

# Study on the Methodology System of Sports Statistics Data in Excel

Shiwei Zhang, Zuocan Wang, and Cheng Liu

**Abstract—Objective:** In the practical activities of physical education, we often collect numerous data, how to fully exploit more valuable information that are hidden among the large amounts of data, we will need to choose reasonable method of data processings, that will be provided the basis for scientific judgment and analysis. **Method:** Based on the introducing of the sports statistics processing need statistics, parameters and its classification, analyzes the type of sports statistical data and processing requirements, The authors analyses sports statistics data types and processing requirements from statistical properties and mathematical statistics group needs. **Result:** The authors describes Excel methodology for sports statistical processing from the four aspects of data analysis: “the function method”, “the chart function equations to obtain statistics and functions”, “pivot tables and perspective”, and “Analysis Tool Pak”. **Conclusion:** Although Excel has not been grouped into statistical software, as long as we take the actual practise into consideration, fully tap the power of Excel statistical functions, improve methodology system of sports statistics data in Excel, enhance data exchange with other statistical software to realize resource sharing, further study and research Excel software development, we can achieve maximum results with little.

**Index Terms—**Excel, sports data, statistical processing, methodology system.

## I. PREFACE

In the practical activities such as physical education teaching, sports training, physical education administration and education physical scientific research, physical education workers often collect numerous data, which are valuable in sports science information for they usually contain a wealth of sports information, however, this information is often hidden among a large amounts of data and not clear. Therefore, to get more valuable information from these numerous and complex data and to fully tap the information reflected in the problem, we need statistical treatment of these data to make scientific judgments and analysis, which makes the method of statistical processing of data particularly important. With the increasing popularity of computer technology, it has become possible for physical education workers to use process and analysis data via

computers thought they are not computer experts. While processing the sports data, although there are some specialized and powerful statistical software such as Minitab, SPSS and SAS, there are also some problems such as high price, learning difficulties and difficulty of understanding statements and low utilization. Excel, though not grouped into statistical software, and not as professional as SPSS, SAS and other statistical software, it is a set of data tables, work function, VAB applications, and powerful report processing data in a statistical tool, which can not only deal with figures, but also with graph and character. For this reason, we need to further tap its statistical data processing functions, sum up and sort out its processing methodology, and describe the achievements of variety of methods in Excel.

## II. PARAMETERS AND STATISTICS IN PHYSICAL STATISTIC PROCESSING

How to simplify and represents a set of data we collected in sports practice, and how to estimate, inspect or forecast some unknown characteristics of samples abstracted from overall data, we need some statistical parameters and statistics to describe the set of data.

### A. Parameters

Also known as overall parameters, is a number that reflects an overall features and is a numerical descriptive data that for the study of overall things, that is, it is used to describe the overall figures, such as the population average, population variance, the overall standard deviation, usually represented with Greek letters.

### B. Statistics

Also known as estimate, is a descriptive statistics used to describe a number of attributes, such as the sample mean, sample variance, sample standard deviation and so on. It is calculated from the samples and usually represented in English. In general, the total of study subjects is mostly unlimited, even if some are generally limited, but its individual content is very large. Therefore, the parameter value on behalf of the general properties is usually not available. We usually obtain data from sample (statistics) to infer and estimate the overall parameter values. Although the parameters are often unknown, it is an objective value, a constant. Statistics are derived from the samples and can be seen as known. It is changed with the change of sampling, so it is random variables. In sports practices we often use statistics to replace general parameters.

### C. The Classification Parameters and Statistics

Statistically, to sort, analysis and reason a set of data, we

Manuscript received February 25, 2012; revised April 27, 2012.

Shiwei Zhang is with the Dept. of Physical Education, Laiwu Vocational and Technical College, Laiwu, Shandong, China, 271100 (e-mail: lwzsw@yahoo.com.cn, lwzyzsw@163.com, tel.: +86-015563419160; fax: +86-0634-6268137).

Zuocan Wang is with the Dept. of Physical Education, Linyi University, Linyi, Shandong, China, 276005.

Cheng Liu is with the Dept. of Educational Administration, Laiwu Vocational and Technical College, Laiwu, China, 271100.

often study the data central tendency and dispersion. Only by combining the focusing indicators and discrete indicators can it fully reflect the distribution characteristics. The central tendency, also known as "concentrated value", reflects the tendency of a group of data approaching a centre number and shows the location of a data centre where all numbers centre around it to show representative number. The statistics usually used are the average (arithmetic mean, including with the internal mean, geometric mean, weighted average, etc.), median, mode, etc., whose role is to measure the central tendency of frequency. Degree of dispersion is another main feature that describes the frequency of number variance and it is usually expressed by variation of indicators. Statics commonly used are range, a quarter potentiometer, percentage values, the average absolute deviation, variance, standard deviation, coefficient of variation and so on.

