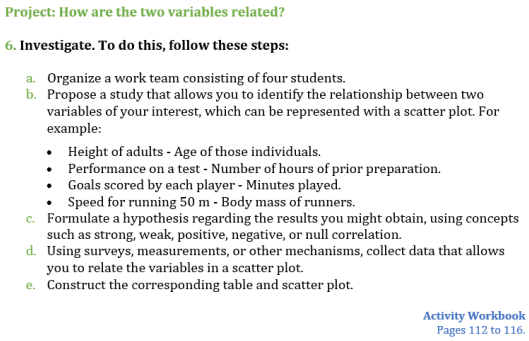


Figure 7: Situation with low relationship intensity (scatter plot on the left) and functional intensity (scatter plot on the right) (T4, p. 16)

The textbooks include some situations where calculating the correlation coefficient is not required or possible, but students are asked to describe the intensity of the relationship. For instance, in Figure 8, task b asks students to propose a study to investigate the relationship between two variables of their choice. In task c, they must provide information about the bivariate distribution: one case with high (strong) association and another with low (weak) association. In these cases, students do not calculate the correlation coefficient but must describe the relationship qualitatively.



Project: How are the two variables related?

6. **Investigate. To do this, follow these steps:**

a. Organize a work team consisting of four students.
b. Propose a study that allows you to identify the relationship between two variables of your interest, which can be represented with a scatter plot. For example:
   - Height of adults - Age of those individuals.
   - Performance on a test - Number of hours of prior preparation.
   - Goals scored by each player - Minutes played.
   - Speed for running 50 m - Body mass of runners.
c. Formulate a hypothesis regarding the results you might obtain, using concepts such as strong, weak, positive, negative, or null correlation.
d. Using surveys, measurements, or other mechanisms, collect data that allows you to relate the variables in a scatter plot.
e. Construct the corresponding table and scatter plot.

Activity Workbook
Pages 112 to 116.

Figure 8. Situation with intensity of relationship verbal description (T1, p. 153)

Table 3 summarizes the distribution of situations classified by the type of relationship intensity involved. All categories of relationship intensity are present in the four analyzed textbooks, except for independence, which does not appear in the Venezuelan textbook. The most common intensities are high (32.5%) and medium (19.7%). The least frequent are functional (3.9%) and independence (6.9%). These findings align with results reported by Gea et al. (2017) in Spanish textbooks.

Table 3: Frequency (percentage) of situations by relationship intensity

Intensity of the relationship	**Chile (T1)**	**Colombia (T2)**	**Ecuador (T3)**	**Venezuela (T4)**	**Total**
Independence	10 (12,0)	2 (3,6)	2 (5,6)	0 (0,0)	14 (6,9)
Low [0.1, 0.5)	21 (25,3)	10 (18,2)	3 (8,3)	2 (6,9)	36 (17,7)
Medium [0.5, 0.8)	21 (25,3)	2 (3,6)	3 (8,3)	14 (48,3)	40 (19,7)
High [0,8; 1)	18 (21,7)	33 (60,0)	14 (38,9)	1 (3,4)	66 (32,5)
Functional	1 (1,2)	3 (5,5)	3 (8,3)	1 (3,4)	8 (3,9)
Verbal description	12 (14,5)	5 (9,1)	11 (30,6)	11 (37,9)	39 (19,3)
Total	83 (100)	55 (100)	36 (100)	29 (100)	203 (100)

The Chilean textbook shows the most balanced inclusion of all types of relationship intensity, with percentages above 12% for each, except for functional, which has only one situation (1 out of 83). Low and medium intensities are the most frequent, accounting for over 50% (42 out of 83) of the situations combined. The Colombian textbook emphasizes high intensity situations (33 out of 55), while medium intensity and independence are the least frequent, with only two instances each. Similarly, the Ecuadorian textbook includes all types of intensity but focuses heavily on high intensity and verbal descriptions, which together account for nearly 70% of the situations (25 out of 36). Like the Colombian textbook, it has the fewest situations with independence. The Venezuelan textbook concentrates on medium intensity and verbal descriptions, which make up over 86% of its situations (25 out of 29), making it the least diverse in terms of relationship intensity.

Noteworthily, unlike other textbooks, the Chilean textbook includes an independence scenario with specific variables (Figure 3), which can help students understand the concept of no linear relationship. In contrast, the other textbooks present independence only through scatter plots, without identifying the variables. While these plots can convey the idea of independence, they may not necessarily foster a deeper understanding of the concept.

