# Research on Calculation Method of Mean in Sports Statistical Data with EXCEL

Aining Li<sup>1</sup>, Shiwei Zhang<sup>2\*</sup>, Zhanbin Feng<sup>3</sup>, Qiyuan Zhang<sup>4</sup>

<sup>1</sup> Dept. of Library and Information Center, Laiwu Vocational and Technical College, Laiwu, Shandong 271100, China.

<sup>2</sup> Dept. of Continuing Education, Laiwu Vocational and Technical College, Laiwu, Shandong 271100, China.

<sup>3</sup> Dept. of Physical Education, Shandong Polytechnic College, Jining, Shandong 272067, China.

<sup>4</sup> Undergraduate of 2018 Mechanical Design and Manufacturing and Automation of School of Mechanical Engineering, Sichuan University, Chengdu, Sichuan 610207, China.

\* Corresponding author. Tel.: + 86-015563419160; email: lwzyzsw@163.com Manuscript submitted February 4, 2020; accepted May 12, 2020. doi: 10.17706/ijcce.2020.9.3.141-148

**Abstract:** The mean is the most commonly used statistical indicator to reflect the central tendency of data distribution, in this study, the mean was reclassified according to whether the effect of extreme value was considered or not. According to the definition and specific classification of mean value, We explain in detail how to calculate the normal mean by ways of "built-in function", "pivot table chart" and "descriptive statistics", symmetry trim mean by TRIMMEAN function and asymmetrical trim mean by ways of nested function and VBA self-built function with the date of the 10 referees' scores in the kindergarten gymnastics competition of normal education department of our college. If we fully exploit the powerful statistical function of Excel, we can better meet our sports practice of statistical processing needs.

Key words: Statistical data, mean, excel, calculation method.

## 1. Introduction

In the practical activities of sports, we often collected numerous data, which are valuable materials in sports science, they usually contain a wealth of sports information, however, these information are often hidden among a large amounts of data and not obvious. Therefore, in order to gain more valuable information from these complex numerous and complicated data, we need the data for statistical processing and analysis, which makes to understand the distribution characteristics of the data accurately and to reveal the inherent law of development [1]. In data distribution characteristics, central tendency is an important aspect, which is used to describe or explain the whole picture of a set of data and the distribution of representative and typicality, mainly including the arithmetic mean, median, mode, harmonic mean, geometric mean and other indicators. Among them, the arithmetic mean is the most commonly used statistical indicators, usually along with the standard deviation, which represents a discrete trend, to describe the distribution characteristics of the data.

## 2. The Definition and Classification of Arithmetic Mean

## 2.1. The Definition of Arithmetic Mean

The arithmetic mean, as we usually call it mean, is the quotient of the sum of observations divided by the

number of observations, and is only called the arithmetic mean when it is distinguished from the harmonic mean, the geometric mean, etc. It's advantage is that representation is easy to accept, all values are used, all values contribute to representation, there is always one and only one, not as there are many as there are modes .It's disadvantage is that it is easily influenced by the extreme value at both ends, which weakens its representativeness [2]. It's usually represented by an  $\frac{1}{\chi}$  or an 'M'. The formula is:

$$\overline{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

#### The Classification of Arithmetic Mean 2.2.

In the practical activities, according to whether we often consider the effect of extreme value on the mean, we divide the arithmetic mean into the normal mean (mean) and the trim mean (also known as the internal mean), and the trim mean into the symmetric trim mean and the asymmetric trim mean with taking out the different extreme values on both end.

## 3. The Calculation Methods of the Mean in EXCEL

Excel is an important component of Microsoft Office, which provides many common data processing functions, including a large number of formula functions, various graphical functions, powerful statistical analysis of data, etc. [3] In this paper, we take the 10 referees' scores in the kindergarten gymnastics competition of normal education department of our college as an example (see Table 1), and explain how to calculate various mean in EXCEL.

