

Challenges to Traditional Statistics in the Age of Big Data

Chen Ming¹, Zhu Xueshuai^{2,*}, Jiang Qian³

¹Hubei College of Chinese Medicine, Hubei, China, 434020

²China University of Mining and Technology, Beijing, China, 100083

³No. Six Middle School of Shashi, Jingzhou, Hubei, China, 434020

Keywords: Traditional statistics; Age of big data; Challenges; Statistical Research; Data analysis

Abstract: The arrival of the era of big data not only broadens the scope of statistical research and enriches the content of statistical research, but also brings great challenges to traditional statistical research. In order to adapt to the development of the era of big data, statistical research should timely adjust the working ideas of sampling survey, deeply understand data information, transform data docking and processing methods, and grasp the key points of data relationship analysis, transforming the focus of statistical work and strengthening analytical thinking, making big data and statistics organically combined, better serving statistics, and injecting vitality into the development of statistical research work.

1. Introduction

Statistics is a highly comprehensive subject, which widely exists in society production and daily life. Making good use of statistics will make our life more convenient. Reviewing the development process of statistics, it has been made continuous progress in actual life and application. With the rise of the era of big data, traditional statistics are also making new changes. The era of big data has injected fresh blood into traditional statistics, changed the inherent mode of thinking, and also brought a lot of challenges. Therefore, how to deal with this challenge and change traditional statistics should be given due attention.

2. The Impact of the age of big data on statistics

The arrival of the era of big data has changed our daily life. Nowadays, big data can be seen everywhere, people will use data to study problems. Big data is a collection of many data. Traditional statistics only rely on paper media to record data. Big data is recorded by advanced software in a limited time, which can operate big data in time and improve the efficiency of data processing. Statistics can help people deal with big data and provide theoretical guidance for big data. Only by fully utilizing the value of big data and rationally integrating resources can great value be created. With the advent of the era of big data, changes in traditional statistics have been reflected from the definition of samples to the thinking and technology of data analysis. It can be seen that the big data make us take more initiative in the use of data, which will promote the rapid development of traditional statistics.

2.1 Deepening the concept of samples and increasing data types

Sample is the basis of statistics. Traditional statistics is to investigate a group, in which a part of the sample is randomly selected for analysis and conclusion. In the era of big data, we omit the steps and use network data to collect, analyze and draw conclusions. This can avoid its error and make the data more accurate. Traditional statistics has a certain model, it has a fixed form of expression, its structure and standard is single. Big data includes not only structured data, but also unstructured data and semi-structured data, which makes its structure no longer single, showing a variety of characteristics. Moreover, large data can better store data for classification and recognition.

2.2 Extending in collect conceptual and increase in software

Traditional data collection is time-consuming, laborious, inefficient and costly. Large data can be recognized and processed in advance, then analyzed effectively, and finally stored. Big data can provide more storage space to identify valuable information. But we should pay attention to the lack of security and cost of big data, so we should reform statistics on the basis of avoiding its shortcomings. At present, traditional statistics analyses and processes data are through two ways. One is to construct the relationship between data by using models, and the other is to analyze data by using relevant software. The process is very complicated. Big data can simplify the process and combine statistical software with big data to simplify its excesses.

2.3 Changes in quantification and analytical thinking

Traditional data is structured data. The way to quantify data is quite mature, and it is relatively easy to get data results that can be analyzed directly. In the era of big data, unstructured data is the main problem. Computer academia has begun to develop technology to deal with unstructured data. It is an important research field to deal with unstructured data directly or quantify it into structured data from a statistical point of view. The traditional statistical inference analysis process is based on the distribution theory. Under the premise of probability assurance, the population is inferred. Generally, the population characteristics are inferred according to the characteristics of the samples. Whether the inference is correct depends on the quality of the samples. Now, the process is based on the actual distribution, according to the overall characteristics of probability judgment, at a static or dynamic point in time, the object of large data processing is the overall data, do not need to infer the overall characteristics based on the distribution theory, but to infer according to the calculation method.

2.4 Different sources of data

In traditional statistics, data are collected according to research purposes. Data sources are usually known, and it is easy to identify the identity of data providers or check them afterwards. But the source of big data is difficult to trace. Because the source of big data is generally information network system, it does not have a strong purpose, it is all the signals recorded by human, and it is difficult to identify the identity of the recorder. In the era of big data, it is particularly important to strive to build the second track of statistical data sources.

