UDC 311
Biljana Popović
Prirodno-matematički fakultet
Niš

THE POTENTIALS AND LIMITATIONS OF STATISTICS AS A SCIENTIFIC METHOD OF INFERENCE *

Rezime

Statistics is a scientific method of inference based on a large number of data that show the so-called statistical homogeneity, regardless of the scientific field the data stem from. Its use is more prominent in sciences, which deal with experiments in the broadest sense of the term. The statistical method is a mathematical method of inference with implications of a completely developed formal system. The possibilities of applying statistics and the importance of thus obtained inferences are irrefutable; however, fallacies and exaggerations are present in the evaluations of such results. The reason for these phenomena, aside from misuse, is certainly insufficient knowledge of the possibilities, i.e. of the advantages of the scientific method in question. This paper presents attempts to stress the potentials of the method, as well as to underline its limitations, i.e. to reveal certain typical fallacies that can be encountered in practice.

Key words: *statistical inference, statistical homogeneity, sample, statistics.*

Introduction

Man has, perhaps, always desired to know what will happen tomorrow, with him personally in the first place, and then with his close ones, with the stock, with crops, orchards, forests, whether there will be a drought or a flood, and he has believed those to be known to God or gods. When he dared to "read" God's thoughts and intentions a 'scientific thought' in toady's meaning of that expression started to develop. He looked for reason, for consequential relationships, confident that, when those mechanisms are understood, he would be able to forecast the future.

One of the most authentic methods developed on the bases of the aforementioned ideas is the statistical inference. That is, from one generation to the next the experience was transferred that, for instance, if the land is prepared for sowing in "such and such" way the harvest will be "such and such". This knowledge was gained based on information exchanged between individuals, from generation to generation and inference of a statistical nature: a) observe an event or perform the experiment (e.g.

* Rad je nastao u okviru projekta Ministarstva za nauku i tehnologije, br. 1341.

sowing) many times under the same conditions independently of each other; b) infer based on those observations (observations that have a characteristic of statistical homogeneity).

Statistical observations create data that have a characteristic of changeability and uncertainty. Mathematical statistics is then used to infer from those data. "There is no statistics without data, neither there is statistics with only data" (Bartholomew [3]). Therefore, it is necessary to conduct the process of mathematical inference, which on the other hand must not be its own purpose: "Theorizing without data leads nowhere" (Bartholomew [3]). Porter [2] says: "Since the invention of the computer there has been no field of mathematics that has found as broad application" as mathematical statistics. Laplace has said that in the beginning there was probability, but the simple human mind reduced it to arithmetics.

Mathematical postulates are concentrated in the probability theory, which is a theoretical foundation of mathematical statistics.

First conscious attempts to define and apply statistical inference are believed to be censii conducted by rulers a few centuries BC in order to determine the number of military conscripts or taxpayers. The beginning of statistics as a scientific method is related to the emergence of the school of "political arithmeticians" in England in 17th century. According to some, "Natural and Political Observations upon the Bills of Mortality" by J. Grant published in 1622 represents the beginning of the statistics as a science. For a long time, the statistics was considered a scientific method for social studies. However, mathematicians who have inevitably been involved in the establishment, formal definition and became responsible for the development of the statistical inference, which is in essence a mathematical method, are responsible for the initial application of statistics in sciences. Among the early adopters of that idea was Galton, an English biologist who applied the statistical method to biology. The introductions of study of the correlation between events using mathematical statistics are contributed to him. His research led to the development of the method of correlation. The mathematical statistics experienced the full blossom in a theoretical sense, as well as in the spectrum of applications, with the development of the probability theory with its axiomatic finally shaped in the nineteen thirties of the previous century.

The Scientific Method

Every scientific method endeavors to achieve ever-greater determinism, so that the conclusions are unambiguous. Despite the endeavor, few scientific methods have such powers. Mathematical statistics is certainly not one of those methods. On the contrary, mathematical statistics never talks with certainty about anything! Therefore, classification into only two categories, like: yes-no, black-white, good-bad, is absolutely foreign to statistical inference.

What is the subject of mathematical statistics?

The subject of mathematical statistics is inference about characteristics or variables of the whole population based on information contained in a sample, giving at the same time a measure of the validity of the inference. Stages through which the inference is conducted are: defining the problem, which incorporates defining the population and one or more variables that should be observed on that population including the goal of the research; defining statistical methods that should lead to achieving the goal; sampling (defining the sample size and method of selection, as well as the selection itself); arranging data gathered from the sample according to the selected method of statistical inference; conducting the statistical inference; and interpretation of obtained results.

Fundamental statistical procedures can be classified according to the way of inference into two groups: estimation of parameters and testing statistical hypothesis.