The bias toward high intensity relationships is also reported in other studies (Estepa & Sánchez, 1998; Gea et al., 2017; Lavalle et al., 2006). Ideally, textbooks should present a complete range of relationship intensities between variables. This would help students avoid misconceptions, such as thinking that only one type of intensity is prevalent or always occurs between variables. Overall, the four textbooks meet this criterion, albeit with varying percentages. The Colombian textbook shows a significant imbalance favoring high intensity, while the Venezuelan textbook focuses on only two types of intensity, which could lead to misconceptions about variable relationships.

Direction of the relationship

In Figure 3, an example is provided where it is not possible to assign a direction to the relationship because no dependency exists between the variables studied. In such cases, the relationship direction is categorized as independent. On the other hand, the scatter plot on the left side of Figure 5 shows an example of a situation with medium relationship intensity ($0.5 \leq r < 0.8$) and a direct direction.

Figure 2 presents an example of an inverse direction. It examines the variables of a fever-reducing medication dose and the time it takes for body temperature to normalize. Students could logically observe the potential relationship between the variables: as the medication dose increases, the fever subsides more quickly, reducing the time needed for body temperature normalization.

The scatter plot on the right side of Figure 9 illustrates a non-linear relationship, identified as a curvilinear correlation. This type of direction was categorized under the other category.

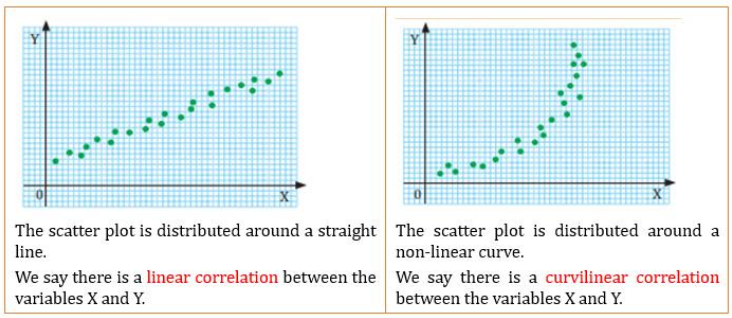


Figure 9. Situation with direct relationship direction (left diagram) and with non-linear relationship (right diagram) (T3, p. 203)

In Figure 8, the third task of the activity asks students to provide information about the bivariate distribution, offering one example with high intensity (strong) and another with low intensity

(weak). Additionally, they must specify whether the relationship direction is direct (positive), inverse (negative), or null (independent). Although they are not required to calculate the magnitude of the correlation coefficient or determine its exact direction, they must describe it qualitatively, which is categorized as verbal description.

Table 4 shows that direct direction is the most frequent relationship direction across situations in the four textbooks, followed by inverse or negative direction. The least common category is other, which includes non-linear relationships. Studies by Estepa and Sánchez (1998), Lavalle et al. (2006), and Gea et al. (2017) similarly report that most tasks in the textbooks they analyzed present relationships with a positive direction.

Table 4. Frequency (percentage) of situations by relationship direction

Direction of the relationship	Chile (T1)	Colombia (T2)	Ecuador (T3)	Venezuela (T4)	Total
Independence	9 (10,8)	2 (3,6)	2 (5,6)	0 (0,0)	13 (6,4)
Direct	36 (43,4)	37 (67,3)	15 (41,7)	16 (55,2)	104 (51,2)
Inverse	25 (30,1)	11 (20,0)	5 (13,9)	2 (6,9)	43 (21,2)
Other	1 (1,2)	2 (3,6)	3 (8,3)	0 (0,0))	6 (3,0))
Verbal description	12 (14,5)	3 (5,5)	11 (30,6)	11 (37,9)	37 (18,2)
Total	83 (100)	55 (100)	36 (100)	29 (100)	203 (100)

Once again, the Chilean textbook exhibits the best balance regarding the direction of relationships between variables in the situations, although direct and inverse directions together account for just over 73% (61 out of 83), with a 13% preference for the former. The Colombian textbook also favors direct and inverse directions, which account for 87% of its situations (48 out of 55), with direct relationships being more than triple the number of inverse ones. In the Ecuadorian and Venezuelan cases, most situations describe either a direct relationship or include verbal descriptions. However, this tendency is more pronounced in the Venezuelan textbook, where they account for 93.2% (27 out of 29) compared to 72.3% (26 out of 36) in the Ecuadorian textbook. In these textbooks, the number of situations with a direct direction is significantly higher than those with an inverse direction. Specifically in Venezuela, the two situations with an inverse direction are scatter plots designed to illustrate relationship types. Additionally, it is the only textbook that does not include situations with independent or non-linear relationships. Including situations with non-linear relationships in textbooks could help students better understand linear relationships by contrast.