### III. TYPES AND PROCESSING REQUIREMENTS OF SPORTS STATISTICS DATA

Statistics, also known statistics data, is generally divided into groups of measurement data, count data and grade data from the view of the Statistics group, and it can also be divided into discrete variables and continuous variables from the perspective of mathematical statistics.

#### A. Classification and Processing of Statistics

Statistics, also known statistics data, is generally divided into groups of measurement data, count data and grade data from the view of the Statistics group, and it can also be divided into discrete variables and continuous variables from the perspective of mathematical statistics. Measurement data, also known as quantitative data, it is obtained by the measurement of certain observation unit. Most of this information is get by measuring observation units one by one via quantitative measurement, and usually with weights and other units of measurement. The measured value of each observation unit is also named as variable values. Count data or the attribute data is not a measurement tool but a method used to collect data, which classifies the observation unit by a certain attribute or category. Ranking data is to classify the observation units by certain attribute groups to varying degrees. These data and count data is different in that there are degree differences of grouping attributes while in ranking data the difference is the size and weight. The difference with the measurement data is that each observation unit is not exactly quantitative, but belongs to a class, which is also known as semi-metering information. Also, according to the analysis needs, measurement data, count data and level data can be transformed into each other.

Different types of data need different approach. In general, the preliminary statistical treatment of count data is seeking relatively numbers, and further statistical inference can calculate confidence intervals and  $\chi^2$  test and so on. Measurement data is usually seeking the average indicator and variance indicator, then takes various method in accordance with different research purposes and under different conditions, or to calculate the value of the normal range of parameter estimates (such as confidence interval estimation), or t-test, F test, etc. to make the right judgments, besides, with the relevant regression, multivariate analysis it

can make deep and detailed analysis to complex issues. But when the measurement data itself does not meet the conditions of application of parametric tests, you can use non-parametric test. As to the ranging data, we often use non-parametric test of the rank sum test.

#### B. Types and Processing Requirements If Data Variables

The data we collect have the generally three features of discreteness, variability and regularity. Because of the variability of the observed data, it is called variable in statistics. Variable is used to describe a certain feature. During the process of data collection, we need to gather all kinds of variables reflecting the characteristics of things. Variable, as the numerical expression of overall attribute, is the basic unit of data processing. Variable is usually divided into two categories: discrete variables and continuous variables. Discrete variable, also known as non-continuous variable or categorical variable, is only used as a code symbol to describe characteristic properties of certain things. Numbers on axis discrete and the numbers themselves are insignificant, without any size or multiple relationships, usually between 1-10. For some measure of attitude or a tendency to scale (or scale evaluation), the number is just a sequence of scales, but in order to facilitate research, it is often regarded as a continuous numerical data, and direct seeking its mean, standard deviation statistics, though not very reasonable, is also a last resort contingency measures. Values of continuous variables are the number of axes in a range of all values, which are unlimited in theory. A third value may exist between any two values. There is a relationship between size and multiples. Continuous variable is divided into packet data and non-packet data.

For the analysis of discrete variables, we usually just seek the number of allocation table, compare the number of frequency they appear (Mode), generally does not seek its mean, standard deviation and median, and sometimes need the chi-square statistics. For the analysis of continuous variable, there are three methods to deal with these data: the first is by descriptive statistics, directly calculate the various statistics quantities, such as all kinds of average, variance, standard deviation, and extreme value, median. Second is to calculate frequency distribution, sometimes it needs to change the continuous data to simple interrupted figures as an array to count the frequency. Third is to make Chi-square test, mean test, Z test, F test, T test, correlation and regression.

### IV. METHODOLOGY SYSTEM OF STATISTICS OF SPORTS DATA IN EXCEL

#### A. Function Method

There are many mathematical functions can be used to solve the problems of statistics and data processing. Excel provides a number of worksheet functions, a total of nine categories of hundreds of functions, including financial, date and time, math and trigonometric, statistical, lookup and reference, database, text, logic, and information. Not only can we use statistical functions to deal with some common statistics, such as all kinds of mean, variance, standard deviation, extreme values, median, chi-square value and related coefficient, but we can also use functions to query,

sort, and other logic operations. In addition, all the parameters distributions (such as normal distribution, F distribution and chi-square distribution, etc.) and test (T test, X2 test and Z test, etc.) require can be generally solved by the function. There are generally two function inputting ways. The first is to directly input the function in formula content from the keyboard. The second is to use "paste function" dialog box, and complete the calculation according to clew. While using the function we should pay attention to its meaning, scope and conditions of use, in order to use functions for fast and accurate statistical data processing.