Number of referees	Class one	Class two	Class three	Class four	
1	7.5	8.8	7.9	7.2	
2	8.1	9.9	7.6	9.9	
3	8.3	7.8	7.1	8.6	
4	9.2	8.5	8.6	7.8	
5	7.6	7.2	8.6	9.2	
6	9.2	8.2	7.8	8.3	
7	8.5	9.1	9.8	8.5	
8	9.1	7.1	6.5	8.9	
9	7.9	7.4	7.9	9.1	
10	8.5	8.6	8.3	7.9	

10

## 3.1. The Calculation of the Normal Mean

## 3.1.1. Using the functional method to calculate the mean

There are many built-in functions in EXCEL that we can use AVERAGE () to calculate the mean, or we can use DAVERAGE () or SUMIF/COUNTIF to calculate the conditional mean. When using a function, we have two functions inputting ways in Excel, one is to directly input the function formula content, if you can enter "=AVERAGE (B2: B11)" in the B12 cell, you will calculate the mean (see Table 2). Another is to use "Insertion function" dialog box, then complete the calculation according to clew. [4]

		0			
Number of referees	Class one	Class two	Class three	Class four	
1	7.5	8.8	7.9 7.2		
2	8.1	9.9	7.6	9.9	
3	8.3	7.8	7.1	8.6	
4	9.2	8.5	8.6	7.8	
5	7.6	7.2	8.6	9.2	
6	9.2 8.2		7.8	8.3	
7	8.5	9.1	9.8	8.5	
8	9.1	7.1	6.5	8.9	
9	7.9	7.4	7.9	9.1	
10	8.5	8.6	8.3	7.9	
	=AVERAGE (B2: B11)				

Table 2. The Mean by Using AVERAGE Function to Calculate

## 3.1.2. Using the pivot table to calculate the mean

For cases where two conditions must be averaged at the same time and the number of persons needs to be averaged together, the easiest way to handle this is to use data pivot table to establish cross list. Data pivot tables are interactive tables that quickly summarize and cross-list large amounts of data by rotating their rows or columns to see different summaries of the source data, and by displaying different row labels to filter the data, which shows the concrete manifestation of Excel's powerful data processing ability. In the application of Pivot table for data summary, we should open the "Pivot Table and Pivot Chart Report" function in the "Data" menu, pressing the "Pivot Table and Pivot Chart Wizard" to finish step by step. [5] While accessing cross-analysis table, we can rationalize the various parts of the pivot table according to the research needs to determine the row, column, field and data content. What should be noted is that if the pre-selected calculation sum does not meet the study requirements, you may change the summary way by double-click "Sum" to select the appropriate calculation options. Summary methods include sum, count, mean, maximum, minimum value, product, value count, standard deviation, population standard deviation, variance, population variance and so on, this research summary way chooses mean.

## 3.1.3. Using "descriptive statistics" in "data analysis" to calculate the mean

Excel provides a set of data analysis tools, called DAD (Data Analysis Database). Procedures can be saved in the establishment of complex statistical analysis, since we only to provide the necessary data and parameters for each analysis tool, then the tool will use appropriate statistics, and the output table will show the corresponding results. Some of these tools can also generate charts as they generate the output table. Using add-ins and analysis Tool Pak installed, you can get "Data Analysis" in "Tools" menu. By using "descriptive statistics" in "data analysis", we can simultaneously obtain the statistics of multiple groups of data in the selected area, such as arithmetic mean, median, mode, standard deviation, variance, extreme value, etc.

In this study, We describe the statistics dialog box in the "input area" to set the data area to be processed (B1: E11) ,select "column by column" in the "grouping mode", click the "flag in the first row", set the output area in H 1 position of the current work table, click on "the summary Statistics", choose confidence level of the mean as "95%", the setting good need to display the most value number, points out "certainly"

then (see Table 3).