As with the intensity of the relationship between variables, it is appropriate for textbooks to present various types of relationship directions. In this regard, all textbooks except the Venezuelan one fulfill this requirement, albeit with varying proportions across categories. In the Colombian textbook, most situations involve high intensity and direct direction. In contrast, the Venezuelan textbook primarily includes situations with medium intensity and positive direction. The Chilean textbook offers the best balance between intensity and direction of relationships between variables.

CONCLUSION

This study aimed to compare the situations presented in the topic of correlation and regression in mathematics textbooks for secondary education in Colombia, Chile, Ecuador, and Venezuela, focusing on the problem field, the intensity, and the direction of the relationship. The analysis of this topic in the reviewed textbooks reveals that they place the greatest emphasis on situations focused on analyzing the existence of a relationship between variables (P1), with modest coverage of data organization and reduction (P0), and limited emphasis on estimating one variable based on another (P2). Regarding the predominance of activities from the P1 field, these results align with those found by Gea et al. (2013, 2017), but the order of presence percentages for P0 and P2 is reversed.

The Colombian textbook demonstrates the best balance among the three problem fields (P0, P1, and P2), although it neglects P2-type activities. These differences could be influenced by the country's curriculum. In general, it is recommended to include more P0-type situations to facilitate understanding of P1 and P2 fields.

The analyzed textbooks share a bias toward activities with high or medium and direct relationships, aligning with findings from other studies (e.g., Gea et al., 2017; Rodríguez-Muñiz et al., 2020). This inclination toward these types of relationships can create challenges in students' learning, particularly potential semiotic conflicts (Salcedo et al., 2018), and limit learning opportunities (Hadar, 2017). While the preference for P1-field situations may stem from the curriculum, the emphasis on high or medium and direct relationships seems to be a conscious or unconscious choice by the textbook authors. In this regard, the textbooks from Chile, Colombia, and Ecuador exhibit similar distributions but maintain the bias toward high and direct relationships. The Venezuelan textbook, however, requires more substantial changes due to significant shortcomings.

The inclusion of correlation and regression in secondary education in Chile, Colombia, Ecuador, and Venezuela is an important step toward enhancing students' statistical education in these countries. However, it is recommended to provide greater variety and balance in the types of problem fields, as well as in the intensity and direction of relationships. Bivariate distributions are a fundamental statistical concept useful for exploring the nature of relationships between statistical variables (both qualitative and quantitative), critical for decision-making and understanding that correlation does not imply causation (Batanero, 2001; Crocker, 1981; Moritz, 2004). Furthermore, regression helps introduce modeling through statistical association (Engel et al., 2022).

A limitation of this study is that it used a non-probabilistic sample of textbooks, meaning the reported results can only be applied to the analyzed books. However, the study can be considered a contribution to the analysis of correlation and regression in textbooks, as this is a topic that has been scarcely addressed in research, particularly in relation to comparisons across different Latin American countries.

AUTHOR'S CONTRIBUTIONS

AS led the process of planning, coordination and execution of the study, participated in the delimitation of the problem to be addressed, data analysis and writing of the manuscript.

JGG participated in the data collection process and in the search and retrieval of the literature according to the study problem and the methodology employed, data analysis and writing of the manuscript.

DD participated in the data collection process and in the search and retrieval of the literature according to the study problem and the methodology employed, data analysis and writing of the manuscript.

YDP participated in the development of the methodology, data analysis and critical revision of the manuscript.

REFERENCES

Batanero, C. (2001). *Didáctica de la estadística*. Universidad de Granada.

Batanero, C., & Borovcnik, M. (2016). *Statistics and probability in high school*. Sense Publishers.

Batanero, C., Díaz, C., Contreras, J. M., & Roa, R. (2013). El sentido estadístico y su desarrollo. *Números*, *83*, 7-18.

Batanero, C., Gea, M.M., Díaz-Levicoy, D., & Cañadas, G.R. (2015). Objetos matemáticos ligados a la regresión en los textos españoles de bachillerato. *Educación Matemática*, *27*(2), 9-35.

