Research on the Development of Statistical Package for the Social Science (SPSS)

Xiaoling Wan[1, a], Yan Cheng[2, b, *]

[1]Shangrao Health School, Jiangxi, 334600, China

[2]Shangrao Preschool Education College, Jiangxi 334000, China

[a]wztgeducation@sina.com, [b]huhuiping@spup.edu.ph

*Corresponding author

Keywords: Statistics; SPSS; Social Science

Abstract: Statistics is a methodological subject based on probability theory. It has important applications in almost all fields. The theory of statistical knowledge is as indispensable as the computers we use every day. It is a necessary quality for modern senior talents. People without statistical knowledge in modern society are like "the-not-educated". Statistics plays an important role in human activities and is applied in many fields. This article tries to present the history and essential terms of statistics, specially of SPSS, to help clarify the theory and application of it.

1. Introduction to the history of statistics

Statistics is the study of how to collect, organize, analyze, and interpret numerical information from data. Statistics is the study of how we make sense of data. The history of statistical development can be traced back to the ancient primitive societies, which is with a history of more than 300 years. From the perspective of the generation and development process of statistics, it can be roughly divided into three periods: the beginning period, the ancient period, and the modern period.

1.The Beginning Period (mid-17th to 18th century)

The main schools of statistics were: School of State (the German H.Conring and G.Achenwall）; School of political arithmetic (the British W.Petty）.

School of State has made many contributions to the establishment and development of statistics:

（1）It is still a world-recognized term for this emerging discipline: "statistics" (statistics);

（2）Some terms are still used by statisticians, such as: "significant matters", "statistical data", "number comparison" and so on.

Main Contribution of the School of Political arithmetic to Statistics:

（1）It is not only satisfied with the quantitative registration, list, summary and description of social and economic phenomena, but also requires a comprehensive and systematic summary of these statistical experience, and to extract some theoretical principles.

（2）In terms of collecting data, the ideas of "a large number of observation methods", "typical investigation" and "regular investigation" are put forward.

（3）In terms of processing data, classification, tabulation and various indicators are widely used to concentrate and show the connotation information of quantitative data.

2. Ancient Period (late 18th century to 19th century)

Main school: School of Mathematical Statistics (P of French.S.Laplace and the Belgian A.Quetelet）; School of Social Statistics (K, German.G.A. Knies and C.L.E.Engel）.

Laplace was the first to introduce probability theory into the field of statistics, an astronomer, mathematician, and statistician whose contribution to statistics:（1）The law of large numbers of statistics is clarified. （2）Attempts to make a large-sample inference were made. It can be said that

Quetelet is the finisher of classical statistics, is also a pioneer of modern statistics, and is the founder of the school of mathematical statistics.

In the second half of the 19th century, the school of social statistics was initiated by the German university professor Knies, and represented by Engel et al.This school believes that the research object of statistics is social phenomena, aiming to clarify the internal connection and mutual relationship of social phenomena; statistics should include data collection, collation and analysis of its research.

3.Modern period (from the early 20th century to the current mathematical statistics period)

Statistics in the late 19th and early 20th centuries were mainly descriptive statistics; In the 1930s, by R.A.Inferred statistics created by Fisher has pushed mathematical statistics into the modern category.

After the 1960s, statistics have developed in three obvious trends:

With the development of mathematics, statistics rely and absorb mathematical methods.

To filtration into other disciplines, that is, statistics-based marginal disciplines are constantly formed.

With the increasingly extensive and deepening application of statistics, especially with the help of computers, statistics are becoming increasingly effective.

2. Statistical Package for the Social Science (SPSS)

SPSS is one of the world's famous statistical analysis software. After nearly 40 years of development, it has a large number of users around the world. At present, SPSS uses the window mode of Windows to display various methods of managing and analyzing data, which can be easily used for specific scientific research statistics. SPSS data editor and SPSS viewer are two windows of this software, in which there are many symbols to indicate some item (Figure 1).

1	A	α	alpha	a:lf
2	B	β	beta	bet
3	Γ	γ	gamma	ga:m
4	Δ	δ	delta	delt
5	E	ε	epsilon	ep`silon
6	Z	ζ	zeta	zat
7	H	η	eta	eit
8	Θ	θ	thet	θit
9	I	ι	iot	aiot
10	K	κ	kappa	kap
11	Λ	λ	lambda	lambd
12	M	μ	mu	mju
13	N	ν	nu	nju
14	Ξ	ξ	xi	ksi
15	O	ο	omicron	omik`ron
16	Π	π	pi	pai
17	P	ρ	rho	rou
18	Σ	σ	sigma	`sigma
19	T	τ	tau	tau
20	Y	υ	upsilon	jup`silon
21	Φ	φ	phi	fai
22	X	χ	chi	phai
23	Ψ	ψ	psi	psai
24	Ω	ω	omega	o`miga

Figure 1. The symbols and pronunciation

SPSS Data Editor is a work environment that integrates data input, data editing and transformation, data file organization, statistical analysis, and statistical charting functions. It has two interfaces: Data View and Variable View.