3. Challenges to statistics in the age of big data

There are many challenges to statistics in the era of big data, but they are mainly manifested in

the production, processing and application of data, as well as in the methods, contents and personnel training of statistics education. In statistics, the production of data is mainly through the relevant statistical departments, such as social security institutions, to carry out the corresponding statistical work to obtain statistical data. The data preprocessing methods are mainly data cleaning, data correction, data filling and so on. In the era of big data, the acquisition of data is realized by means of advanced instruments such as sensors, measuring instruments, etc. Data processing is realized by compiling databases through all kinds of software.

Traditional statistics still stays in the use of sampling technology to collect data in general, and then establish a model for statistical analysis of data. But with so much data today, traditional statistics face enormous challenges in completing this. If statistics wants to get rid of the old and bring forth the new, it must reform itself.

4. Countermeasures of traditional statistics in the age of big data

With the development of big data, many application areas are inextricably linked with big data. If statisticians are proficient in the processing of big data, sum up the data in time, and calculate the feasibility data, it will facilitate the improvement of quality of life. Therefore, it is necessary to strengthen the training of statisticians, invite the company's statistical professionals to give lectures, and guide the processing methods of corporate employees' big data. In the context of big data, statisticians are required to have stronger professional knowledge and operational ability. Statisticians should actively learn new knowledge and make better use of big data to create economic benefits for companies.

4.1 Reform the teaching method of statistics

Higher education institutions must follow the pace of the big data era and make reforms to the traditional teaching model. Introduce advanced teaching facilities in time, improve the statistical teaching environment, and teach statistical teaching methods. In the actual reform process, statistical teaching content should use big data as a case to analyze the development process of big data in the modern era and integrate big data into modern life. In the reform of the teaching of statistics, it is necessary to start from the curriculum and teaching, and reform the statistics in order to integrate the development of statistics and the era of big data. Now statistics must not only meet the challenges brought by the era of big data, but also the opportunity to seize the era of big data.

4.2 Change the function of sampling survey

For traditional statistics, sampling survey is one of the most important methods to collect data. However, sampling survey is not a comprehensive survey after all. It needs enough good samples to correctly reflect the overall characteristics, which has the shortcomings of instability and large error. In the era of big data, although network data has many advantages, it is not perfect. The bias of data is inevitable. When performing statistical analysis, there will be a lot of data that is not available. Therefore, the two can be complemented, the sampled data is corrected for Internet data, and the Internet data is supplemented by the sampled data. Secondly, the sample survey can be regarded as a clue to find rules or relationships from the mixed data, so as to better perform data mining and rapid detection and analysis.

4.3 Adopt new big data sorting and classification method

Traditional statistics are sorted and classified according to a preset scheme. The indicators and

the classifications obtained are structured. The sorting and classification of data is a necessary step for data preprocessing and is a component of statistical analysis. . However, for big data, because the source of the data, the form of the data, and the way the data is expressed are diversified, they can only be supplemented and improved according to the characteristics of the data itself after preprocessing the data. At this time, the traditional data grooming and classification methods are no longer practical. It is necessary to create and develop data grooming and classification methods adapted to the era of big data, and open up new data analysis paths accordingly.

4.4 Associate structured and unstructured data

Structured data and unstructured data are two forms of data in the era of big data. In modern society, various types of data have contributed to the era of big data. We need to improve the expression level of large data. Traditional statistics pays more attention to the inference of data, and the aggregation of large data is more inclined to the description of data, because the form of large data is diverse. Big data can describe data more completely. To collect data, extract effective information with useful value from unstructured data. Unstructured data is very abstract, and late development should focus on transforming unstructured data into structured data, making information more concise and clear.

5. Conclusions

In the era of big data, traditional statistics is not only deeply influenced, but also faced with challenges. It is required to change its understanding of samples, change its understanding of uncertainty, establish new methods of data sorting and classification, strengthen the connection between structured data and unstructured data, change the function of sampling survey, combine inductive deduction with inferential deduction, and combine statistical ideas. In short, we should firmly seize the opportunities brought by big data and organically combine big data with statistics in order to actively meet the challenges, and improve the adaptability of statistics to the era of big data, so that big data and statistics can be integrated and promote each other's development. Make full use of the opportunities brought to statistics by the era of big data, make statistics a tool to promote big data processing and combing, and make the era of big data a ladder to speed up the reform of statistics.

References

- [1] Huang Wenbin, Guo Mengnan. *Reflections on the Reform of Statistics Teaching in the Age of Big Data*[J]. *Modern Communication: Academic Edition*, 2017(19):16 0.
- [2] Davenport T H, Barth P, Bean R. *How big data is different* [J]. *MIT Sloan Management Review*, 2012, 53(5).
- [3] Zhu Jianping,Zhang Yuehan. *Thoughts on the Transformation of Traditional Statistics in the Age of Big Data* [J].*Statistical Research*, 2016.2(33)3-9