International Electronic Journal of Mathematics Education
2023, 18(4), em0748
e-ISSN: 1306-3030
https://www.iejme.com

MODESTUM

Research Article

OPEN ACCESS

The conceptual field of measures of central tendency: A first approximation

Irene Mauricio Cazorla [1]*, Miriam Cardoso Utsumi [2], Sandra Maria Magina [1]

[1] Universidade Estadual de Santa Cruz, Ilhéus, BA, BRAZIL
[2] Universidade Estadual de Campinas, Campinas, SP, BRAZIL
***Corresponding Author:** icazorla@uesc.br

Citation: Cazorla, I. M., Utsumi, M. C., & Magina, S. M. (2023). The conceptual field of measures of central tendency: A first approximation. *International Electronic Journal of Mathematics Education, 18*(4), em0748. https://doi.org/10.29333/iejme/13571

ARTICLE INFO

Received: 04 May 2023

Accepted: 03 Aug. 2023

ABSTRACT

This article aims to present a first approximation of the conceptual field of measures of central tendency (MCT), grounded in the theory of conceptual fields. We propose six situations according to type of variable, data presentation (raw or grouped) and amount of data. We revisit specific situations for the mean and exemplify several possibilities for introducing these measures. This is a theoretical reflection and a systematization of the knowledge produced in the area with a view to aiding it teaching in Brazilian basic education (PreK-12). We argue that from the early school years, students should collect data, use concrete and manipulable material, analyze the types of variable and context of data, transform them into different representations and based on the visual observation of data distribution, intuit the location of MCT to then calculate them.

Keywords: measures of central tendency, theory of conceptual fields, basic education, statistics education, teaching of statistics

INTRODUCTION

The teaching of statistics in Brazilian basic education (PreK-12) was made official in 1997 with the issuance of the Brazilian national curriculum standards (Ministério da Educação [Ministry of Education] [MEC], 1997) and ratified by the Brazilian national common core curriculum (BNCCC) (MEC, 2018) with its inclusion in one of the five main areas of mathematics, called probability and statistics.

In this context, the academic community has focused on research involving the teaching and learning of the main concepts of statistics to be taught since the beginning of elementary school (K-5). Likewise, textbooks have advanced in the introduction of those concepts and there is evidence of their teaching in schools.

Consequently, the measures of central tendency (MCT)–mode (*Mo*), median (*Md*), and mean (*M*)–have receive a lot of attention, as they are fundamental concepts of statistics that provide a foundation for more complex statistical concepts. However, we still find students and teachers facing difficulties to grasp and use these measures, as we will see in the review below.

In our view, these difficulties stem from the nature of statistics. Although its concepts and procedures use basic mathematical tools (arithmetic operations), grasping them, especially MCT, requires understanding other concepts, such as the different types of variables, data configuration and the purpose of collecting data. In addition, it requires having an overall idea of the peculiarities of MCT and the different situations in which they can be calculated in order to identify which one is more appropriate in a specific situation.

We note that in a decision-making situation we may be challenged to choose one of MCT, and such a choice can involve, besides mathematical properties, ethical and idiosyncratic aspects that may vary from person to person. In this sense, it is important to have a comprehensive view of MCT to understand the different situations that generate them, the concepts involved, and their properties and representations. Several researchers have therefore carried out studies with different theoretical frameworks, among whom we highlight the group from Granada (Spain), who use the onto-semiotic approach (Batanero, 2000; Cobo, 2003; Cobo & Batanero, 2000; Mayén et al., 2017); Andrade (2013), who used the anthropological theory of the didactic and the theory of conceptual fields (TCF); and the Brazilian group, who use TCF (Cazorla et al., 2019, 2021b, 2021c; Magina et al., 2022).

Based on what has already been investigated, the goal of this article is to briefly present a first approximation of the conceptual field of MCT, aimed at their teaching in Brazilian basic education, grounded in TCF. Thus, we propose six situations according to type of variable, data presentation (raw or grouped) and amount of data, revisit specific situations for the mean that do not apply to the median and mode and exemplify several possibilities for introducing these measures. This is a theoretical reflection and a systematization of the knowledge produced in the area with a view to aiding its teaching in Brazilian basic education.

This work consists of five sections, including this introduction. In the second section we present TCF; the third section features a brief review of the main studies that address MCT; in the fourth section we describe our proposal for the framework of the conceptual field for MCT; and the fifth section comprises our conclusions.

THEORY OF CONCEPTUAL FIELDS: A BRIEF REVIEW

TCF is a cognitive theory developed in the field of psychology within mathematics education, strongly influenced by genetic epistemology (Piaget, 1966) and social constructivism (Vygotsky, 2001), with various of their concepts applied in classroom. This is the case of the concepts of schema, proposed by Piaget and Inhelder (1995), and zone of proximal development (Vygotsky, 2001).

TCF assumes that while the content of mathematics to be taught is the same anywhere in the world, its teaching is not, since it takes place in a determined society within a given institution, and in a particular classroom with a specific teacher. This means that although the mathematical content has universal traits, it is taught in different ways and levels, as it is directly related to the culture and technological advancement of each society, the psychoeducational approach of the educational institution and also the background (with specific and pedagogical content) of the teacher. These factors lead to different goals and outcomes.

Based on this epistemological position, Vergnaud (1990, p. 3) raises the following questions:

> What is the nature and function of a new concept, a new procedure, a new kind of reasoning, a new representation? More precisely, what relationship exists between the new mathematical skills and conceptions and the practical and theoretical problems that give them value and meaning?

We agree with Vergnaud (1990) as to the importance of these types of questions as they help teachers choose the most appropriate situations with which to approach the various concepts with their students. For him, it is essential in the teaching process that teachers know how to choose the situations that will help their students develop a given concept, while identifying the different concepts present in each of the problem situations. In addition, they must observe the properties that are relevant to the resolution (invariants), thinking about possible symbolic representations that will allow students to establish relationships between the situations and their invariants.

This process is key to the cognitive analysis of the understanding of behaviors assumed by students during the resolution (Vergnaud, 1990). Furthermore, in several texts Vergnaud (1982, 1987, 1996) declares that the learning of any concept involves three elements: a set of situations (S) in which the concept to be learned is inserted; a set of invariants (I) present in this set of situations; and the set of symbolic representations (R) that express the invariants of situations. This triad can be described, as follows: C=(S, I, R). Regarding this triad, Vergnaud (1987, 1996, 1998) explains that the situations (S) give **meaning** to the concept. Thus, the more situations the learner interacts with, the broader the meaning of the content will be. The invariants (I) are the (mathematical) **properties** that define the content and the **procedures** adopted by the student to solve the situations. Finally, the symbolic representations (R) allow students to express themselves about the concept, relating the **meaning** to the **properties** of the content they are learning.

In this sense, Vergnaud (1987) borrows the terms “referent,” “signified” and “signifier” from linguistics and uses them from a psychological perspective. Thus, the situation (S) is the “referent,” which is linked to the real world from which emerge people’s experiences, that is, it relates to the task, the activity and the problem in which the mathematical object (content) is present. The “signified” relates to the set of invariants (I), which, in turn, consists of the properties that belong to the object and also of the procedures that the student uses to deal with the situation. It is understood that both properties and procedures may be implicit for the student who is using them. In this case, it would be what Vergnaud (1996, 1998) calls “theorem-in-action.” Finally, there is the symbolic representation (R), which Vergnaud (1987) considers arising, from an epistemological viewpoint, from the interaction between the set of situations (S–referent) and the set of invariants (I–signified). It is associated with the signifier of the concept.

Figure 1 was designed from the representation of the triangle proposed and featured in Magina et al. (2022), highlighting the relationships (S, I, R) present in the formation of the concept and associating them with elements of semiotics. Throughout this article we will refer to this triad to support the concepts in MCT, thus drawing on this theoretical framework to reflect on the concepts present in MCT.

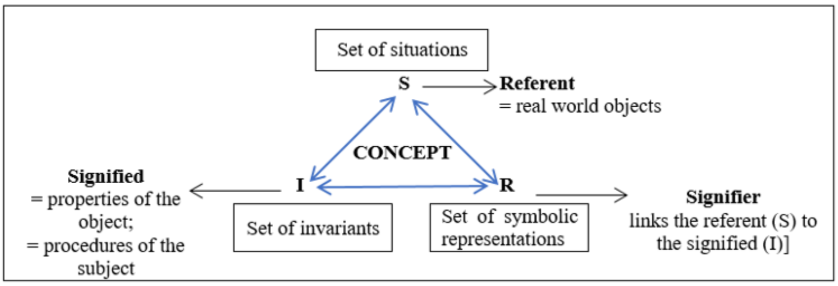


Figure 1. Concept formation triangle (Magina et al., 2022)

MEASURES OF CENTRAL TENDENCY

MCT–mode (*Mo*), median (*Md*), and mean (*M*)–are measures that summarize and represent a data set using only one or few number(s)/category(ies). They indicate the site, where the data tend to concentrate.

