The Unity of Social Statistics and Mathematical Statistics

Wang jianding

CHINA ECONOMIC PUBLISHING HOUSE BEIJING

The Unity of Social Statistics and Mathematical Statistics

Wang jianhui

CHINA ECONOMIC PUBLISHING HOUSE BEIJING

图书在版编目（CIP）数据

社会统计学与数理统计学的统一/王见定著
北京：中国经济出版社，2010.6
ISBN 978-7-5017-9692-2
Ⅰ.①社… Ⅱ.①王… Ⅲ.①社会统计—研究 ②数理统计—研究 Ⅳ.①C91-03 ②O212
中国版本图书馆 CIP 数据核字（2009）第 228242 号

英文翻译　吕丹丹
责任编辑　张玲玲
责任印制　张江虹
封面设计　北京然则创意设计公司

出版发行　中国经济出版社
印 刷 者　北京市人民文学印刷厂
经 销 者　各地新华书店
开　　本　710mm×1000mm　1/16
印　　张　12.75
字　　数　168 千字
版　　次　2010 年 6 月第 1 版
印　　次　2010 年 6 月第 1 次
印　　数　1-3000 册
书　　号　ISBN 978-7-5017-9692-2/C·42
定　　价　80.00（RMB）

中国经济出版社 网址 www.economyph.com 社址 北京市西城区百万庄北街 3 号 邮编 100037
本版图书如存在印装质量问题，请与本社发行中心联系调换（联系电话：010-68319116）

FOREWORD

It is recorded by the history of statistics that "political arithmetic" and "staatenkunde" date back to the 17th century—primary social statistics, originated in Britain and Germany. Almost at the same time, "gambling mathematics"—primary probability theory, appeared in Italy. Primary mathematical statistics did not come into being until the 19th century when law of large numbers and error theory were discovered and added to probability theory. So we can draw the conclusion that social statistics came into use two centuries earlier than mathematical statistics.

Because social statistics is widely used in economy and politics, the governments of all nations have been attaching great importance to it and made it develop systematically. After 1940s, mathematical statistics developed rapidly owing to the development of probability theory. After 400 years' development, two branches of learning have been formed—social statistics and mathematical statistics, both of which are in constant competition and neither of them can win easily.

After 30 years' research on these two subjects, I found that social

statistics and mathematical statistics are at once related and different. Their relationship is similar to that between Newton-mechanics and Einstein-mechanics.

Einstein-mechanics is used in the case near to velocity of light, while in most cases it is both accurate and convenient to apply Newton-mechanics when velocity of the moving body is far slower than that of light. Social statistics is used when we explain variable while mathematical statistics is applied when we describe random variable.

It is known that variable and random variable are both related and different. In the case of variable, when probability taking on each of its values is not the number of 1, variable will change into random variable; in the case of random variable, when probability taking on each of its values is 1, random variable will change into variable.

When knowing the connection and difference between variable and random variable, we know the relationship between social statistics and mathematical statistics. We can use social statistics to depict variable and mathematical statistics to describe random variable. Mathematical statistics will be overused to depict variable just like using a steam-hammer to crack nut.

Over 70 years, due to the rapid development of mathematical statistics, social statistics is in danger of being replaced by mathematical statistics. In some advanced countries, especially in the U. S. A., statistics is almost thought to be the same as mathematical

statistics. In fact, this is a misunderstanding. Through research, I found that mathematical statistics will never take the place of social statistics and both of them will co-exist and complement each other.

It is time for the battle between social statistics and mathematical statistics to come to an end. In order to further explain my viewpoint, I wrote this book for the reference of all the counterparts in the world.

This book sorts out and standardizes in a scientific way the development of the almost 400-year-old statistics and puts an end to the argument of over 100 years between social statistics and mathematical statistics. Since economic index is measured and analyzed through statistics, the unity of social statistics and mathematical statistics will certainly improve the analysis made by economics on an overall basis.

Note: Professor Wang Jianding is member of IASS (registration No. 4498)

Contents

Chapter 1 Data Collecting and Data Sorting

Statistics is a kind of science about data collecting, data sorting and data analyzing. So data is the base of statistics. Without data, statistical methods would be impossible. But where do data come from and how are they sorted after being obtained? This issue would be discussed in the following.

Section 1 Data Measure and Types

1. Data Quantitative Measure

Data quantitative measure is divided into four levels according to our level of precise measure of objective things.

(1) Fixed Type Scale

Fixed type scale, also called column names scale, is the most rough and lowest quantitative measure level. It is only a parallel

classification of objective things in accordance with a certain attribute. For example, it classifies the population into male and female by gender. In addition, it classifies enterprises into state-owned, private, foreign companies based on the nature of ownership. Fixed type scale measures simply the differences between categories. All categories are parallel relationships with no distinctions between good and bad or big and small. We have to fit the principles of exhaustive and mutually exclusive principles when using the method of fixed class scale. That is, when we classify all categories, we have to ensure that each individual is in a certain category, and only belongs to this one. For example, a person is either a male or a female, and can only belong to one gender.