Batanero, C., Gea, M.M., López-Martín, M.M., & Arteaga, P. (2017). Análisis de los conceptos asociados a la correlación y regresión en los textos de bachillerato. *Didacticae*, *1*, 60-76. https://doi.org/10.1344/did.2017.1.60-76

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. En 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

Conejero, S., Moraga, K., García-García, J.I., & Arredondo, E.H. (2023). Comprensión gráfica de estudiantes chilenos de educación media: un estudio exploratorio. *Investigación e Innovación en Matemática Educativa*, *8*, 1-18. https://doi.org/10.46618/iime.201

Cordero, F., & Flores, R. (2007). El uso de las gráficas en el discurso matemático escolar. Un estudio socioepistemológico en el nivel básico a través de los libros de texto. *Revista Latinoamericana de Matemática Educativa, 10*(1), 7-38.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin, 90*(2), 272-292. https://doi.org/10.1037/0033-2909.90.2.272

Díaz-Levicoy, D., Giacomone, B., & Arteaga, P. (2017). Caracterización de los gráficos estadísticos en libros de texto argentinos del segundo ciclo de Educación Primaria. *Profesorado. Revista de Currículum y Formación del Profesorado, 21*(3), 299-326.

Díaz-Levicoy, D., Vásquez, C., & Molina-Portillo, E. (2018). Estudio exploratorio sobre tablas estadísticas en libros de texto de tercer año de Educación Primaria. *Tangram. Revista de Educação Matemática, 1*(2), 18-39. https://doi.org/10.30612/tangram.v1i2.7574

Engel, J. (2019). Statistical literacy and society. En J.M. Contreras, M.M. Gea, M.M. López-Martín, & E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística* (pp. 1-17). FQM126.

Engel, J., & Sedlmeier, P. (2011). Correlation and regression in the training of teachers. En C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics challenges for teaching and teacher education: A Joint ICMI/IASE study* (pp. 247-258). Springer. https://doi.org/10.1007/978-94-007-1131-0_25

Engel, J., Schiller, A., & Martignon, L. (2022). Estadística cívica en la formación de profesores de matemáticas. En A. Salcedo y D. Díaz-Levicoy (Eds.), *Formación del Profesorado para Enseñar Estadística: Retos y Oportunidades* (pp. 21-46). CIEMAE.

Esquivel, E. C. (2016). La enseñanza de la Estadística y la Probabilidad, más allá deprocedimientos y técnicas. *Cuadernos de Investigación y Formación en Educación Matemática, 11*(15), 21-31.

Estepa, A., & Sánchez-Cobo, F.T. (1998). Correlation and regression in secondary school textbooks. En L. Pereira-Mendoza, L. Seu, T. Wee, & W. Wong (Eds.), *Proceedings of the Fifth Internacional Conference on Teaching of Statistics* (pp. 671-676). ISI.

Estepa, A., & Sánchez-Cobo, F.T. (2003). Evaluación de la comprensión de la correlación y regresión a partir de la resolución de problemas. *Statistics Education Research Journal*, *2*(1), 54-68. http://www.stat.auckland.ac.nz/~iase/serj/SERJ2(1).pdf

Fan, L., Zhu, Y., & Miao, Z. (2013). Textbook research in mathematics education: development status and directions. *ZDM Mathematics Education, 45*, 633-646. https://doi.org/10.1007/s11858-013-0539-x

Fergusson, A., & Pfannkuch, M. (2022). Introducing high school statistics teachers to predictive modelling and APIs using code-driven tools. *Statistics Education Research Journal, 21*(2), 1–25. https://doi.org/10.52041/serj.v21i2.49 Fraenkel, J.R., & Wallen, N.E. (2006). *How to design and evaluate research in education*. McGrawHill.

Gal, I. (2004) Statistical Literacy: meanings, components, responsibilities. En D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 47 -78). Springer. https://doi.org/10.1007/1-4020-2278-6_3

Gal, I., Grotlüschen, A., Tout, D., & Kaiser, G. (2020). Numeracy, adult education, and vulnerable adults: A critical view of a neglected field. *ZDM*, *52*(3), 377-394. https://doi.org/10.1007/s11858-020-01155-9

García-García, J.I. (2021). El contagio de los datos. La importancia de alfabetización estadística. En J. E. Sagula (Comp.), *II Simposio de Educación Matemática Virtual. Educación Matemática enriquecida por Interdisciplinariedad con la Tecnología. Tomo I: conferencias, paneles* (pp. 199-204). EdUnLu. https://doi.org/10.13140/RG.2.2.18297.34408

Garfield, J., & Ben-Zvi, D. (2008). *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice*. Springer. https://doi.org/10.1007/978-1-4020-8383-9

Gea, M.M., Batanero, C., Arteaga, P., & Contreras, J.M. (2017). Variables characterizing correlation and regression problems in the Spanish high school textbooks. En G. Gueuedet, & T. Dooley (Eds.), *Proceedings 10th Congress of the European Society for Research in Mathematics Education* (pp. 764-771). CERME.