As Cobo and Batanero (2000) point out, MCT are summary measures, that is, they refer to the entire data set (collective) rather than the particular elements of the set, which implies a shift from the deterministic perspective to the statistic perspective:

> Saying that a (collective) data set has a tendency or referring to a summary measure implies that collective is a collection of ***identical*** elements that vary in relation to the variable in question. Such understanding also implies understanding the variability of the data in relation to the summary measure (Cobo & Batanero, 2000, p. 3, emphasis added).

MCT are apparently simple concepts since their calculation involves basic arithmetic operations that are developed and expanded over the school years. However, several researchers show that students face difficulties solving problems that involve MCT.

Batanero (2000) carried out an extensive review of research on the difficulties of students in understanding MCT, among which we highlight: using the simple mean rather than the weighted mean; not considering "weighting" in data presented in frequency distribution tables (FDT) at intervals; calculating the mean of frequencies; using the algorithm mechanically without understanding its meaning; not being able to find an unknown value based on knowledge of the other values and the mean; confusing mode with modal frequency; and not sorting data to find the median. For the author, the reason may be that students are used to a single calculation method and a single solution to mathematical problems.

Regarding the median, Cobo and Batanero (2000) undertook a comprehensive study explaining four definitions involving raw data, data grouped in an FDT, and data based on the cumulative frequency graph. The authors point out that the main difficulty in understanding the median is that there is no single calculation algorithm, as it depends on the type of data, the way they are presented and their amount.

Cobo (2003) expanded the two previous studies, presenting the meanings of MCT for high school. Subsequently, Mayén et al. (2017) investigated the resolution of nine MCT-related problems by Mexican students, comparing them with the resolution by Spanish students. All these studies were grounded in the onto-semiotic approach system.

In his doctoral thesis Andrade (2013) undertook a comparative study on the teaching of MCT in basic education in Brazil and France using the anthropological theory of the didactic and TCF. Figueiredo (2017) carried out a review of how the median is introduced and calculated in 7th and 8th grade textbooks and in the mathematics core standards in Portugal.

Using TCF, Cazorla et al. (2019) presented a first approximation of the conceptual field of the arithmetic mean, restricted to the empirical field, explaining the network of concepts involved and the triad (S, I, R), distinguishing three classes of mean: simple, aggregated and weighted. Later, Magina et al. (2022) presented the network of concepts and a first proposal for teaching MCT in the early years of elementary school. Cazorla et al. (2021b) presented three situations for teaching MCT with raw data and Cazorla et al. (2021c) presented a first approach to the conceptual field of the median.

Considering the results of those investigations, we present below our proposal for MCT.

FRAMEWORK OF MEASURES OF CENTRAL TENDENCY: AN ADVANCE IN THEIR APPROXIMATION

We begin by briefly presenting the definition of the three MTC. Firstly, however, we note that when we refer to type of variable (qualitative or quantitative) it is always post- operationalization, since, as Cazorla et al. (2021d) stress, qualitative variables can be operationalized quantitatively and vice versa.

Measures of Central Tendency

Mode (***Mo***) is the value (x_i) that appears most frequently (f_i) in a data set. If the variable is qualitative, the mode will be a category; if the variable is discrete and takes on few values (unique values), the mode will be a unique value; and if the variable is continuous or discrete and takes on many values, the mode may not exist or be meaningless. Finding it requires constructing the FDT in class intervals and, in this case, the mode will be "given by the maximum frequency density" (Andrade, 2013, p. 127), as we will detail later. We note that the mode may not exist (amodal distribution) or take on more than one value/category in the case of bimodal or multimodal distributions.

Median (***Md***) divides the amount of data (*n*) into two equal parts, with half of the data taking on values lower than or equal to it, and the other half comprising values greater than or equal to it. If the data are raw, they must be sorted. Its calculation depends on whether the amount of data elements (*n*) is odd or even. If *n* is odd, then the central position will be (n+1)/2 and the median will take on the value that occupies that position: $Md=X_{((n+1)/2)}$; if *n* is even, then the central position will be between the position (n/2) and the subsequent position (n/2)+1, and the median will be the mean of those two values: $Md=(X_{(n/2)}+X_{(n/2)+1})/2$. In the case of grouped data, the cumulative frequency (F_i) must be calculated, which will depend on whether the FDT comprises unique values or class intervals, as we will see in the situations.

Mean (***M***) is defined as ratio between sum of variable values and number of data elements, whose algorithm depends on kind of data, which can be simple, aggregated or grouped, as described by Cazorla et al. (2019), and which we will present in the situations.

Given these definitions, we expand below the triad (S, I, R) to MCT, which so far merely involved the raw data proposed by Cazorla et al. (2021b), to include the situations for grouped data.

Symbolic Representations (R) Present in MCT

In the case of quantitative variables, MCT can be represented in their numerical, verbal, iconic (the triangle has been used as the "scale tipper" to represent the mean), algebraic or graphic (one line, when the variable is represented in a bar or line graph) format, as presented by Cazorla et al. (2019). In the case of qualitative variables, MCT can be represented in their verbal or iconic format, related to the most frequent category or median category.

We note that data representation must be distinguished from MCT representation. In the case of a variable (univariate analysis) whose data are represented in graphs, the variable will generally be on the abscissa, where the MCT will be, and so we can use arrows or triangles that touch the number line or categories to indicate, where they meet. When the variable is indexed by time (bivariate analysis), time will generally be represented on the abscissa and the variable on the ordinate; therefore, MCT will be on the ordinate, as seen in **Figure 1**.

However, it is not uncommon to have the variable on the ordinate. In this sense, Arteaga (2011, p. 162) draws attention to the proper choice of data representation and how it can help in the visual analysis of where MCT are located. The author analyzed the representation of experimental data of 45 students who flipped a coin 20 times. One student used the bar/rod graph, recording the students on the abscissa and the variable values on the ordinate (part a in **Figure 2**). This means of representation is arbitrary, as the order of students on the abscissa has no sense or meaning. It makes it difficult to visually locate MCT. In this type of representation, in order to locate the mean the student will have to imagine a straight line between the tops of the bars/rods, in such a way that the segments above the line and the segments missing to reach the line cancel each other out. For the median the student would have to mentally order the bars and check the height of the one that occupies position 23, which is complex. To find the mode he or she would have to visually examine the greatest number of bars with the same height.

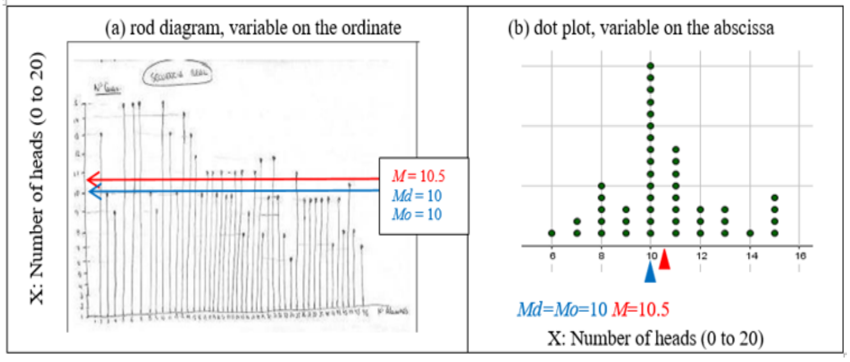


Figure 2. Representation of raw data & MTC (Source: Authors'own elaboration)

The adequate representation for this type of data is the dot plot (part b in **Figure 2**), where we can visualize that the mode is 10. For the median we can use the value of 10 as a reference point, observing that there are 11 data elements below it and 19 above it, hence the position 23 will be at the value of 10. In this case, we can intuit that the mean will be a little higher than 10 due to the presence of four values of 15.

Thus, we can see that the representation with the bar/rod graph (part a in **Figure 2**) is not suitable for this purpose. The only advantage of this type of representation for the mean is that one of its properties would be activated, since the deviations of the variable values in relation to the mean cancel each other out, resulting in zero and helping its understanding.

In the specific case of the mean and seasonal time series (weekly, monthly, annual, etc.), the moving average is used to mitigate the impact of seasonal variability and is represented by a line graph (**Figure 3**). This type of representation has become very popular in disclosing COVID-19 data, as we will detail later on.