Apart from this it is possible to make other classifications according to other criteria such as the kind of variable observed on a population; complexity of the method of inference and so on. The classification emhasized in [3], that should be presented here, is the one based on the complexity of inference. According to this classification, there are four types of statistical procedures.

Statistical procedures of the first type comprise simplest inferences performed at the level of arranging and presenting samples, in the first place by graphical means.

The second type relates to applying the statistics (functions of a sample with certain characteristics), of so called "small" samples, but also those that lead to the same type of inferences based on "large" samples. From the point of view of the potentials and limitations of statistics as a science, a special attention should be paid to the sense of "small" and "large" samples, since the sample size impacts the ability to infer on the selected statistics in the most direct way. It should be stressed here that there is no "universal" sample size, which would be large enough for all statistics. The sample size determines the speed of convergence of the selected statistics towards the real value of the parameter as well as the accuracy of the approximative estimation which is only possible upon a part of the total population – a sample.

Statistics of the third type follow dynamical stochastic systems (as opposed to those of the second type, which mainly concerned with static systems). Dynamical stochastic systems are by definition very complex systems comprising a large number of variables – characteristics that are often not well differentiated between. Dynamical systems can range from natural to production, to social. An example of a social dynamical system is, for example, the quality of life of a social group. The complexity of such a problem from the point of view of the statistical inference is obvious. The number of factors – variables that constitute this system; types of variables in the system (qualitative and quantitative variables); possibility of their measurement – direct or indirect; criteria, i.e. the range of their values that could be used to describe the quality of life, are only some of the questions that face the statistician on his way to understanding of this system. However, it should be kept in mind that the statistician is he who provides a good service to other scientists or simply to orderers of the service and that without those orders and a well defined project goal his services are not usable. This third type of statistical inference includes the so-called statistical modeling.

The fourth type of statistical inference comprises statistical decision– the decision theory. That is making decisions on all levels of management of social, production, even natural processes by means of the statistical methods.

The Potentials and Limitations

When potentials and limitations of a scientific method are discussed it is assumed that the method has been applied correctly, or properly in the first place, and then its limitations should be considered as well as the usability of obtained inferences. All these phases will be analyzed in the context of a statistical method. A special emphasis will be given to psychological research related to errors in applying the statistical inference.

First, let us consider conditions under which a statistical method can be applied. It has been stated already that the statistical method can be applied on all events that have a property of the so-called statistical homogeneity. That means that they must be equivalent to the statistical experiment. The statistical experiment meets the following criteria: 1) it can be repeated any number of times under the same conditions; 2) it is defined in advance what to observe in the experiment and all possible outcomes of the observation are known; 3) the outcome of any individual experiment is not known in advance. In this context we will, from now on, only talk about the statistical experiment, or experiment, regardless of whether it is a natural or perhaps a social event or even an artificially triggered experiment. It follows directly from such a definition that a large number of events those psychology and social sciences are interested in can be categorized as statistical experiments. Hence a large interest of social scientists and psychologists for this area of science. However, it should be kept in mind that if even a single element of the above definition is missing while observing an event, the statistical method is rendered useless. This is even more so since conditions of the definition are not always obvious, i.e. it is not always possible at first glance to determine if the observed event is a statistical experiment or not. Therefore, the applicability of the statistical method is questionable, i.e. conclusions that would eventually be drawn from its application.

By definition, a statistical experiment is conducted on a sample, rather than the total population. In order to be able to make conclusions about the total population based on a sample, it is necessary that this sample be a representative one. The importance of the selection of a representative sample is often illustrated by Fisz's example [1]:

The population comprised all employed health insurance policyholders in Poland, them $N = 2,757,131$ - the population size; and the observed variable was the type of work they did. All types of employment were grouped in four categories (a qualitative variable with four possible values) and the following table shows the absolute frequency of these categories ($N_i, i = 1,2,3,4$) in the total population of employees, as well as the absolute frequency of employment types in the sample ($n_i, i = 1,2,3,4$) of size $n = 230,433$; which has been selected from the total population based on the criterion that an employee's surname starts with the letter "P":

		N_i	n_i
1	laborers, except miners and steel mill workers	1,778,446	152,812
2	miners and steel mill workers	250,397	22,493
3	social workers and public servants	564,147	44,040
4	services	164,141	11,088

By comparing the distribution of the variable across the total population and the distribution of the same characteristic across the selected sample testing the hypothesis about the equality of those distributions, it was shown that the distribution of the variable on the sample was significantly different from the distribution across the total population. So, the sample was not representative, i.e. it was shown that the choice of profession in Poland was not independent of the staring letter of the surname.