Gea, M.M., Batanero, C., Cañadas, G., & Contreras, J.M. (2013). *Un estudio empírico de las situaciones-problema de correlación y regresión en libros de texto de bachillerato.* En A. Berciano, G. Gutiérrez, A. Estepa, & N, Climent (Eds.), *Investigación en Educación Matemática XXI* (pp. 283-291). SEIEM.

Gea, M.M., López-Martín, M.M., & Roa, R. (2015). Conflictos semióticos sobre la correlación y regresión en los libros de texto de Bachillerato. *Avances de Investigación en Educación Matemática, 8*, 29-49. https://doi.org/10.35763/aiem.v1i8.113

Hadar, L.L. (2017). Opportunities to learn: Mathematics textbooks and students' achievements. *Studies in Educational Evaluation*, *55*, 153-166. https://doi.org/10.1016/j.stueduc.2017.10.002

Herbel, B.A. (2007). From intended curriculum to written curriculum: Examining the «voice» of a mathematics textbook. *Journal for Research in Mathematics Education, 38*(4), 344-369

Hernández, R., Fernández, C., & Baptista, M.P. (2014). *Metodología de la Investigación*. McGraw Hill.

Lavalle, A.L., Micheli, E.B., & Rubio, N. (2006). Análisis didáctico de regresión y correlación para la enseñanza media. *Revista Latinoamericana de Investigación en Matemática Educativa*, *9*(3), 383-406.

Olsher, S., & Even, R. (2014). Teachers editing textbooks: Changes suggested by teachers to the math textbook they use in class. En K. Jones, C. Bokhove, G. Howson & L. Fan (Eds.), *Proceedings of the international conference on mathematics textbook research and development* (pp. 43-48). University of Southampton.

Qi, C., Zhang, X., & Huang, D. (2018). Textbook use by teachers in junior high school in relation to their role. En L. Fan, L. Trouche, C. Qi, S. Rezat, & J. Visnovska (Eds.), *Research on Mathematics Textbooks and Teachers' Resources, ICME-13 Monographs* (pp. 29-51). Springer. https://doi.org/10.1007/978-3-319-73253-4_15

Rezat, S., Fan, L. & Pepin, B. (2021). Mathematics textbooks and curriculum resources as instruments for change. *ZDM,* 53, 1189–1206. https://doi.org/10.1007/s11858-021-01309-3

Rodríguez Muñiz, L.J., & Díaz, P. (2018). Las investigaciones sobre la estadística y la probabilidad en los libros de texto de Bachillerato. ¿Qué se ha hecho y qué se puede hacer? *AIEM*, 14, 65–81. https://doi.org/10.35763/aiem.v0i14.218

Rodríguez-Muñiz, L.J., Corte, Á., & Muñiz-Rodríguez, L. (2020). ¿Evolucionan los libros de texto de matemáticas con los cambios curriculares? Estudio de la regresión y la correlación lineal en la Educación Secundaria en España. *Números*, *103*, 65-79.

Salcedo, A., González, J., & González, J. (2021). Lectura e interpretación de gráficos estadísticos, ¿cómo lo hace el ciudadano? *PARADIGMA*, *41*(e1), 61-88. https://doi.org/10.37618/PARADIGMA.1011-2251.2021.p61-88.id1018

Salcedo, A., Molina-Portillo, E., Ramírez, T., & Contreras, J. M. (2018). Conflictos semióticos sobre estadística en libros de texto de matemáticas de primaria y bachillerato. *Revista de Pedagogía*, *39*(104), 223-244.

Stempler, S. (2001). An overview of content analysis. *Practical Assessment, Research & Evaluation, 7*(17), 1-6. http://doi.org/10.7275/z6fm-2e34

Stylianides, G.J. (2009). Reasoning-and-proving in school mathematics textbooks. *Mathematical Thinking and Learning*, *11*(4), 258-288. https://doi.org/10.1080/10986060903253954

Valverde, G., Bianchi, L., Wolfe, R., Schmidt, W., & Houang, R. (2002). *According to the book: using TIMSS to investigate the translation of policy into practice through the world of textbooks*. Springer. https://doi.org/10.1007/978-94-007-0844-0

Watson J., Fitzallen N., Fielding-Wells, J., & Madden, S. (2018). The practice of statistics. En D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 105-137). Springer. https://doi.org/10.1007/978-3-319-66195-7_4

Wijaya, A., van den Heuvel-Panhuizen, M., & Doorman, M. (2015). Opportunity-to-learn context-based tasks provided by mathematics textbooks. *Educational Studies in Mathematics*, *89*(1), 41-65. https://doi.org/10.1007/s10649-015-9595-1