Fonte: Consórcio G1, O Globo, Extra, O Estado de S. Paulo, Folha de S. Paulo e UOL com dados das secretarias estaduais de Saúde

Figure 3. Representation of moving average by a line graph (Rodrigues et al., 2021, p. 197)

In the algebraic representation of MCT, we use a notation system whose symbols are explained below:

X: generic designation of the variable being studied,

*x*i: designates the i-th value/category of the variable. In a class interval it denotes the class mark or midpoint,

n: sample size, amount of data,

*x*min: minimum value of the quantitative variable (in the case of the ordinal qualitative variable: lower category),

*x*max: maximum value of the variable (in the case of the ordinal qualitative variable: higher category),

AT: total amplitude *AT*=xmax- xmin,

*P*m: midpoint, the value that divides AT into two equal parts, *Pm*=*x*min +AT/2,

k: number of categories (ordinal qualitative), of unique values (discrete that take on few values) or of classes (continuous or discrete that take on many values),

*f*i: absolute frequency, number of occurrences of the variable in the i-th category, in the i-th value or in the i-th class; such that: $n = \sum_{i=1}^{k} f_i$,

*F*j: cumulative absolute frequency, number of occurrences of the variable up to the i-th category, value or class; such that: $F_j = \sum_{i=1}^{j} f_i$,

Md: Median,

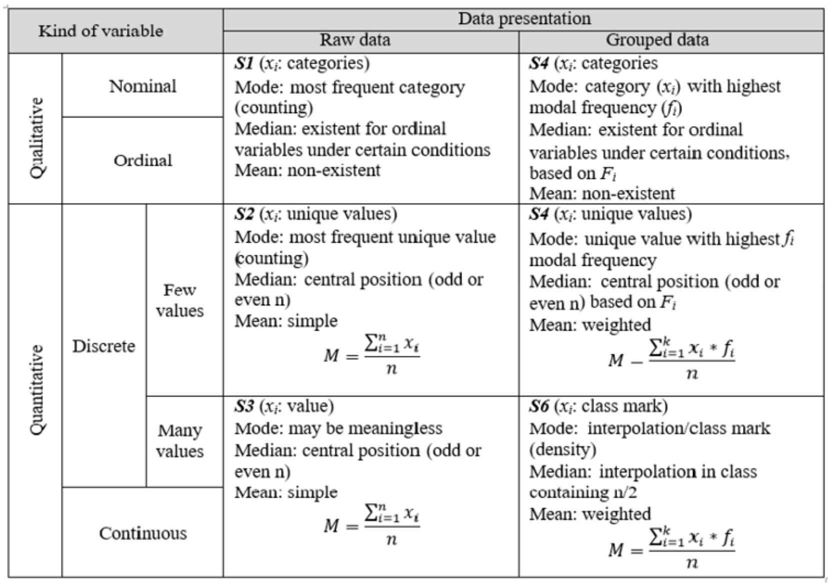
[*L*i; *L*i+1[: designates the i-th class interval (or simply class), half-open (closed on the left and open on the right. Includes all xi values: such that $L_i \leq x_i < L_{i+1}$), and

*L*i: designates the lower limit, and *L*i+1 designates the upper limit of the i-th class interval; *a*: class interval amplitude, such that: *L*i+1:= *L*i+*a*.

Situations (S) Present in MCT

The algorithm to find MCT depends on the type of variable, the amount of data and how they are arranged, whether in their original form (raw data) or grouped. In other words, the situations depend on the data format, whether raw or organized in FDT or graphs that allow access to absolute or relative frequency.

Cazorla et al. (2021b, p. 9) proposed three situations for MCT related to raw data: ***S1*** for qualitative variables, ***S2*** for discrete variables that take on few values and ***S3*** for continuous and discrete variables that take on many values, featuring in the middle column of **Figure 4**. Expanding this proposal to grouped data, we present three additional situations, namely: for qualitative variables (***S4***), for discrete variables that take on few values (***S5***) and for continuous and discrete variables that take on many values (***S6***). In addition to these, we revisited the specific situations for the mean proposed by Cazorla et al. (2019).

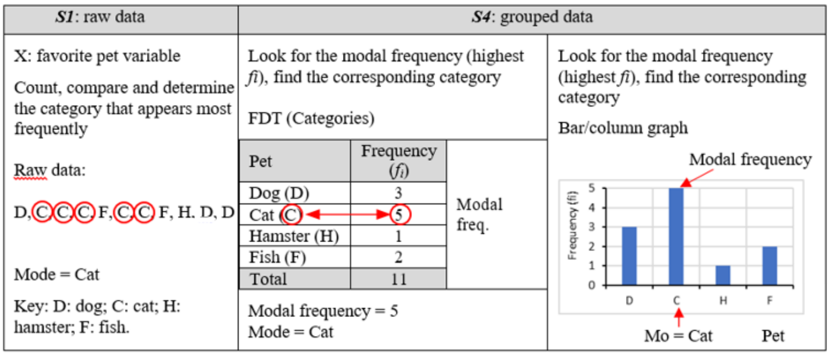


Kind of variable			Data presentation	
			Raw data	Grouped data
Qualitative	Nominal		***S1*** (x_i: categories) Mode: most frequent category (counting) Median: existent for ordinal variables under certain conditions Mean: non-existent	***S4*** (x_i: categories Mode: category (x_i) with highest modal frequency (f_i) Median: existent for ordinal variables under certain conditions, based on F_i Mean: non-existent
	Ordinal			
Quantitative	Discrete	Few values	***S2*** (x_i: unique values) Mode: most frequent unique value (counting) Median: central position (odd or even n) Mean: simple $M = \frac{\sum_{i=1}^{n} x_i}{n}$	***S4*** (x_i: unique values) Mode: unique value with highest f_i modal frequency Median: central position (odd or even n) based on F_i Mean: weighted $M - \frac{\sum_{i=1}^{k} x_i * f_i}{n}$
		Many values	***S3*** (x_i: value) Mode: may be meaningless Median: central position (odd or even n) Mean: simple $M = \frac{\sum_{i=1}^{n} x_i}{n}$	***S6*** (x_i: class mark) Mode: interpolation/class mark (density) Median: interpolation in class containing n/2 Mean: weighted $M = \frac{\sum_{i=1}^{k} x_i * f_i}{n}$
	Continuous			

Figure 4. Situations involving MCT (Source: Authors'own elaboration)

Below we present the situations according to the type of variable.

Situations 1** and **4 (***S1*** *and* ***S4***). In these two situations we included all nominal (gender, favorite pet, ...), ordinal (food taste, social class, ...) and sequential (month of birth, weeks, ...) qualitative variables. All these variables have a mode. If the data are raw (**S1**), then it will be necessary to count how many times each category is repeated and find the category that was repeated the most. If the data are grouped in the FDT (***S4***), just search the frequency column for the modal frequency (f_i), which is the one with the highest value, and the mode will be the corresponding category. If the data are in a bar/column, circular or pictogram graph, just look for the largest bar/column or sector and the mode will be the corresponding category, as shown in **Figure 5**.



S1: raw data

X: favorite pet variable

Count, compare and determine the category that appears most frequently

Raw data:

D, C, C, C, F, C, C F, H, D, D

Mode = Cat

Key: D: dog; C: cat; H: hamster; F: fish.

S4: grouped data

Look for the modal frequency (highest *fi*), find the corresponding category

FDT (Categories)

Pet	Frequency (f_i)
Dog (D)	3
Cat (C)	5
Hamster (H)	1
Fish (F)	2
Total	11

Modal freq.

Modal frequency = 5
Mode = Cat

Look for the modal frequency (highest *fi*), find the corresponding category

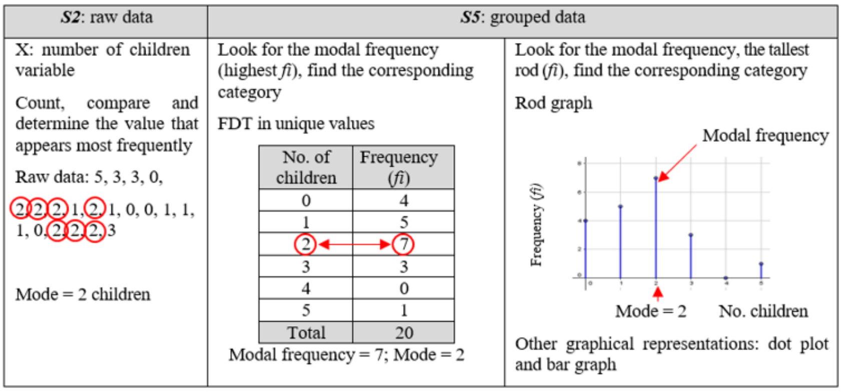
Bar/column graph

Figure 5. Situations involving mode for qualitative variables (***S1*** & ***S4***) (Source: Authors'own elaboration)

For some authors there is a median for ordinal variables. However, if the number of data elements is even and the positions (n/2) and ((n/2)+1) are contained in different categories, the median is undetermined, as shown by Cazorla et al. (2021c). There is no mean for qualitative variables.

Situations 2 and 5 (***S2*** and ***S5***) involve discrete variables that take on few values, such as the number of children per employee, and have the three MCT. In basic education they have been treated and represented like an ordinal variable, using FDT in unique values and bar/column graphs. Here we will use data from the example by Morettin and Bussab (2010, p. 16) related to the number of children of 20 employees in a factory.