A lot of psychological research has been dedicated to understanding the representative sample [4]. Let us list some of the most common misconceptions. Often, the validity and likelihood of a sample is estimated according to its similarity with the actual population. For example, in samples of six successive births in a hospital realized in the following way: MFFMFF and MMMMMF (M-male sex, F-female sex), the first is layman accepted as more representative than the second one. Applying this heuristics also leads to the following error. That is, those small samples represent the total population equally well as large samples. For example, it is erroneously concluded that 70% of obtained tails is an equally representative outcome for both the batch of 100 and the batch of 10 coin tosses. (However, it is easily calculated that the probability of obtaining 70 tails in a batch of 100 trials is approximately 0.000023, while the probability of obtaining 7 tails in a batch of 10 tosses is approximately 0.117187.)

Therefore, the choice of a sample is one of the phases of the statistical method of inference, which in a fundamental way determines the worthiness of inferences obtained using this method. The logical principle that "from a wrong premise an arbitrary conclusion (correct as well as incorrect) can be drawn" gets a practical confirmation in this case in the sense that from an erroneously selected sample it is possible to draw any, i.e. totally useless inferences. The importance of this phase of statistical inference is best illustrated by the fact that the sampling theory represents a special field of mathematical statistics.

Hereafter it will be assumed that the sample, when one is discussed, has been selected in an appropriate way.

A sample size is the particular question. Actually, the sample size is associated with the limit distribution of the appropriate statistics used according to the method of inference. So, only those statistics corresponding the available sample size should be taken into account, or we should choose the sample size satisfactory enough having decided in advance about the method of inference, i.e. the appliance of the certain statistics.

The third aspect concerns the question of chance. Not only that chance is often erroneously equated to probability, but also it is also unjustifiably applied to an individual case. In general, whenever an individual case is considered, statistical inferences should not be used unreservebly. This will be discussed in more details later on.

Let us discuss chance on the example of a claim of a pharmaceutical firm that probability of the antibiotic "A" being effective on the bacterium "B" is 0.9. The question is what is the chance of each of 30 patients infected by the bacterium "B" to be cured by the antibiotic "A". Based on the manufacturer's claim regarding the therapeutic values of the antibiotic it should be expected that if all patients were treated with it, 27 would be cured. However, apart from the fact that the medicine will help 27 out of 30 patients, there is no way to know in advance, who the 27 lucky ones will be. So, a patient can rightly believe that he or she will be cured based on the chance given by the manufacturer. From the point of view of the doctor who has successfully cured 27 patients in a row it does not mean that 28^{th} has a lesser chance to be cured than any of already cured patients individually, but neither has a greater chance also. It should be indicated here that psychological research shows that majority of people consider chance as a self-controlling process. For example, after a long sequence of heads in a batch of coin toss, most people believe that the next toss should be a tail. This is because such a sequence would be more representative for the experiment of coin toss than if the outcome were a head again. However, it is clear that in that particular toss it is equally probable to get a head as it is to get a tail, naturally, as long as the coin is not fake.

In this context we should also mention the idea of the so called "outcome orientation" [5] where people use a model of probability, i.e. statistical inferences, to make a decision of the type "yes-no" about an individual case rather than a series of events. For example, the weather forecast states that the probability of having rain in the next 10 days is 60%. If it indeed rained 6 out of 10 days for which the forecast had been made, the question is how worthy the forecast was. Many would conclude that the forecast was not good, since along all these 10 days the forecast has been given concerning raining with the probability of 0.6, it should rain for certain. It is the consequence of focusing at a single elementary event included in some event (experiment), not at the sequence of elementary events (experiments) having the probability of 0.6 to rain, assuming that it would rain in that 10 day's period. Accordingly, the forecast of 40% chances to rain means that in the given period it should not rain anyway. The forecast of 50% chances to rain in the next 10 days does not give any possibility to conclude if it will rain or not.

It is well known that statistics is concerned with comparative questions of real events to the supposed ideal model, ex. testing the variable distribution fitting to the normal distribution. The normal distribution is theoretically assumed ideal model, which is really nowhere realized completely. In other words, the distribution of the variable being considered on population can more or less resemble the normal distribution. The assumption of existence of ideal model, i.e. the exact measure lies at the basis of all methods for approximate calculation, and mathematical statistics as well. The fact that you can just approach to the ideal model with a certain accuracy, actually to the acceptable error, is one of the aspects under which the question of potentials and limitations of this scientific method should be discussed. Hence, the statistical hypothesis testing procedure assumes the possibility of accepting the hypothesis, which is actually wrong, i.e. rejecting the correct one. The risk of wrong conclusion in the statistical hypothesis testing procedure is measurable. The possible inference errors are the type I error and the type II error. Their measures are their probabilities.

