

MATHEMATICAL ERROR IN THE MICROSOFT EXCEL PROGRAM “MODE” FUNCTION, ITS IMPLICATIONS AND POSSIBLE CORRECTIONS

 <https://doi.org/10.56238/arev7n1-135>

Submission date: 12/15/2024

Publication date: 01/15/2025

Rafael Alberto Gonçalves¹ and Anderson Michel Hornburg²

ABSTRACT

INTRODUCTION: Mathematics is a science that permeates several areas of knowledge and is fundamental to understanding the world around us. Among its branches, we have statistics, whose function consists of organizing a list (set of raw data) and extracting significant data from them, such as the arithmetic mean, mode, median, sample standard deviation, and population standard deviation, among other calculations. This article discusses the importance of statistics in the contemporary world, with an emphasis on the concept of the mode means, also focusing on its practical applications for educators, students, and various professionals; it also addresses the impacts of calculation errors arising from Microsoft's Excel spreadsheet. Finally, practical solutions will be presented to correct these inconsistencies, using VBA (Visual Basic for Applications).

Keywords: Mathematical Error. Microsoft Excel. Mode Function (mode).

¹ Master's Degree in Teaching Natural Sciences and Mathematics (FURB).

LATTES: <http://lattes.cnpq.br/1469248630990193>

² Specialization in Business Law (PUC Minas Gerais).

LATTES: <http://lattes.cnpq.br/6637887531082916>

INTRODUCTION

Mathematics is a science that permeates several areas of knowledge and is fundamental to understanding the world around us. Among its branches, we have statistics, whose function consists of organizing a list (set of raw data) and extracting significant data from them, such as the arithmetic mean, mode, median, sample standard deviation, and population standard deviation, among other calculations. This article discusses the importance of statistics in the contemporary world, with an emphasis on the concept of the mode means, also focusing on its practical applications for educators, students, and various professionals; it also addresses the impacts of calculation errors arising from Microsoft's Excel spreadsheet. Finally, practical solutions will be presented to correct these inconsistencies, using VBA (Visual Basic for Applications).

2 STATISTICS: FASHION.

The field of exact sciences is known for its expertise in the logical and exact demonstration of certain cases, such as statistics, where raw data is transformed into meaningful information, from which it is possible to extract bases for establishing definitions, patterns, and trends that support complex decision-making.

As we live in an information age, in which we are bombarded with data at every moment, understanding statistics and correctly applying its fundamentals is essential for defining business strategies that can determine the success or failure of a product or even the company itself, the inclusion or exclusion of a certain agent in situations such as selection processes or competitions, or even the creation of statistics at educational levels, among many other uses. Statistics allows us not only to understand what the numbers mean but also to communicate these results efficiently and effectively.

According to Silva (2015, p. 10),

statistics is a field of study focused on the production of methodology for collecting, organizing, describing, analyzing, and interpreting data, as well as obtaining valid conclusions and making reasonable decisions based on such analyses.

Statistics can be divided into two main categories: descriptive statistics and inductive statistics, also called inferential statistics. The first focuses on collecting, organizing, and describing data, while the second uses this information to make inferences (hence the name), that is, to analyze and interpret the data collected. It is in descriptive statistics that we will have the so-called measures of central tendency, divided into arithmetic mean, mode, and median (harmonic and geometric).

The arithmetic mean represents the average, typical value. It is calculated by adding all the numbers and dividing them by the overall quantity. The median, in turn, consists exactly of the value that is in the middle of the analyzed list, if the universe is odd, or the two central data, if it is even.

The mode (Mo), in turn, is the measure of central tendency that represents the value or parameter that appears most frequently in a data set. In other words, the mode is the number that repeats itself the most. It can be classified into five classes: amodal when there are no repetitive numbers; unimodal, when there is only one value that repeats itself more than the others; bimodal, when there are two values that repeat themselves with the same frequency; trimodal, when there are three values that repeat themselves with the same quantity; and polymodal, when four or more values repeat themselves in the same quantity.

The calculation of the mode can be done manually, simply by counting the frequency of the data classified in ascending and descending order and counting the quantity that each number appears. However, depending on the volume of information to be analyzed, spreadsheets, such as Excel, may be used. In this program, developed by Microsoft, there is the "MODE" function, whose command shows the most frequent value in a range of selected or typed cells [such as, for example, the mode of numbers in cells 1 to 16 of column A "=MODE(A1:A16)"]. The function is intuitive and very practical, but unfortunately, it presents configuration and calculation errors, as will be seen throughout the article.