Table 5. The Mean by Using Descriptive Statistics to Calculate							
	Class one	Class two	Class three	Class four			
Mean	8.39	8.26	8.01	8.54			
Standard Error	0.199694211	0.284487844	0.285365263	0.248193473			
Median	8.4	8.35	7.9	8.55			
Mode number	9.2	#N/A	7.9	#N/A			
Standard Deviation	0.631488541	0.899629553	0.902404196	0.784856675			
Variance	0.398777778	0.809333333	0.814333333	0.616			
Kurtosis	-1.32391354	-0.507984144	1.068454799	-0.070870881			
Degree of deviation	0.045600704	0.329762311	0.357801642	-0.017788011			
Region	1.7	2.8	3.3	2.7			
Minimum Value	7.5	7.1	6.5	7.2			
Maximum Value	9.2	9.9	9.8	9.9			
Make Peace	83.9	82.6	80.1	85.4			
Number of observations	10	10	10	10			
Maximum 1	9.2	9.9	9.8	9.9			
Minimum 1	7.5	7.1	6.5	7.2			
Confidence level(95.0%)	0.451739688	0.643556213	0.645541072	0.561452641			

Table 3. The Mean by Using "Descriptive Statistics" to Calculate

## 3.2. Calculation of Trim Mean

## 3.2.1. Calculation of the symmetrical trim mean

The symmetrical trim mean refers to the average of an array after removing the same numbers extreme value in the head and tail, which is in order to eliminate the extreme's impact on the mean. The calculation of the trimmed mean is also by the TRIMMEAN function. The grammar of function is TRIMMEAN (array, percent), in which the "Array" refers to the data array or the data field , and "Percent" refers to the percent of the removed extreme, which is the multiple of 2 to remove same numbers of extreme in both ends [6]. In this example, the 1 highest score and the 1 lowest score should be removed from each of the 10 judges, the percent is: Percent=2/10=0.2, we can input it in the corresponding in cell B12: "=TRIMMEAN(B2:B11,0.2)", then the symmetrical trimmed mean will be gotten (see Table 4).

Table 4. The Symmetrical Thin Mean by Osing TRIMMEAN Function to Calculate							
Number of referees	Class one	Class two	Class three	Class four			
1	7.5	8.8	7.9	7.2			
2	8.1	9.9	7.6	9.9			
3	8.3	7.8	7.1	8.6			
4	9.2	8.5	8.6	7.8			

Table 4. The Symmetrical Trim Mean by Using TRIMMEAN Function to Calculate

5	7.6	7.2	8.6	9.2	
6	9.2	8.2	7.8	8.3	
7	8.5	9.1	9.8	8.5	
8	9.1	7.1	6.5	8.9	
9	7.9	7.4	7.9	9.1	
10	8.5	8.6	8.3	7.9	
	=TRIMMEAN (B2:B11,0.2)				

## 3.2.2. Calculation of the asymmetrical trim mean

In sports practice, we sometimes encounter another complex situation, such as in a set of data in the head and tail after removing different numbers of extreme value to calculate the mean. This situation is called asymmetric trim mean, and it can not be simply achieved through a statistical function, the following is the discussion of several methods to achieve.

## 3.2.2.1. Nested function method

A nested function is a function that the researcher may need to use as an argument to another function under certain circumstances. When dealing with a complex problem, the use of one or two functions alone can not effectively solve the problem, or even can not solve it, the problem can be easily solved by using nested functions. [7] For this problem, it is easy to realize with nested function method, and train of thought is clear: to calculate the sum of all the data, remove the need to eliminate the different number of maximum and minimum, and then divided by the number of remaining data, then we can take the asymmetrical trim mean. We will only need to use a function or formula language expression for this process. In this example, we want to get rid of one maximum and two minimum, and input the following formula in cell B12: =(SUM(B2:B11)-LARGE(B2:B11,1)-SMALL(B2:B11,1)-SMALL(B2:B11,2))/ (COUNT(B2:B11)-3) (see Table 5).