(2) Ordinal Scale

Ordinal scale, known as the order of scale, is a kind of measure of order difference between things. For example, product rating is an ordinal measure of product quality. It rates products into first-class, second-class, third-class, and defective. Examination results can also be rated into excellent, good, fair and poor. Obviously, ordinal scale is more precise than fixed type scale. It shows the order between categories, but fails to show the exact differences. So it cannot do some basic mathematical operations such as addition, subtraction, multiplication and division.

(3) Interval Scale

Interval scale is also called inter-zone scale. It can not only

divide things into different categories, arrange their order, but also accurately measure out the gaps between them. For instance, we use hundred-mark system to measure test grades, degrees Celsius to measure temperatures and unit of meter to measure lengths. Therefore, the difference between two counts would be accurately pointed out only if given one unit of measure. For instance, the grade gap between 70 and 90 is 20; the temperature of one region is 10℃ higher than that of another area, etc. In other words, the data of interval scale can operate addition and subtraction.

(4) Fixed Ratio Scale

Fixed ratio scale is also named ratio scale. In addition of all features of the above three data measures, it possesses another trait, that is, the ability to calculate the ratio between the two measured values. This requires a fixed zero in the fixed ratio scale, which is the only difference between the fixed ratio scale and the interval scale. Like, a earns 6000 RMB monthly, while B earns 3000 RMB. The income of A is twice of B. So we can conclude fixed ratio scale can run multiplication and division besides addition and subtraction. Of course, number zero may appear in interval scale, but it simply means a zero level and does not mean no-existence.

2. Types of Data

Data can be divided into qualitative data and quantitative data

according to the results of the above four kinds of measure. Qualitative data, known as the quality of data, shows the quality characteristic of a phenomenon. It is the result of nominal scale measure and ordinal scale measure. However, quantitative data, noted as magnitude data, shows the quantity characteristic of a phenomenon. It is obtained from the result of interval scale measure and fixed ratio scale measure.

In statistics, the concept that shows some feature of a phenomenon is called a variable. And a concrete manifestation of this variable is called variable value. Variable is divided into quality variable and digital variable.

For example, "Gender" is a quality variable, for it presents itself as "male" or "female"; "Product Level" is also a quality variable, for it presents itself as "first-class product", "second-class product", "third-class product" and "defective". On the contrary, "output", "commodity sales revenue", "age" and "time" are all digital variables.

Digital variable is divided into discrete variables and continuous variables according to their different values. Values of discrete variables are discrete and tend be enumerated, whereas values of continuous variable are continuous, such as "temperature", "weight", etc. The variables in subsequent chapters are all digital variables.

Section 2 Data Collecting

Since we've been clear about data measure, the next question we are going to handle is where data come from. Originally data are from the direct investigation and experimentation.

However, from the user's point of view, data come from two channels: one channel is the investigation and experimentation on their own; the other is the investigation and experimentation of others. Generally speaking, the second channel is also very common.

Now we are going to talk about the methods of data collecting from the user's point of view.

1. Survey Plan

Before data collecting, a complete survey plan is required to guide the entire investigation and guarantee a smooth process.

A complete survey plan should include the following contents.

(1) Survey Purpose

First the purpose, mission and meaning of this investigation need to be explained by answering questions like why we conduct this investigation and what problem we would solve. Only after making clear these questions can we determine whom to survey, what to survey and in what ways.

For example, the purpose of the fourth population census in 1990 is to "make a thorough investigation of changes of the population of our country in aspects of amount, geographical distribution, composition and quality since the third national census. The investigation is expected to provide reliable material for making national economy and society development strategies scientifically, improving the people's material and cultural living standards and checking the actual performance of population policies".

(2) Survey Object and Survey Unit

Survey object is the range of the survey according to the survey purpose.

Survey unit is every unit making up of subject of the survey object.

In a practical survey, survey unit is either all units of one survey object or parts of it. For instance, in a census, all units of the survey object are survey unit. Whereas, in a sample survey, only part units of the survey object are survey units.

(3) Survey Project and Survey Questionnaire

Survey project is to answer "what to survey" and "what form to take".

Usually, a survey project presents with the form of a questionnaire. A questionnaire generally includes table header, table body and table legs. Table header shows the title of the questionnaire and the name, nature and affiliation of the investigated unit. Table

body shows the main contents. Table leg includes the signature of the informant and filling in date, etc.

2. Survey Methods

Common methods of a survey are census, sample survey, statistical report and so on.

(1) Census

A census is a one off comprehensive survey for a particular purpose, such as population census, industry census and agriculture census. Many countries conduct surveys regularly in order to hold basic data of their national conditions and power, which serves as a basis for governments to establish related policies. Census covers various aspects, involves many units and costs abundant manpower, material and financial resources and time.

Census holds the following traits: (a) Census has periodicity. Hold our China for example: population census is held every year with unit digit zero; industrial census is held every year with unit digit five; agricultural census is held every year with unit digit seven. (b) Census has a set unified investigation period in order to ensure the accuracy of results and avoid duplication and omission of data. (c) Census data is relatively accurate. They can provide basic references for other surveys. (d) Census commonly investigates some essential phenomena.