ERRORS IN THE "MODE" FUNCTION IN THE MICROSOFT EXCEL PROGRAM

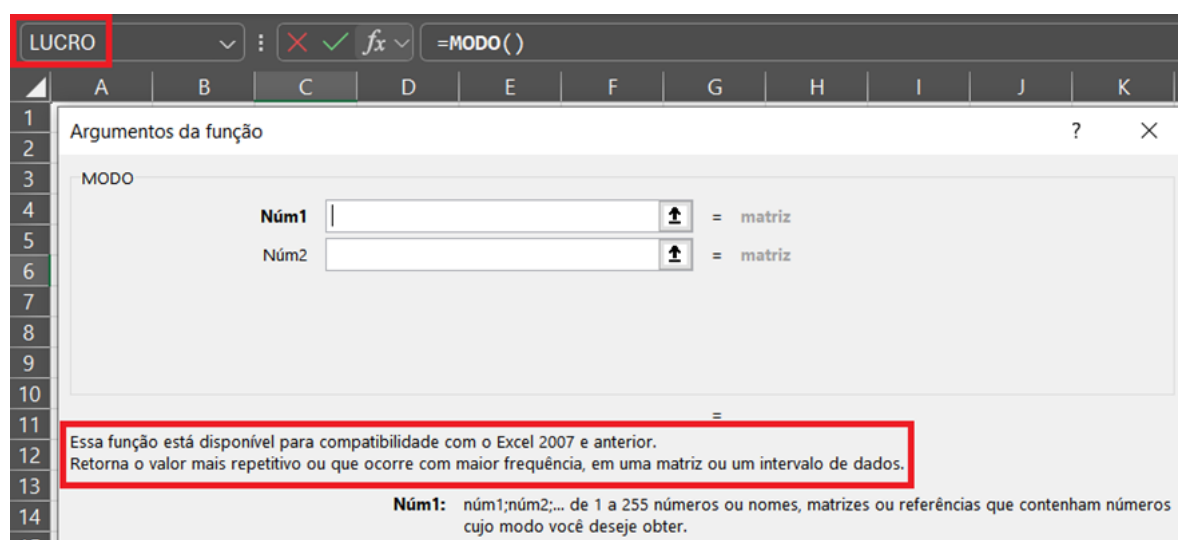
The advancement of technology has profoundly changed our society, enabling greater precision in calculations or project development in a shorter period. However, there have been programming errors in software or applications, or even usage errors, due to lack of knowledge or incorrect application of concepts, whether mathematical or linguistic. Such errors are much more than mere errors, because the inferred data ends up being propagated, generating successive mistakes.

Understanding statistics, including the concept of mode, is fundamental in the globalized world, in fact, in all areas: in education, where educators can identify which contents are being best understood and which ones need more attention; in the government sphere, by obtaining socioeconomic data from civil society, the education system, health, etc., as well as in the implementation of public policies; or even in the

private sector, when analyzing market trends, customer satisfaction, and operational efficiency.

Therefore, the need for accurate tools for statistical analysis is evident, and this is where Microsoft Excel becomes a popular choice worldwide, extremely useful, but not infallible.

The Office package was launched in 1990 and, since then, has been receiving continuous updates. However, it is clear that, to date, Microsoft has not made the necessary adjustments to this very fundamental MODE (Statistical Mode) function. When entering data into this function, the name box displays the word “PROFIT”, an inconsistency that causes a real breakdown in understanding, especially for laymen, since it has no relation whatsoever to the statistical purpose of the function.



The discrepancy between the designed function and the presented response compromises the user experience, leading them, especially laypeople, to infer that it addresses the incorrect function, or even that the data may represent a profit.

There are also other compromising mathematical errors, let's see:

B3				
=MODE(A1:A4)				
	A	B	C	D
1	1			
2	2	Excel	Fórmula	Observação
3	3	#N/D	=MODE(A1:A4)	#N/D (acrônimo de) não disponível
4	4			

Based on the concepts presented above, which relate to basic statistics, it can be seen that Excel presents an incorrect calculation since in this specific case the expected result is 'amodal'. That is, there is no value that appears more frequently in the analyzed data set. Instead of dealing with this situation appropriately, the program informed that the formula was unable to return a valid result.

Amodal data, although less frequent in a set, are highly relevant in decision-making, since they highlight the absence of patterns and, as such, demonstrate the need for a new approach or new method of executing a project, or even problems in redirecting public resources in a given location.

That said, it is understood that the exclusion of the modal classification in the spreadsheet is inappropriate, since an amodal response is a valid condition that should be recognized and communicated appropriately, and not due to a supposed calculation or information insertion error. The inability to identify and treat amodal results compromises the integrity of the statistical analyses and the reliability of the data presented, therefore, improving the software is crucial.