To find the mode with raw data (***S2***) we use the same procedure used with qualitative variables, that is, we count how many times each value is repeated and find the most frequent value (**Figure 6**). If the data are grouped in FDT (***S5***), just search the frequency column for the modal frequency (f_i), which is the one with the highest value, and the mode will be the corresponding value (x_i). If the data are in a bar/column graph, rod graph or dot plot, just look for the largest bar/column/rod or number of dots and the mode will be the corresponding value.



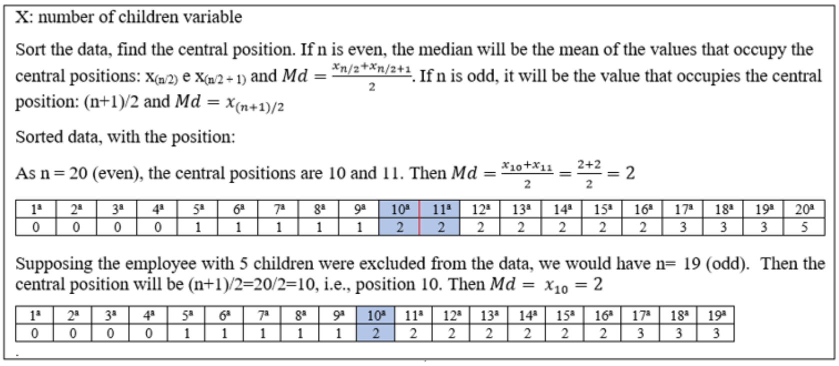
S2: raw data	***S5***: grouped data	
X: number of children variable Count, compare and determine the value that appears most frequently Raw data: 5, 3, 3, 0, 2, 2, 2, 1, 2, 1, 0, 0, 1, 1, 1, 0, 2, 2, 2, 3 Mode = 2 children	Look for the modal frequency (highest *fi*), find the corresponding category FDT in unique values Modal frequency = 7; Mode = 2	Look for the modal frequency, the tallest rod (*fi*), find the corresponding category Rod graph Other graphical representations: dot plot and bar graph

No. of children	Frequency (*fi*)
0	4
1	5
2	7
3	3
4	0
5	1
Total	20

Figure 6. Situations involving mode for discrete variables that take on few values (***S2*** & ***S5***) (Source: Authors'own elaboration)

We note that the proper representation for this type of variable is the dot plot or rod graph. Unfortunately, these two kinds of graph were not included in BNCCC, and in school and textbooks the bar graph is used, which is not suitable, as the bar width exceeds the unique value, which can be misleading. Furthermore, if we want to represent the mean with the triangle, it will be much more difficult to locate it on the number line.

To find the median, we draw on the example presented by Cazorla et al. (2021b). In the case of raw data (***S2***), they must be sorted and it must be checked whether the number of data elements is odd or even, as in **Figure 7**.



X: number of children variable

Sort the data, find the central position. If n is even, the median will be the mean of the values that occupy the central positions: $x_{(n/2)}$ e $x_{(n/2+1)}$ and $Md = \frac{x_{n/2}+x_{n/2+1}}{2}$. If n is odd, it will be the value that occupies the central position: (n+1)/2 and $Md = x_{(n+1)/2}$

Sorted data, with the position:

As n = 20 (even), the central positions are 10 and 11. Then $Md = \frac{x_{10}+x_{11}}{2} = \frac{2+2}{2} = 2$

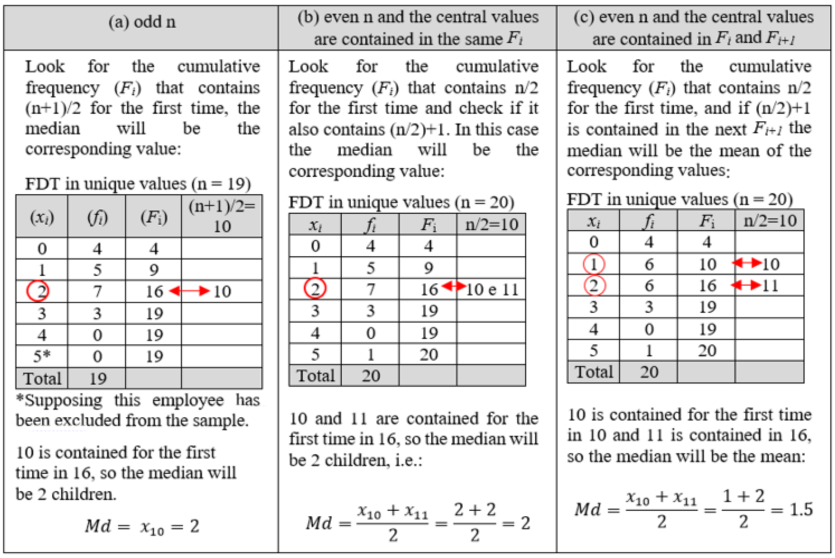
1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	9ª	10ª	11ª	12ª	13ª	14ª	15ª	16ª	17ª	18ª	19ª	20ª
0	0	0	0	1	1	1	1	1	2	2	2	2	2	2	2	3	3	3	5

Supposing the employee with 5 children were excluded from the data, we would have n= 19 (odd). Then the central position will be (n+1)/2=20/2=10, i.e., position 10. Then $Md = x_{10} = 2$

1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	9ª	10ª	11ª	12ª	13ª	14ª	15ª	16ª	17ª	18ª	19ª
0	0	0	0	1	1	1	1	1	2	2	2	2	2	2	2	3	3	3

Figure 7. Median for discrete variables that take on few values & raw data (***S2***) (Cazorla et al., 2021c, **Figure 5**, p. 10)

If the data are grouped in an FDT (***S5***), we must calculate the cumulative frequency (*Fᵢ*) and find the one that contains n/2 for the first time. If n is odd, the median is the corresponding value (part a in **Figure 8**). If n is even, we must examine whether n/2 and ((n/2)+1) are contained in the same *Fᵢ*, in which case the median will be the corresponding value (part b in **Figure 8**); but if those numbers are contained in consecutive *Fᵢ*, then the median will be the mean of the corresponding values (part c in **Figure 8**).



(a) odd n

Look for the cumulative frequency (F_i) that contains (n+1)/2 for the first time, the median will be the corresponding value:

FDT in unique values (n = 19)

(x_i)	(f_i)	(F_i)	(n+1)/2= 10
0	4	4	
1	5	9	
2	7	16	10
3	3	19	
4	0	19	
5*	0	19	
Total	19		

*Supposing this employee has been excluded from the sample.

10 is contained for the first time in 16, so the median will be 2 children.

$$Md = x_{10} = 2$$

(b) even n and the central values are contained in the same F_i

Look for the cumulative frequency (F_i) that contains n/2 for the first time and check if it also contains (n/2)+1. In this case the median will be the corresponding value:

FDT in unique values (n = 20)

x_i	f_i	F_i	n/2=10
0	4	4	
1	5	9	
2	7	16	10 e 11
3	3	19	
4	0	19	
5	1	20	
Total	20		

10 and 11 are contained for the first time in 16, so the median will be 2 children, i.e.:

$$Md = \frac{x_{10} + x_{11}}{2} = \frac{2+2}{2} = 2$$

(c) even n and the central values are contained in F_i and F_{i+1}

Look for the cumulative frequency (F_i) that contains n/2 for the first time, and if (n/2)+1 is contained in the next F_{i+1} the median will be the mean of the corresponding values:

FDT in unique values (n = 20)

x_i	f_i	F_i	n/2=10
0	4	4	
1	6	10	10
2	6	16	11
3	3	19	
4	0	19	
5	1	20	
Total	20		

10 is contained for the first time in 10 and 11 is contained in 16, so the median will be the mean:

$$Md = \frac{x_{10} + x_{11}}{2} = \frac{1+2}{2} = 1.5$$

Figure 8. Median for discrete variables that take on few values & data grouped in one FDT (***S5***) (Cazorla et al., 2021c, **Figure 6**, p. 11)

If the data are represented in graphs that allow access to the frequencies, just reconstruct FDT and follow the procedure described. In some cases, we can find the median directly from the graphs by using the same strategy used in FDT: calculate n/2 and count the data elements cumulatively until n/2 is found for the first time. Let's look at three cases. In the first case, when n is odd (n=19), then the median occupies position 10. If we are using standardized paper slips, each student pastes their slip on the board, in the corresponding place (part a in **Figure 9**), and then counts how many slips there are up to the first "bar" (4), up to the second (nine), then position 10 will be in the third "bar" and the median will be two. In the second case, when n is even (n=20), then the median will be between positions 10 and 11. If we are using the statistics cube (Cazorla et al., 2020), each student stacks their cube and counts how many there are in each stack, observing that *x10*=2 and *x11*=2, so the average is equal to 2, i.e., the value is immediately apprehended (part b in **Figure 9**). In the third case, when n is even and the central positions 10 and 11 take on different values (*x10*=1 e *x11*=2), then the median will be the mean of those values. In part c in **Figure 9** we use the rod diagram by pasting matchsticks or straws. In these three (iconic) graphical representations of the data we observe that students are able to count the data elements, as the concrete manipulable materials (paper slip, cube or straw) have a one-to-one correspondence with them. Such counting is not possible with the bar graph (part d in **Figure 9**) and rod graph (part f in **Figure 9**) since the bar or rod is continuous and the data elements can no longer be identified. The dot plot still allows the pairing between each data element and its representation (part e in **Figure 9**).