The SPSS result output window is named Viewer, which is a window for displaying and managing SPSS statistical analysis results, reports and graphics. Readers can save the content in this window in

the form of result file *.sav. The result output part is divided into left and right parts. The left part is the index output area, which is used to display the title and content index of the existing analysis results; the right part is the specific results of each analysis, which is called the detailed explanation output area. This is very similar to Word's document structure view. The output area is a view of the detailed output area, reflecting each content item in the detailed output area in a concise way, making it easy for users to find the operation results. The table in the detailed output area can be edited and other operations.

The creation of data files generally goes through two steps: 1). Define the structure of the data file, that is, define variables under the Variable View interface; 2). Input analysis data, that is, input various types of data under the Data View interface. With SPSS, researchers are able to be proficient in using arithmetic mean, harmonic mean, geometric mean, standard deviation and standard deviation coefficient formulas (Figure 2).

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N} \qquad \bar{X} = \frac{\sum_{i=1}^{N} X_i F_i}{\sum_{i=1}^{N} F_i} \qquad \bar{X} = \sum XF \Big/ \sum F = \sum X \frac{F}{\sum F} \qquad H = \frac{N}{\sum_{i=1}^{N} \frac{1}{X_i}} \qquad H = \frac{\sum_{i=1}^{K} M_i}{\sum_{i=1}^{K} \frac{M_i}{X_i}}$$

$$\bar{X}_G = \sqrt[N]{X_1 \cdot X_2 \dots X_N} = \sqrt[N]{\Pi X} \qquad \bar{X}_G = \sqrt[\sum_{i=1}^{K} F_i]{X_1^{F_1} \cdot X_2^{F_2} \dots X_K^{F_K}} = \sqrt[\sum_{i=1}^{K} F_i]{\prod_{i=1}^{K} X_i^{F_i}}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}} \qquad \sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \bar{X})^2 F_i}{\sum_{i=1}^{N} F_i}} \qquad V_\sigma = \frac{\sigma}{X} \times 100\%$$

Figure 2. The formulas in SPSS

3. The importance and practice of statistics

Statistics enable us to: characterize persons, objects, situations, and phenomena; explain relationships among variables; formulate objective assessments and comparisons; and, more importantly, make evidence-based decisions and predictions. There were two widely divergent influences on the early development of statistical methods. Statistics had a mother who was dedicated to keeping orderly records of government units (states and statistics come from the same Latin root status) and a gentlemanly gambling father who relied on mathematics to increase his skill at playing the odds in games of chance. The influence of the mother on the offspring, statistics, is represented by counting, measuring, describing, tabulating, ordering, and the taking of censuses-all of which led to modern descriptive statistics. From the influence of the father came modern inferential statistics, which is based squarely on theories of probability.

Describing collections involves tabulating, depicting and describing collections of data. these data may be quantitative such as measures of height, intelligence or grade level---variables that are characterized by an underlying continuum---or the data may represent qualitative variables, such as sex, college major or personality type.

Large masses of data must generally undergo a process of summation or reduction before they are comprehensible. Descriptive statistics is a tool for describing or summarizing or reducing to comprehensible form the properties of an otherwise unwieldy mass of data.

Inferential statistics is a formalized body of methods for solving another class of problems that present great of problems characteristically involves attempts to make predictions using a sample of observations. For example, a school superintendent wishes to determine the proportion of children in a large school system who come to school without breakfast, have been vaccinated for flu, or whatever. Having a little knowledge of statistics, the superintendent would know that it is unnecessary and

inefficient to question each child: the proportion for the sample of as few as 100 children. Thus, the purpose of inferential statistics is to predict or estimate characteristics of a population from a knowledge of the characteristics of only a sample of the population.

4. Summary

A knowledge of statistics can help anyone discriminate between fact and fiction in everyday life. For teachers, education statistics is a smart helper to do education research. Statistics helps in describing these measurements more precisely. Different policies of the education management are based on statistics. Statistical data are now widely used in taking all administrative education management decisions, such as schools, education bureaus or education research centers.

References

[1] Jordi Vallverdú. Ancient Statistics History in a Nutshell. Bayesians Versus Frequentists pp19-36. Springer Science. 07 November 2015.

[2] Yu Zhongyi. A Concise Outline of the Academic History of Statistics [J]. Statistical Research. 2009 (06).

[3] Stephen M. Stigler. Studies in the history of probability and statistics, L: Karl Pearson and the Rule of Three[J]. Biometrika. 2012,99(1).