B3					=MODOS.ÚNICO(A1:A4)	
	A	B	C	D		
1	2					
2	1	Excel	Fórmula	Observação		
3	1	2	=MODOS.ÚNICO(A1:A4)	Retorna o 1º número repetido & registrado		
4	2					

Although Excel has undergone improvements, it still does not adequately meet the properties of this function, as it is possible to notice a serious error in the data obtained. Now, as can be seen in the image above, the numbers 1 and 2 repeat in the same quantity, however, Excel gives preference to the first number recorded which repeats, resulting in a real collapse in the interpretation of the data. This behavior is unacceptable, because, in this specific case, the response should be bimodal, recognizing both numbers as modes. Excel's inability to correctly identify and report bimodal cases significantly compromises the accuracy and usefulness of statistical analyses. Correcting this flaw is also crucial to ensure that the results are representative and reliable, avoiding distortions that could negatively impact decisions based on the data provided.

B3					=MOD0.MULT(A1:A4)				
	A	B	C			D			
1	1								
2	2	Excel	Fórmula			Observação			
3	2	1	=MOD0.MULT(A1:A4)			Retorna os valores em células distintas			
4	1	2							
5									

Although it is considered that the latest software update attempted to implement bimodal classification in some way, it can be seen that this was done in order to expand the boundaries of the calculation cells. In detail, the algorithm inserts values into sequential cells, but, as can be seen in the image, cell B4, which contains the number two, does not follow the same format as cell B3.

This oversight not only compromises the consistency and aesthetics of the spreadsheet but can also interfere with data or other functions correlated in the rows below, creating not only interpretation problems but also spreadsheet configuration problems.

In the following image, it will be seen that the error not only perpetuates itself, but evolves, and incessantly sows chaos in its path.

B3		=MOD0.MULT(A1:A6)		
	A	B	C	D
1	1			
2	2	Excel	Fórmula	Observação
3	3	#DESPEJAR!	=MOD0.MULT(A1:A6)	(Despejar)?
4	1	Na célula B4, digitamos para demonstrar o padaroxo da função		
5	2			
6	3			

In this case, the spreadsheet intends to DUMP (to vacate, evacuate, empty - Priberam Dictionary of the Portuguese Language), that is, to dump the requested information in cells below the spreadsheet boundaries, which certainly makes the column of the aforementioned tab completely unfeasible, depending on the result of the fashion. It is quite clear that this function is full of flaws, reminiscent of F1 cars that gain fins in desperate attempts to evolve. It would be better if the function used the verb EXPAND.

	A	B	C	D
1	1		Coluna1	
2	2			
3	3		Média	2,5
4	4		Erro padrão	0,423
5	1		Mediana	2,5
6	2		Modo	1
7	3		Desvio padrão	1,195
8	4		Variância da amostra	1,429
9			Curtose	-1,456
10			Assimetria	-4E-17
11			Intervalo	3
12			Mínimo	1
13			Máximo	4
14			Soma	20
15			Contagem	8

The data analysis feature was introduced in Excel in the 1990s, however, after more than twenty years and in version 365, it still has significant flaws, such as the mode function. For example, when calculating the standard deviation, the summary does not specify whether the value is sample or population. This flaw is particularly serious since distinguishing between sample and population standard deviation is crucial for the accuracy of statistical analyses. The lack of this information can lead to wrong decisions, impairing students' learning and providing incorrect data to other institutions. It is regrettable and incomprehensible that after so many updates and improvements these flaws persist in paid software.

PROGRAMMING CORRECTIONS

Aware of the possibility of correcting and truly meeting the properties of the mean model, the authors developed an efficient calculation in VBA (Visual Basic for Applications) for hypotheses in which there is amodal data.

	A	B	C	D
1	1			
2	2	Autores	Fórmula	Observação
3	3	Amodal	=Moda_Autores(A1:A4)	Correto
4	4			

For other modal classes:

B3 X ✓ fx =Moda_Autores(A1:A10)				
	A	B	C	D
1	1			
2	2	Autores	Fórmula	Observação
3	3	1; 2; 3; 4; 5	=Moda_Autores(A1:A10)	Correto
4	4	Na célula B4, digitamos para demonstrar a correção da função		
5	5			
6	1			
7	2			
8	3			
9	4			
10	5			