### B. Using Charts Feature to Get Statistical Tables and Function Formulas

In order to make the data more intuitive in sports research, we need to draw a variety of charts and solving equations. By opening the "Insert" menu under the "Chart" option, in accordance with the Chart Wizard step prompts, according to the research need to select relevant content, then available we need charts, and the formula.

### C. Using the PivotTable in Excel Database to Make Cross-Analysis Table.

To explore the correlation between two or more variables, we can quickly summarize large amounts of data in data tables via pivot table and make cross analysis to identify the linkages between them. In the application of Pivot table for data summary, we should open the "Data" menu in the "PivotTable and PivotChart Report" function, pressing the "PivotTable and PivotChart Wizard" to finish step by step. 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, average, maximum, minimum value, product, value count, standard deviation, population standard deviation, variance, population variance and so on. If necessary, we can also choose the display order in a pivot table, which include ordinary display, the difference, percentage, a summary of the field according to certain field, representing the sum of the percentage of peer data, accounting data with the sum of the column the percentage of the total and index. While using the overall data in Pivot table, you can change the layout of the pivot table, mainly the exchange of field position and the change of the PivotTable field position. In the pivot table the content that doesn't no repeatedly appear will be classified into a category table to calculate cross-correlation statistics. If there is a low repetition rate of the date or number, it is very likely that each value is unique so that it is impossible to reduce its class of case. On this occasion, we can add a row in the original data table, use the IF function to packet the original data, and then make cross-analysis. Also, if you need to create formulas using data elements or other worksheet, you can create a calculated field or calculated item in the field.

### D. Using Excel Data Analysis Database.

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, 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 and you can use the "data analysis" for statistical processing. Before using the analysis tools you must organize data in a worksheet in accordance with rows or columns, which is Excel's input area. Using these tools requires us to be familiar with specific areas of statistical analysis. Analysis tool library contains information on: 1, analysis of variance (ANOVA, repeated two-factor analysis of variance, two-factor analysis of variance without repetition), 2, correlation coefficient, 3, covariance, 4, descriptive statistics, 5, exponential smoothing, 6, F-test: two-sample variance, 7, Fourier analysis, 8, histogram, 9, moving average, 10, a random generator, 11, rank and Percentile, 12, regression, 13, sampling, 14, t-test (paired two-sample analysis of the average, two-sample equal variance assumptions, double-sample Assuming Unequal Variances).

Although Excel has not been grouped into statistical software, and its function of statistics is not comparable with SPSS and SAS and other professional statistical software, in sports research, as long as we take the actual practise into consideration, fully tap the power of Excel statistical functions, improve methodology system of sports statistics data in Excel to increase its practicability and operability, enhance data exchange with other statistical software to realize resource sharing, further study and research Excel software, comprehensively improve its statistical functions, can we achieve maximum results with little effort.

### ACKNOWLEDGEMENTS

I would like to express my appreciation, first and foremost, to my postgraduate classmate Zuocan Wang and my colleague Cheng LIU, who had provided exciting insights to our creative process. Second to my colleague Xiuzhi He, who had helped us for translating paper into English. I would also thank the entire faculty and the exemplary staff in my department for all their assistance.

Thank you all for your hard work.

### REFERENCES

- [1] D. Zhou, *Sports Scientific Survey*. Beijing Sport University Press, 2001.
- [2] G. Qi, *Multiple Elements Analysis in Sports*. Beijing Sport University Press, 1998.
- [3] S. Yang, *Application Sample of the Statistical Data and Analysis*. Young Chinese Press, 2004.
- [4] S. Tang, "Research of applications of Excel statistical function," *Journal of Chengdu University of Information Technology*.2002, vol. 17, no. 4, pp. 255-260.
- [5] L. Pan, Y. Yao, and J. Zhang, *et al.*, "Realized statistical functions extension in Excel," *Journal of Chinese health statistics*.2007, vol. 24, no. 5, pp. 543-544.
- [6] Z. Wang and S. Zhang, "Research on the statistics of tested data of the National students' Physical Fitness and Health Standards with Excel," *Journal of Honghe University*.2009, vol. 7, no. 2, pp. 90-92.
- [7] S. Zhang and C. Liu, "Application research on statistical processing of the sports questionnaire data in Excel," in *Proceeding of the 2010 Conference on Computer Science in Sports*. Liverpool: World Academic Union. 2010, pp. 272-276.