Number of referees	Class one	Class two	Class three	Class four	
1	7.5	8.8	7.9	7.2	
2	8.1	9.9	7.6	9.9	
3	8.3	7.8	7.1	8.6	
4	9.2	8.5	8.6	7.8	
5	7.6	7.2	8.6	9.2	
6	9.2	8.2	7.8	8.3	
7	8.5	9.1	9.8	8.5	
8	9.1	7.1	6.5	8.9	
9	7.9	7.4	7.9	9.1	
10	8.5	8.6	8.3	7.9	
	=(SUM(B2:B11)-LARGE(B2 :B11,1)-SMALL(B2:B11,1)- SMALL(B2:B11,2))/ (COUNT(B2:B11)-3)				

Table 5. Asymmetrical Trim Mean by Using Nested Function Method to Calculate

## 3.2.2.2. Using VBA self-built function

The above nested function is suitable for cases with few numbers of data. If the competition had dozens of classes or 200 tables to process, we would have to do the same thing many times. This competition is to remove one maximum score and the corresponding two minimum points, and the future competition to remove a number of maximum score and minimum points, or the number of referee are different, and need to repeat the input formula to complete, which is relatively cumbersome.

Excel also has a powerful function – VBA. VBA Technology (Visual Basic for Applications) is a programming language developed by Microsoft to perform automated tasks in its desktop Applications. VBA programming in Excel does not require the installation of any other proprietary programming language development platform, it is a macro programming language of Visual Basic, which can make full use of the functions contained in the Excel itself to program, reduce the complexity of programming, can achieve automation, batch processing tasks program. [8]-[10] According to their own needs, the researcher can write the function through the VBA, and it can be ported to other EXCEL files or any computer.

In this study, assuming that the function to achieve the above function is named "mean\_way ", the following code can be written in VBA to calculate the asymmetrical trim mean. After opening the VBA editor window, right-click on This work book in the Engineering File Explorer Window, select insert module, and enter the following code in the inserted module:

The function is defined as "mean\_way " and ran is the region of the data being processed, in this case the region is (B2: B11), La is the maximum number of numbers to be removed, in this case the maximum number is 1, sm is the minimum number of numbers to be removed, and in this case the minimum number is 2.

```
Function mean_way (ran As Range, la As Integer, sm As Integer) As Variant
Dim sum, i, j, count As Variant
'Find the sum of all the data
Sum = Application. Work sheet Function. Sum (ran)
'Subtract minimums of sm
For i = 1 to sm Step 1
Sum = sum – Application .Work sheet Function. Small (ran, i)
Next
'Subtract maximums of la
For i = 1 to la Step 1
Sum = sum - Application. Work sheet Function. Large (ran, i)
Next
'Find the number of data left after the minimum and maximum number is removed
Count = Application. Work sheet Function. count (ran) - la - sm
'Calculate the mean
mean_way = sum / count
End Function
```

After completing the code input and returning to the EXCEL table, using the self-built function "mean\_way" in Cell B12, you can calculate the specific value of the mean of the asymmetric trim mean. If you want to use the built in function on other EXCEL files or computers, you can export the edited module file, copy it, and import the module file into the desired EXCEL file.

## 4. Conclusion

Although Excel is not classified as statistical software, and its statistical functions can not be compared

with professional statistical software such as SPSS and SAS, but it is a data statistics tool with a collection of data table, work function, VAB applications and report processing, which can not only deal with figures, but also with graph and character. The data in Excel can be shared with and converted to those in Word and SPSS and any other databases. Therefore, it is necessary to further explore the data statistical processing method of EXCEL, study the expansion software of EXCEL, comprehensively improve the data statistical function of EXCEL software, and better serve sports practice.

## **Conflict of Interest**

The authors declare no conflict of interest.

## **Author Contributions**

Aining Li completed the implementation of various mean calculation methods. Shiwei Zhang was responsible for the conception and framework design of the whole paper. Zhanbin Feng assisted in the implementation of various mean calculation methods. Qiyuan Zhang translated and proofread the thesis in English.