Bar graph (n = 9, odd) Central position: (n+1)/2=10ª $Md = x_{10} = 2$	Dot plot (n = 20, even) Central positions n/2 and (n/2)+1; 10 and 11 are in the same F_i $Md = \frac{x_{10}+x_{11}}{2} = \frac{2+2}{2} = 2$	Rod graph (n =20, even) Central positions are in consecutive F_i $Md = \frac{x_{10}+x_{11}}{2} = \frac{1+2}{2} = 1.5$

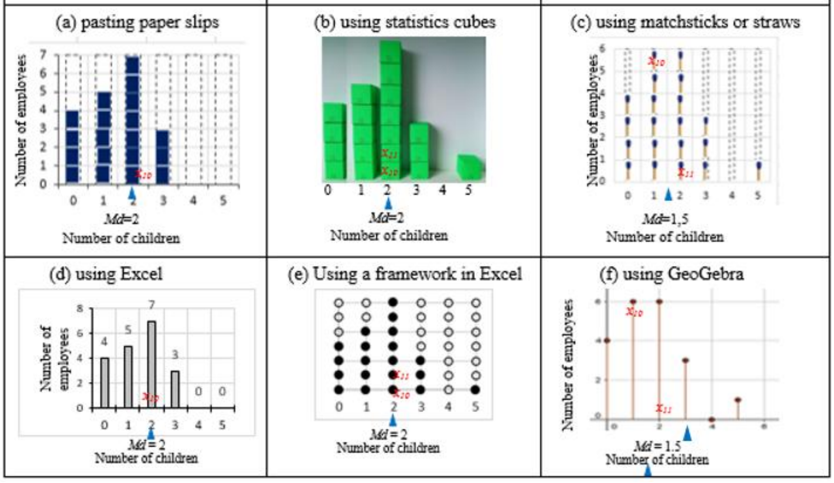


Figure 9. Data & median representations for discrete variables (***S5***) (Source: Authors'own elaboration)

We believe that the various representations help students understand both data distribution and MCT location. Therefore, we favor the use of concrete and manipulable materials in treating statistical data for qualitative variables, as indicated by Cazorla et al. (2020), and for quantitative variables, as pointed out by Cazorla et al. (2021a).

For the mean, in the case of raw data (***S2***), just add up all the values of the variable and divide it by the number of data elements (part a in **Figure 10**); if the data are in an FDT or graph that allows access to frequencies (***S5***), the mean will be calculated as a "weighted" mean, in which the frequency is the weighting factor, as seen in part b in **Figure 10**. We note that, strictly speaking, this is not a weighted mean, but only an abbreviation of calculations. In part c in **Figure 10** the data are represented in a rod graph and the mean in the triangle.

(a) Raw data

X: number of children variable

Simple mean: add up values and divide by the number of elements:

$$M = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{33}{20} = 1.65$$

Add:
5+3+3+0+2+2+2+1+2+1+0+0+1+1+1+0+2+2+2+3 = 33

Divide: 33/20 = 1.65

Mean = 1.65 children/employee

(b) Data grouped in an FDT

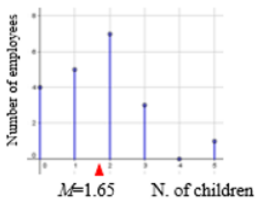
"Weighted" mean: multiply the unique values (x_i) by their frequencies (f_i), add and divide by the number of data elements:

$$M = \frac{\sum_{i=1}^{k} x_i * f_i}{n} = \frac{33}{20} = 1.65$$

FDT in unique values

x_i	f_i	x_i *f_i
0	4	0
1	5	5
2	7	14
3	3	9
4	0	0
5	1	5
Total	20	33

(c) Iconic representation of the mean

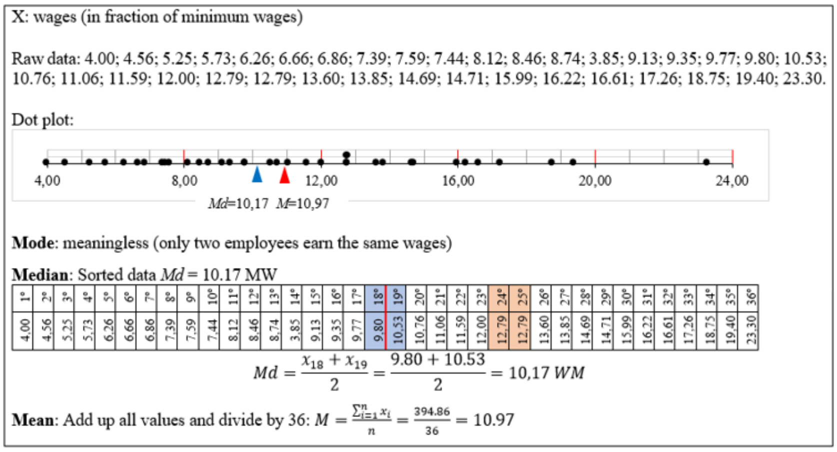


The mean is 1.65 children per employee

Figure 10. Mean for discrete variables that take on few values (**S5**) (Source: Authors'own elaboration)

***Situations 3** and **6** (**S3** and **S6**)*. In these situations, we include discrete variables that take on many values, for example, number of students per class, and continuous variables, such as people's height. To understand how these variables are distributed, since they take on many values, it is interesting to construct a dot plot, which provides an overview of data distribution, minimum and maximum values and amplitude, thus affording a better idea of partitioning for the construction of class intervals and the histogram. The dot plot also enables the construction of the box plot.

In **Figure 11** we present MCT for raw data based on the example by Cazorla et al. (2021c), who use data of the wages of 36 employees. We can see that the mode is meaningless; the median is the mean of the values that occupy the central positions (18 and 19); and the mean is the sum of all values divided by 36.



X: wages (in fraction of minimum wages)

Raw data: 4.00; 4.56; 5.25; 5.73; 6.26; 6.66; 6.86; 7.39; 7.59; 7.44; 8.12; 8.46; 8.74; 3.85; 9.13; 9.35; 9.77; 9.80; 10.53; 10.76; 11.06; 11.59; 12.00; 12.79; 12.79; 13.60; 13.85; 14.69; 14.71; 15.99; 16.22; 16.61; 17.26; 18.75; 19.40; 23.30.

Dot plot:

Mode: meaningless (only two employees earn the same wages)

Median: Sorted data *Md* = 10.17 MW

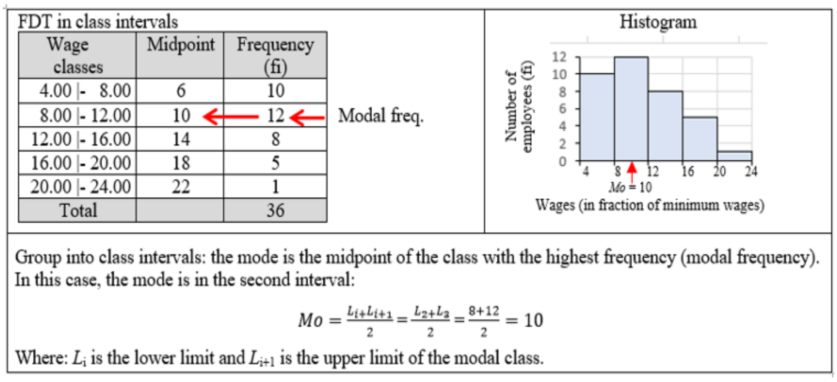
1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°	12°	13°	14°	15°	16°	17°	18°	19°	20°	21°	22°	23°	24°	25°	26°	27°	28°	29°	30°	31°	32°	33°	34°	35°	36°
4,00	4,56	5,25	5,73	6,26	6,66	6,86	7,39	7,59	7,44	8,12	8,46	8,74	3,85	9,13	9,35	9,77	9,80	10,53	10,76	11,06	11,59	12,00	12,79	12,79	13,60	13,85	14,69	14,71	15,99	16,22	16,61	17,26	18,75	19,40	23,30

$$Md = \frac{x_{18} + x_{19}}{2} = \frac{9.80 + 10.53}{2} = 10{,}17\ WM$$

Mean: Add up all values and divide by 36: $M = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{394.86}{36} = 10.97$

Figure 11. MCT for continuous variables with raw data (S3) (Source: Authors'own elaboration)

If the data are already grouped in an FDT in class intervals, just identify the interval components: lower limit (*Li*), upper limit (*Li+1*), interval amplitude (a), and class mark or midpoint (*Li+ Li+1*)/2. To calculate the mode, just find the modal frequency and the mode will be the midpoint of the class interval that contains the modal frequency (**Figure 12**). We note that there are other methods to find the mode. For this we suggest seeing Cobo (2003, p. 56).