This algorithm calculates amodal, unimodal, bimodal, trimodal, and also polymodal cases, with the efficiency of providing the answer in a single cell, a problem that persists in the original Excel programming. As can be seen, VBA itself was used to correct the errors that the developers missed.

```
Function Moda_Autores(Intervalo As Range) As String
    Dim Frequencia As Object
    Dim Valor As Variant
    Dim Resultado As String
    Dim MaxFreq As Long
    Dim Contagem As Long
    Set Frequencia = CreateObject("Scripting.Dictionary")

    For Each Celula In Intervalo
        Valor = Celula.Value
        If Not IsEmpty(Valor) Then
            If Frequencia.Exists(Valor) Then
                Frequencia(Valor) = Frequencia(Valor) + 1
            Else
                Frequencia.Add Valor, 1
            End If
        End If
    Next Celula

    MaxFreq = Application.WorksheetFunction.Max(Frequencia.Items)

    For Each Key In Frequencia.Keys
        If Frequencia(Key) = MaxFreq And MaxFreq > 1 Then
            Resultado = Resultado & Key & "; "
            Contagem = Contagem + 1
        End If
    Next Key

    If Contagem = 0 Then
        Moda_Autores = "Amodal"
    ElseIf Contagem > 0 Then
        Resultado = Left(Resultado, Len(Resultado) - 2)
        Moda_Autores = Resultado
    End If
End Function
```

It is essential to remember that the authors have already identified dozens of basic and compromising flaws in the Excel spreadsheet, which were the subject of previous publications. The correction, in addition to being a significant victory, highlights the need to review and correct these critical functions, ensuring the reliability of the overall results.

FINAL CONSIDERATIONS

Statistics, in this case the mode, is an important tool for understanding diverse data, in several areas of knowledge. However, the effectiveness of this analysis requires the reliability of the calculation results presented by software such as Microsoft Excel, as well as an understanding of basic mathematical concepts by the users who are making the observation.

In the meantime, it was seen that Excel, although extremely popular and practical, has limitations and errors in the "MODE" function, since it does not understand the basic concepts of amodal results and others, in addition to presenting a series of flaws regarding the use of cells. This is unfortunate, considering that it is paid software, and dangerous since the results of the erroneous calculations induce interpretations that will be the basis for important decision-making, whether in the public or private sphere and even in education, which is the essence of this research.

However, these limitations and flaws were, with great effort and dedication by the authors, overcome through VBA programming, as shown in the results presented in this article. Algorithms were developed that analyze all modal cases, with results in single cells, which does not compromise other possible calculations in the spreadsheet. These improvements ensure the accuracy of the statistical calculations generated by Excel, as well as aligning the software with correct mathematical and semantic standards.

Finally, the improved algorithm in VBA was subjected to classroom testing and demonstrated extreme accuracy in the statistical calculations, in addition to promoting, of course, a clearer understanding of the mathematical concepts involved by the students.

REFERENCES

1. Cormen, T. (2012). Algoritmos: Teoria e prática (3ª ed.). GEN LTC. E-book. p. 840. ISBN 9788595158092. Disponível em: <https://app.minhabiblioteca.com.br/reader/books/9788595158092/>. Acesso em: 19 out. 2024.
2. Gonçalves, R. A., & Hornburg, A. M. (2024). Erro semântico e matemático na função “ALEATORIOENTRE” do programa Microsoft Excel, suas implicações e possível correção. In C. Bianchessi (Ed.), Temas em educação e ensino: Olhares interdisciplinares, reflexões e saberes (1ª ed., Vol. 5, pp. 333–340). Editora Bagai. ISBN 978-65-5368-484-3.
3. Levine, D. M., Stephan, D. F., Krehbiel, T., & Berenson, M. L. (2014). Estatística e teoria e aplicações: Usando o Microsoft® Excel em português (6ª ed.). LTC.
4. Medeiros, J. de, & Gonçalves, R. A. (2018). Aplicações tecnológicas em ambiente acadêmico: Um olhar sobre o uso de planilhas eletrônicas e seus impactos sócio-mercadoológicos. In R. M. Carrara & M. A. Orth (Eds.), Educação tecnológica na América Latina (pp. 51–77). Contexto Digital Tecnologia Educacional.
5. McFedries, P. (2012). Fórmulas e funções: Microsoft Excel 2010 (3ª ed.). Alta Books.
6. Priberam. (2025). Dicionário Priberam da Língua Portuguesa. Disponível em: <https://dicionario.priberam.org/despejar#:~:text=verbo%20transitivo&text=Desocupar%3B%20evacuar%3B%20esvaziar>. Acesso em: 6 jan. 2025.
7. Silva, J. L. de C. e, Fernandes, M. W., & Almeida, R. L. F. de. (2015). Estatística e probabilidade (3ª ed.). EdUECE.