## Acknowledgment

This work is supported by the Teaching Reform Research Project of Laiwu Vocational and Technical College, the project number is 2019014. I would like to express my appreciation to Shiwei Zhang, Zhanbin Feng and Qiyuan Zhang, they have provided exciting insights to our creative process. I would also thank the authors of the references for their research has helped us a lot.

#### References

- [1] Zhang, S. W., Wang, Z. C., & Liu, C. (2012). Study on the methodology system of sports statistics data in Excel. *International Journal of Computer and Communication Engineering*, *1*(*1*), 16-18.
- [2] Yang, S. Y. (2004). Application sample of the statistical data and analysis. Beijing: Young Chinese Press.
- [3] Zhu, Q. W. (2017). Research on function extension method of Excel. *Journal of Information Systems Engineering, (12),* 153-156
- [4] Zhang, S. W., & Liu, C. (2010). Application research on statistical processing of the sports questionnaire data in Excel. *Proceeding of the 2010 Conference on Computer Science in Sports, Liverpool: World Academic Union* (pp. 272-276). Nanjing, Jiangsu, China.
- [5] Wang, Z. C., & Zhang, S. W. (2009). Research on the statistics of tested data of the national students' physical fitness and health standards with Excel. *Journal of Honghe University*, *7*(*2*), 90-92.
- [6] Wang, Z. J. (2019). Introduction of various average function in Excel. *International Journal of Computer Knowledge and Technology, (7),* 40-41.
- [7] Qian, X. F. (2010). Practical application of function nesting feature in Excel. *Journal of Human Resources Management, (4),* 136.
- [8] Bao, X. S., Liang, B., Zhou, H. Y., *et al.* (2009). Programming with VBA and Excel functions in data processing. *Computer Applications of Petroleum*, *64(4)*, 9-12.
- [9] Lv, H. S. (2008). Put the frequent numerical calculus into VBA practice of Excel. *Journal of Chaohu College*, *10(6)*, 6-11.
- [10] Fang, R. (2018). Application of VBA technology in Excel form data statistics. *Journal of Jiamusi Vocational College*, (8), 455-457.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (<u>CC BY 4.0</u>).



**Ai-ning Li** was born in Shandong Province of China on April 24, 1968. He received his bachelor of science degree from Qufu Normal University in 1994. He is a librarian of Laiwu Vocational and Technical College, majoring in library management.



**Shi-wei Zhang** was born in Shandong Province of China, on July 1, 1970. He received his master's degree from Beijing Sport University in 2003. He is a professor of Laiwu Vocational and Technical College. His research direction is sports statistics and measurement evaluation. He's a Jiusan Society, a member of the Chinese Society for Sports Science, the trustee of Shandong Province Higher Vocational Sports Committee. He's a member of Academic Committee and Professors Committee of Laiwu Vocational and Technical College. His research projects have won the Shandong soft science outstanding achievement award, the Laiwu social science outstanding achievement

award, and the Shandong staff education and vocational education outstanding research achievement award. He presided over the science and technology project of the Chinese Society for Sports Science, the sub-project of the key project of the ministry of education of the national education science "tenth five-year Plan", the research project of the teaching reform of vocational education in Shandong Province, and the scientific research projects of his college. He has published more than 20 academic papers in International Journal of Computer and Communication Engineering, Sports Science and Journal of Beijing Sport University, et al., among which 2 papers were indexed by ISTP.



**Zhan-bin Feng** was born in Shandong Province of China on January 15, 1969. He received his bachelor's degree in education from Shandong University of Physical Education in 1994. He is an associate professor of Shandong Polytechnic College. His research direction is physical education teaching and training.



**Qi-yuan Zhang** was born in Shandong Province of China on December 20, 1999. He is an undergraduate student of Grade 2018 Mechanical Design and Manufacturing and Automation, School of Mechanical Engineering, Sichuan University.