FDT in class intervals

Wage classes	Midpoint	Frequency (fi)
4.00 |- 8.00	6	10
8.00 |- 12.00	10 ←	12 ← Modal freq.
12.00 |- 16.00	14	8
16.00 |- 20.00	18	5
20.00 |- 24.00	22	1
Total		36

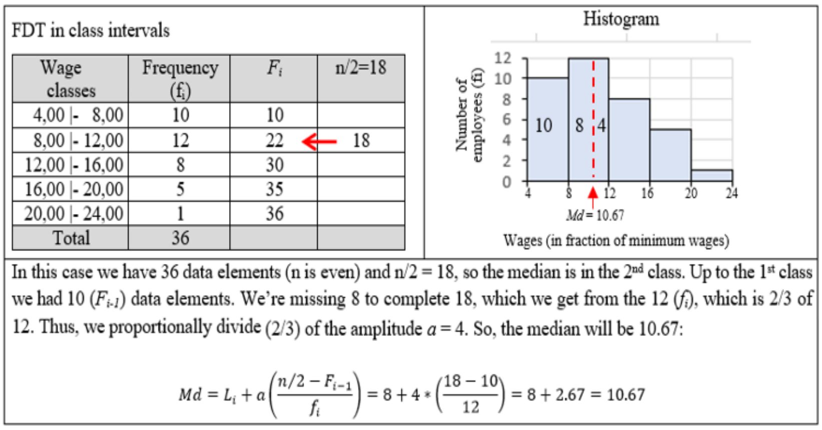
Group into class intervals: the mode is the midpoint of the class with the highest frequency (modal frequency). In this case, the mode is in the second interval:

$$Mo = \frac{L_i + L_{i+1}}{2} = \frac{L_2 + L_3}{2} = \frac{8+12}{2} = 10$$

Where: L_i is the lower limit and L_{i+1} is the upper limit of the modal class.

Figure 12. Mode for continuous variables with grouped data (S6) (Source: Authors'own elaboration)

To find the median we must calculate n/2 and look for the cumulative frequency (F_i) that contains n/2 for the first time and divide the amplitude proportionally to the frequency to complete n/2, according to the formula (**Figure 13**). We note that here it no longer matters whether n is odd or even, as we are in continuous space.



FDT in class intervals

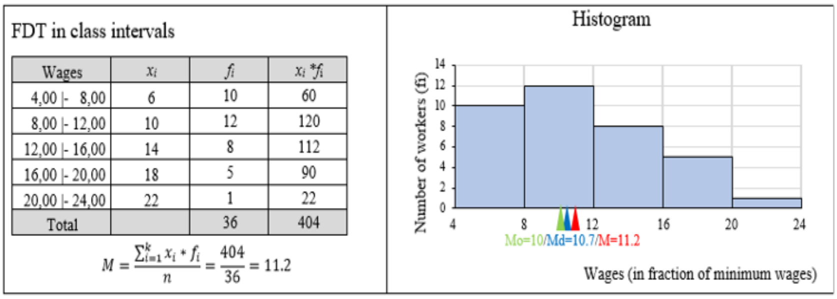
Wage classes	Frequency (f_i)	F_i	n/2=18
4,00 |- 8,00	10	10	
8,00 |- 12,00	12	22 ←	18
12,00 |- 16,00	8	30	
16,00 |- 20,00	5	35	
20,00 |- 24,00	1	36	
Total	36		

In this case we have 36 data elements (n is even) and n/2 = 18, so the median is in the 2nd class. Up to the 1st class we had 10 (F_{i-1}) data elements. We're missing 8 to complete 18, which we get from the 12 (f_i), which is 2/3 of 12. Thus, we proportionally divide (2/3) of the amplitude *a* = 4. So, the median will be 10.67:

$$Md = L_i + a\left(\frac{n/2 - F_{i-1}}{f_i}\right) = 8 + 4 * \left(\frac{18 - 10}{12}\right) = 8 + 2.67 = 10.67$$

Figure 13. Median for continuous variables with grouped data (***S6***) (Cazorla et al., 2021c, **Figure 8**, p. 12)

The mean will be calculated by approximation, as in theory we do not have access to the raw data and the class mark (midpoint) represents all the data contained in this interval. So we "weight" the class mark by frequency, as shown in **Figure 14**.



FDT in class intervals

Wages	x_i	f_i	x_i *f_i
4,00 |- 8,00	6	10	60
8,00 |- 12,00	10	12	120
12,00 |- 16,00	14	8	112
16,00 |- 20,00	18	5	90
20,00 |- 24,00	22	1	22
Total		36	404

$$M = \frac{\sum_{i=1}^{k} x_i * f_i}{n} = \frac{404}{36} = 11.2$$

Figure 14. Mean for continuous variables with grouped data (***S6***) (Source: Authors'own elaboration)

We observe that calculated this way, the mean approximates the mean calculated with the raw data (10.97), as the class mark replaces the values contained in the intervals, and we find that the value of 11.22 is quite close.

In addition to these six situations, the mean has four other particular situations. In **Figure 15** we present the scheme proposed by Cazorla et al. (2019), with situations ***S2***, ***S3***, ***S5***, and ***S6*** in the blue boxes; the green boxes, in turn, feature the specific situations of the mean, which are the aggregated mean (***S7***), when you have the sum of all values but not the individual values; the actual weighted mean (***S8***), when the weighting factor of the unique values is not frequencies, but valuations of another nature (some of them subjective); and the mean of the means or overall mean (***S9***), formed by means of groups. To these situations we have added the simple moving average (***S10***), which we detail below.

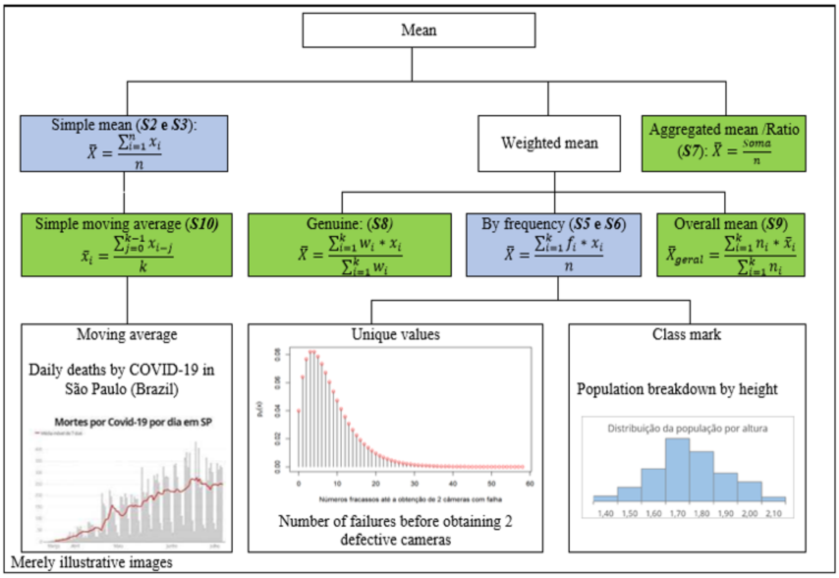


Figure 15. Situations for mean based on **Figure 2** of Cazorla et al. (2019, p. 12)

The moving average, of length k, is widely used to understand the behavior of time series. It is a simple average of k elements that move over time. Its amplitude depends on the seasonality, which can be daily, weekly, monthly, annual or any other type of criterion. For COVID-19, the moving average of the last seven days has been used, including day 7. That is because the daily records fall on weekends and holidays and peak at the beginning of the week due to reporting bureaucracy. In this case, we can calculate the average from day 7, which will be the simple average of the previous seven days, including the day in question. The average of day 8 is the average of day 2 through day 8, and so on:

Moving average of length k

$$\bar{x}_i = \frac{\sum_{j=0}^{k-1} x_{i-j}}{k}$$

For every i, so that: k≤i≤n

7-day rolling average

$$\bar{x}_i = \frac{\sum_{j=0}^{6} x_{i-j}}{7}$$

For every i, so that: 7≤i≤n

In **Figure 16** we illustrate the use of the moving average using data of deaths by COVID-19 in Brazil in April 2021, available on the Brazilian Ministry of Health website (https://covid.saude.gov.br/). On the bar graph we can see that numbers are much lower on Sundays and Mondays, and then on Tuesday there is an increase that persists over the week, falling again as the weekend approaches. In this graph we present the average for the month of April (30 days), the red line, which captures the monthly level, and the moving average, the green line, which mitigates the weekly variation and indicates the downward trend at the end of the month.