Let us discuss the importance of these two types of errors in the light of the limitation aspects of the statistical method of inference. If the null hypothesis is accepted with the probability α of type I error, equal to 0.05 and some probability of type II error, β, it means that from 100 realized samples of the same size n, the 5% of them the most, i.e. at most 5 all together, will fall in the rejecting region if the null hypothesis is really true. Meanwhile, when testing a hypothesis, we are to accept the null hypothesis in the case that the one realized sample didn't fall into the rejecting region, i.e. believing that it was the one realized having the high probability (0,95) since such event is more probable.

The similar principle of confidence is applicated with the interval estimating of parameters, where such intervals are called the *confidence intervals* with the confidence level γ. The point is in *confidence* that the interval we got for the only one realized sample covers the real value of the estimated parameter.

Besides this formal limitative aspect of statistical testing inference there is another one, quite informal from mathematical point of view, but certainly essential from scientific point of view where statistical method is applied. We will discuss this method at one example. We can often find in the press the following remark: "Unsocialized youth behavior concerning its environment is more frequent than 20 years ago". If delinquency is followed, it could be easily established that it is getting worse with youngsters. However, is delinquency inadequate behavior or is it a reflection of adapting, and if it is – is it the only criteria of inadequacy, i.e. that argues the delinquency to be measure of inadequate behavior. Statistics is actually the scientific method which has to tilt with the fact that there are no two persons who could be expected to have the same attitude about the same concerning position.

*

The series of limitations having been discussed here cannot be the arguments for not to apply the statistical method of inference. On the contrary, each scientific method is getting more applicable if it is known better. The proper recognition of the item that could not be achieved by the certain scientific method, or where the method cannot be or must not be applied, can only encourage that method, especially if there is none better existing. Such situation is exactly with the mathematical statistics as the processing data method and the inference on the data basis about the items whose outcomes could be considered as random.

Finally, it could be pointed out that it was not our goal here to talk about the misuse of statistics, which means its deliberate wrong appliance, which is also worth considering for sure.

Conclusion

In the appliance of the statistical method of inference, there are objective and subjective limitations. In this paper, the analyzed aspects are certainly not the only ones and it should be taken into account when the appliances of statistical method are concerned. It is important, however, to conclude that at the today's level of scientific thought and knowledge about objective and subjective reality the statistical method is the irreplaceable one and the only one being concerned with data and their processing and the only one accepting the randomness as the objectively existing. Concerning the last it should be pointed out that in scientific world today – it is getting more talked about determined or deterministic chaos. In the case it is proved that the whole nature and society are arranged this way, mathematical statistics will definitely lose on its importance. For the deterministic chaos it is, however, important to discover the first element of a sequence. In that case, it should be thought of mathematical statistics as the available scientific procedure in the process of recognizing that first one in the sequence.

References

1. Fisz, M. (1968): Probability theory and mathematical statistics, New York.
2. Porter, T. (1986): The Rise of Statistical Thinking 1820-1900, Princeton:Princeton University Press.
3. Bartholomew, D.J. (1995): What is Statistics?, in: J.R.Statist.Soc., A, 158, Part 1, 1-20
4. Garfield, J. (1995): How Students Learn Statistics, in: Int. Stat. Review, 63, 1, 25-34
5. Konold, C. (1989): Informal Concepts of Probability, in: Cognition and Instruction, 6, 59-98

Biljana Popović

MOĆ I NEMOĆ STATISTIKE KAO NAUČNOG METODA ZAKLJUČIVANJA

Summary

Statistika je naučni metod izvođenja zaključaka na osnovu velikog broja podataka koji pokazuju tzv. statističku homogenost, bez obzira na naučnu oblast iz koje podaci potiču. Njena primena je im izraženija u naukama koje se bave eksperimentom i to u najširem značenju. Statistički metod je matematički metod zaključivanja pri čemu se podrazumeva potpuno izgrađen formalni sistem. Mogućnosti primene statistike i značaj tako dobijenih zaključaka su nesporni, međutim, zablude i preterivanja su veoma prisutni u vrednovanju tako dobijenih rezultata. Uzrok ovim pojavama je svakako nedovoljno poznavanje, izuzimajući zloupotrebu, mogućnosti, odnosno, moći naučnog metoda o kome je reč. U ovom radu su izlo'ena nastojanja da se istaknu mogućnosti ovog metoda, ali i da se istaknu ograničenja, tj. da se razotkriju neke tipične među brojnim zabludama koje se u praksi susreću.

Ključne reči: statistički metod zaključivanja, statistička homogenost, uzorak, statistika