Raw data on deaths by COVID-19 in Brazil in April 2021

T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
3762	2922	1987	1240	1319	4195	3829	4249	3693	2616	1803	1480	3808	3459	3560	3305	2929	1657	1347	3321	3472	2027	2914	3076	1305	1139	3086	3163	3001	2595
						2750.6	2820.1	2930.3	3020.1	3100.6	3123.6	3068.3	3015.4	2917.0	2861.6	2906.3	2885.4	2866.4	2796.9	2798.7	2579.7	2523.9	2544.9	2494.6	2464.9	2431.3	2387.1	2526.3	2480.7

Example of moving average calculation for day 7:

$$\bar{x}_7 = \frac{\sum_{j=0}^{6} x_{i-j}}{7} = \frac{x_7 + x_{6+}x_{5+}x_{4+}x_{3+}x_{2+}x_1}{7} = \frac{3762 + 2922 + 1987 + 1240 + 1319 + 4195 + 3827}{7} = \frac{19254}{7} = 2750.6$$

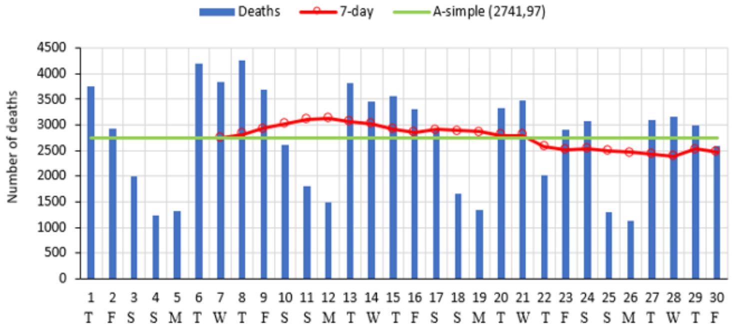


Figure 16. Example of moving average calculation & its graphical representation (Source: Authors'own elaboration)

In **Figure 17** we present the dot plot and box plot, removing the effect of the time variable. Here the median is represented by the line segment inside the box. Some software programs use X to represent the mean. This is a good example of a discrete variable that takes on many values, forcing it to be treated as a continuous variable. We note here that the mode does not exist and that the median does not contribute much to the analysis of the evolution and trend of the number of deaths, the moving average being the most adequate MCT for its understanding.

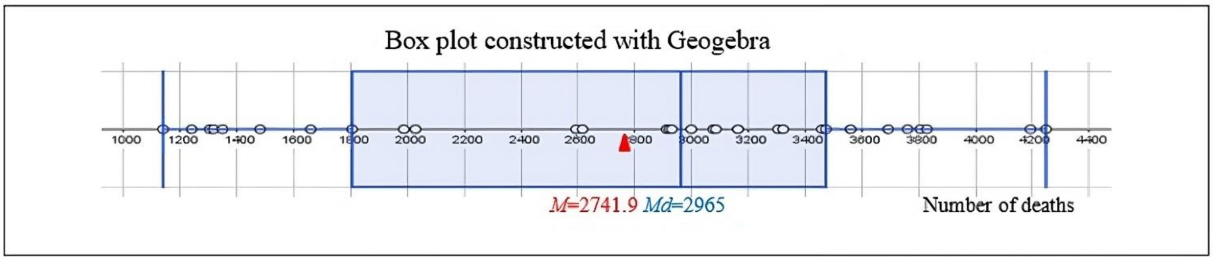


Figure 17. Graphical representation using dot & box plot (Source: Authors'own elaboration)

Invariants

Invariants (I) are formed by the network of concepts in which the conceptual field is grounded, as well as the MCT properties that unify and distinguish the situations that support it, as shown in **Figure 18**.

Properties (Invariants)	Mean	Median	Mode
It is located between the extreme values ($X_{min} \leq MTC \leq X_{max}$)	Yes	Yes	Yes
It is influenced by each and every value (or is it influenced by extreme values)	Yes	No	No
It coincides with one of the values that make it up	Sometimes	Sometimes	Sometimes
It can be a number with no equivalent in physical reality	Sometimes	Sometimes	Sometimes
Its calculation takes into account all values, including nulls and negatives	Yes	No	No
It is a representative value of the data from which it was calculated	Yes	Yes	Yes
It is the equilibrium point (center of gravity)	Yes	No	No
It is closer to all values	Yes	No	No
It can be calculated from partial results	Yes	No	No
The sum of the values can be obtained by multiplying MCT by the number of data elements	Yes	No	No
The sum of deviations from MCT is zero ($\sum_{i=1}^{n}(x_i - MCT) = 0$)	Yes	No	No

Figure 18. MCT properties (invariants) for quantitative variables (Source: Authors'own elaboration)

In the case of nominal qualitative variables, the mode will be one of the categories; in the case of ordinal qualitative variables, the mode and the median will be between the lower category and the higher category.

As we can observe from the situations, understanding MCT requires working with them comprehensively, preferably with a reasonable amount of data that makes it possible to analyze how they are distributed. The strategy of presenting each MCT with few data, focusing on the algorithm or addressing specific aspects is valid to introduce the concepts, but such fragmented and isolated teaching does not contribute to the apprehension of their meanings.

At this point, besides variable type and data format and amount, we must develop the skills to recognize data distribution, as only then will students be able to understand the real meaning of MCT. By distribution we mean recognizing how the variable behaves. There are variables whose distribution is symmetrical by nature, such as the measurement errors of an object or the amount of liquid in a one-liter water bottle. In these cases, the mean, median and mode coincide with the midpoint (P_m) and are easily recognizable, as shown in part as in **Figure 19**.

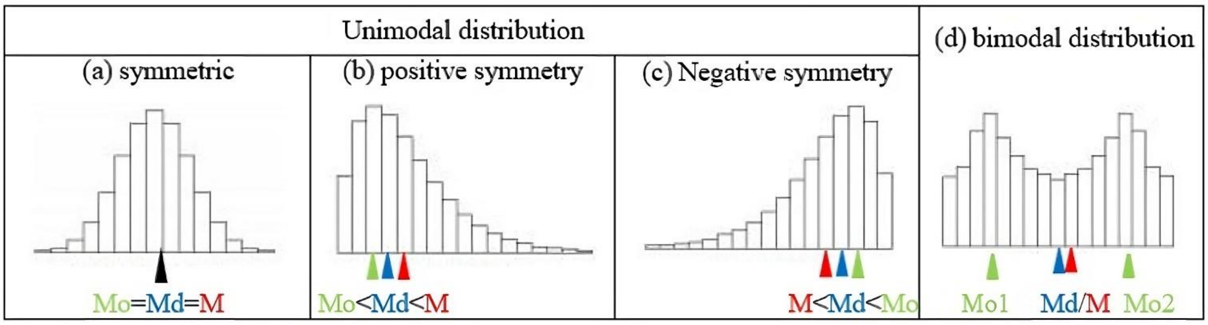


Figure 19. Types of data distribution & MCT (Source: Authors'own elaboration)

However, other variables are asymmetric, such as the breakdown of per capita income of countries or the mathematics score in the Brazilian national secondary education examination (ENEM), in which most countries/students have incomes/scores in the lowest values, and few have income/scores in the highest values. These distributions are called positive asymmetry (part b in **Figure 19**), and in this case the mode is lower than the median, which is lower than the mean, as it is influenced by extreme values. There are also variables that focus on the highest values, for example, the humanities score of ENEM. In this case the mean is lower than the median, which is lower than the mode (part c in **Figure 19**). In these three examples we note that we have a unimodal distribution. In part d in **Figure 19**, in turn, we present a bimodal distribution, typical of variables that are influenced by another variable, as shown by Magina et al. (2022) when describing shoe size by gender, since, in general, girls wear smaller shoe sizes than boys.

In the early school years, when we work with variables such as number of goals per team/match or brothers, etc., they tend to be distributed with positive asymmetry, i.e., it is more common to score few goals and have few brothers. If we were to investigate the number of letters in the first name, this tends to have a more symmetrical distribution, as shown by Cazorla et al. (2021b).

CONCLUSIONS

This article expands on and synthesizes studies by the authors and other researchers aiming to present a framework of the conceptual field of MCT. Its originality lies in systematizing situations that may help students and teachers understand these measures.

We proposed six situations for MCT according to type of variable, date presentation (raw or grouped) and amount of data. Also, we revisited specific situations for the mean that do not apply to the median and mode and we added the moving average. Throughout the situations we exemplified the different possibilities for presenting these measures.

In analyzing the situations, we identified that to understand MCT it is important to consider variable distribution, which should involve a reasonable amount of data in order to explore its full potential.

Furthermore, the most relevant aspect is to recognize the most suitable MCT and algorithm in a specific situation. The lack of such variety of situations, representations and invariants can lead students to make mistakes that endure even after almost 20 years from the introduction of statistics teaching in Brazilian basic education, as shown in the work by Oliveira (2020), in which undergraduate math education students, both freshmen and seniors, cannot find the mean from a symmetrical histogram or have no convincing arguments to choose an MCT in a decision-making situation.

We believe that this results from the way MCT have been taught in Brazilian basic education and basic education teacher training courses, as already indicated by Lopes (2008, 2013), Rodrigues and Silva (2019), and Silva (2011). Therefore, we argue that students should collect data from the early school years, using concrete manipulable materials, organize them in tables and graphs and, from visual observation, intuit the likely location of MCT to later perform the calculations and check how accurate their intuition is.

Finally, it is an article focused on theoretical reflection and systematization of knowledge produced in the field of statistics education. We hope it will help teachers, students and researchers interested in understanding MCT from the perspective of TCF and contribute to the development of statistics reasoning and literacy among students in basic education.

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

Funding: No funding source is reported for this study.

Acknowledgements: The authors thank Espaço da Escrita - Pró-Reitoria de Pesquisa - UNICAMP - for the language services provided.

Ethical statement: Authors stated that the study does not require an ethics committee approval because it is a theoretical article.

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

Andrade, V. (2013). *Os conceitos de medidas de tendência central e de dispersão na formação estatística no ensino médio no Brasil e na França. Abordagem exploratória no quadro da teoria antropológica do didático e da teoria dos campos conceituais* [*The concepts of measures of central tendency and dispersion in statistical training in high school in Brazil and France. Exploratory approach within the framework of the anthropological theory of didactics and the theory of conceptual fields*] [PhD thesis,. Universidade Federal Rural de Pernambuco].

Arteaga, J. (2011*). Evaluación de conocimientos sobre gráficos estadísticos y conocimientos didácticos de futuros profesores* [*Assessment of knowledge about statistical graphs and didactic knowledge of future teachers*] [PhD thesis, Universidade de Granada].

Batanero, C. (2000). Significado y comprensión de las medidas de posición central [Meaning and understanding of measures of central position]. *Uno, 25*, 41-58.

Cazorla, I., Henriques, A., & Santana, C. (2020). O papel dos Ostensivos na representação de variáveis estatísticas qualitativas [The Role of the Ostensives in Understanding Qualitative Statistical Variables]. *Bolema, 34*(68), 1243-1263. https://doi.org/10.1590/1980-4415v34n68a19.

Cazorla, I., Henriques, A., Correia, G., & Santana, C. (2021a). The role of the Ostensives in understanding quantitative statistical variables. *Acta Scientiae, 23*(4), 46-70. https://doi.org/10.17648/acta.scientiae.6532

Cazorla, I., Magina, S., & Santana, C. (2021b). Potencialidades de uma sequência para ensinar as medidas de tendência central nos anos iniciais do ensino fundamental [Potentialities of a didactical sequence to teach measures of central tendency in the early school years]. *Em Teia, 12*(3), 1-26. https://doi.org/10.51359/2177-9309.2021.250551

Cazorla, I., Santana, E., & Utsumi, M. (2019). O campo conceitual da média aritmética: uma primeira aproximação conceitual [The conceptual field of arithmetic media: A first conceptual approach]. *Revemat, 14, Edição Especial Educação Estatística*, 1-21. http://doi.org/105007/1981-1322.2019.e62827

Cazorla, I., Utsumi, M. & Magina, S. (2021c). Revisitando o conceito de Mediana na perspectiva dos Campos Conceituais: uma aproximação teórica [Revisiting the concept of Median from the perspective of conceptual fields: a theoretical approach]. *Seminário Internacional de Pesquisa em Educação Matemática*. https://www.even3.com.br/anais/viiisipemvs2021/380382-revisitando-o-conceito-de-mediana-na-perspectiva-dos-campos-conceituais--uma-aproximacao-teorica/

Cazorla, I., Utsumi, M., & Monteiro, C. (2021d). Variáveis estatísticas e suas representações em gráficos: reflexões para seu Ensino [Statistical variables and its representations in graphs: reflections for teaching]. *Números, 106*, 23-32. http://funes.uniandes.edu.co/23578/

Cobo, B. (2003). *Significados de las medidas de posición central para los estudiantes de secundaria* [*Meanings of central position measures for high school students*] [PhD thesis, Universidade de Granada].

Cobo, B., & Batanero, C. (2000). La mediana en la educación secundaria obligatoria: ¿Un concepto sencillo? [The median in compulsory secondary education: A simple concept?] *Uno, 23*, 85-96.

Figueiredo, T. (2017). *Mediana e quartis: Um caso de estudo das dificuldades de aprendizagem de alunos do 8.° ano de escolaridade* [*Median and quartiles: A case study of the learning difficulties of 8th grade students*] [Master's theis, Universidade de Aveiro].

Lopes, C. (2008). O ensino da estatística e da probabilidade na educação básica e a formação dos professors [The teaching of statistics and probability in basic education and teacher training]. *Caderno Cedes* [*Cedes Notebook*]*, 28*(74), 57-73. https://doi.org/10.1590/S0101-32622008000100005

Lopes, C. (2013). Educação estatística no curso de licenciatura em matemática [Statistical education in the degree course in mathematics]. *Bolema, 27*(47), 901-915. https://doi.org/10.1590/S0103-636X2013000400010

Magina, S., Lautert, S., & Cazorla, I. (2022). A teoria dos campos conceituais na sala de aula [The theory of conceptual field in the classroom]. In S. Magina, S. Lautert, & A. Spinillo (Eds.) *Processos Cognitivos e Linguísticos na Educação Matemática. Teorias, Pesquisas e Salas de Aula* (pp. 55-97). Biblioteca de Educador.

Mayén, S., Cobo, B., Batanero, C., & Balderas, P. (2017). Comprensión de las medidas de posición cesntral en estudiantes Mexicanos de bachillerato [Understanding measures of central position in Mexican high school students]. *Unión, 9*, 187-201.

MEC. (1997). *Parâmetros curriculares nacionais matemática* [*Mathematics national curriculum parameters*]. Ministério da Educação [Ministry of Education].

MEC. (2018). *Base nacional comum curricular* [*Common national curriculum base*]. Ministério da Educação [Ministry of Education].

Morettin, P. A., & Bussab, W. O. (2010). *Estatística básica* [Basic statistics]. Saraiva.

Oliveira, T. (2020). *Contribuições das disciplinas de estatística na formação do futuro professor de matemática para a educação básica* [*Contributions of the disciplines of statistics in the formation of the future teacher of mathematics for basic education*] [Master's thesis, Universidade Estadual de Santa Cruz].

Piaget, J. (1966). *A epistemologia genética* [*The genetic epistemology*]. Martins Fontes.

Piaget, J., & Inhelder, B. (1995). *A psicologia da criança* [*Child's psychology*]. Bertrand Brasil.

Rodrigues, C., Ferreira, A., Carrara, A., Silva, H., & Leite, V. (2021). Educação estatística: O conceito de média móvel no ensino fundamental na pandemia da COVID-19 no Brasil [Statistical education: The concept of moving average in elementary education in the COVID-19 pandemic in Brazil]. *Educação Matemática em Pesquisa: Perspectivas e Tendências* [*Mathematics Education in Research: Perspectives and Trends*]*, 3*, 186-204. https://doi.org/10.37885/21050451

Rodrigues, M., & Silva, L. (2019). Disciplina de estatística na matriz curricular dos cursos de licenciatura em matemática no Brasil [Discipline of statistics in the curriculum of undergraduate courses in mathematics in Brazil]. *Revemat, 14*, 1-21. https://doi.org/10.5007/1981-1322.2019.e62829

Silva, M. A. (2011). A presença da estatística e da probabilidade no currículo prescrito de cursos de licenciatura em matemática: Uma análise do possível descompasso entre as orientações curriculares para a educação básica e a formação inicial do professor de matemática [The presence of statistics and probability in the prescribed curriculum of undergraduate courses in mathematics: An analysis of the possible mismatch between curricular guidelines for basic education and the initial training of mathematics teachers]. *Bolema, 24*(40), 747-764.

Vergnaud, G. (1982). A classification of cognitive tasks and operations of thought involved in addition and subtraction problems. In T. Carpenter, & J. Moser (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 39-59). Lawrense Erlbaun. https://doi.org/10.1201/9781003046585-4

Vergnaud, G. (1987). Conclusion. In C. Janvier (Ed.), *Problem of representation in the teaching and learning of mathematics* (pp. 227-232). LEA Publishers.

Vergnaud, G. (1990). Epistemology and psychology of mathematics education. In P. Nesher, & J. Kilpatrick (Eds.), *Cognition and practice* (pp. 14-30). Cambridge University Press. https://doi.org/10.1017/CBO9781139013499.003

Vergnaud, G. (1996). A teoria dos campos conceituais [The theory of conceptual fields]. In J. Brun (Ed.), *Didáctica das matemáticas* [*Didactics of mathematics*] (pp. 155-191). Instituto Piaget.

Vergnaud, G. (1998). Comprehensive theory of representation for mathematics education. *Journal of Mathematics Behavior, 2*(17), 167-181. https://doi.org/10.1016/S0364-0213(99)80057-3

Vygotsky, L. (2001). *A construção do pensamento e da linguagem* [The construction of thought and language] (Trans. P. Bezerra). Martins Fontes.