Chapter 1

Introductory Chapter: Statistics

Türkmen Göksel

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.79773

1. Introduction

"Statistics is the grammar of Science," a famous quote by Karl Pearson who was the British statistician and leading founder of the modern field of statistics. Pearson highlights the importance of statistics and particularly emphasizes the significance of quantification for various fields of scientific study in his publication, The Grammar of Science [1].

Statistics is defined as the study of the collection, analysis, interpretation, presentation, and organization of data by the Oxford Dictionary of Statistical Terms [2]. Since the data grow faster than ever and information is increasing tremendously nowadays, the role of statistics becomes more crucial.

In general, we can group statistical analysis into two parts: (i) descriptive statistics and (ii) inferential (analytical) statistics. Descriptive statistics is used to summarize and/or describe a collection of data. Therefore, descriptive statistics provides a powerful way to summarize what already exists in data. However, inferential statistics focuses on the patterns in the data and then draws inferences from these patterns. In other words, by analyzing data gathered from samples (smaller subsets of the entire population), statistical methods infer about populations.

The field of statistics is the science of learning from data. In other words, statistics is the tool we use to convert data into information. Decisions based on data and information will provide better outcomes than those just based on intuition or gut feelings. In our daily life, there is almost no human activity where the application of statistics is not needed. Therefore, application of statistics plays a very significant role in almost every field such as Mathematics, Physics, Chemistry, Biology, Botany, Medicine, Economics, Education, Public Policy, Psychology, Astronomy, Zoology, Bio-Technology, Information Technology, Manufacturing, Service Industry, Business, and Commerce, among many other fields.

Since the application of statistics is very wide, different and multidisciplinary fields have evolved over time. These are some examples of application of statistics to other disciplines: Astrostatistics, Biostatistics, Econometrics, Business Analytics, Environmental Statistics, Statistical Mechanics, Statistical Physics, Actuarial Science, and so on. For example, Astrostatistics is the field which applies statistics to astronomical data which indicate that astrostatistics is a combination of astrophysics and statistical analysis. Biostatistics is the application of statistics to a wide range of topics in biology. Econometrics is the field where statistical tools are used to explain economic theories, and business analytics is the branch in which the statistical analysis applied to understanding of business performance and opportunities. Statistical physics is the branch which uses statistical methods to answer physical problems, and actuarial science is the field that uses statistical methods to analyze the risk insurance and some other financial issues.

To underline the importance of statistics in our daily life, we can look at the following examples: (i) *weather forecasting*: most of the people watch weather news and make decisions according to this news. Ineluctably, there are statistical models behind these forecasts that predict the weather conditions. (ii) *Insurance*: most of the people have some kind of insurance, such as medical, home, car, etc. Most of the insurance companies use statistical models to calculate the risk of giving insurance. (iii) *Medical field*: before any drugs prescribed, scientist should demonstrate a statistically valid rate of effectiveness. (iv) *Financial markets*: traders and businessmen use data and statistics to invest money. Of course one can easily extend the number of examples in which we use statistics in our daily life.

Among many others, some benefits of statistical analysis can be summarized as follows. First of all, it helps to present and compare the facts from data in a definite form. In other words, expressing results and/or conclusions in numbers develops a necessary and common form of communication for scientists, policy makers, and many others. Secondly, it helps us to formalize our thinking. In particular, statistical methods are used in formulating/testing the hypotheses or a new theory. By using these methods, we can determine the likelihood that a hypothesis should be either rejected or not. Thirdly, statistical methods help us to draw conclusions about populations based only on sample results. Last but not least, statistics is very important especially when it comes to the conclusion of the research, and in this sense statistical methods allow us for forecasting and policy making.

Among many benefits, of course, there are also some misuses of statistics. Main examples of misuses among others are overgeneralization, biased samples, insufficient sample size, and spurious correlations. All these mistakes may give us misleading conclusions. Let me just give some intuition about these misuses. (i) *Overgeneralization*: the results from one population may not be valid to another population. For example, a result/conclusion which is valid for one age group may not be true for other age groups. Hence, one should be careful about his conclusions when it comes to generalization. (ii) *Biased samples*: a random sample should be used, since a non-random sample can bias the results from the beginning. Making an argument or claim about an entire population based on a sample that is not representative of the whole is an important example of misuse of statistics. If one is analyzing whether or not the school lunch program provided enough food for students and conducts a survey only with

the basketball team then this would be a very good example of biased sample. The reason for this is that most likely a player on a basketball team burns more calories and eats more than an average student. In this sense, a basketball player is not representing the population very well which indicates that we have a biased sample. (iii) *Insufficient sample size*: when measuring a population, it may not be possible collect information from every member of that population. However, good news is that a "sample" can perform the job just as well. However, the important issue is to determine the right size for a sample to get accurate results. Using various methods, one can determine the right sample size which makes the data collection statistically significant. On the other hand, if the sample size is not sufficiently big enough, than the data may give us misleading conclusions. (iv) *Spurious correlations*: it does not necessarily mean that because two factors are correlated, one of these factors caused the variations in the other. Many statistical methods help us to analyze if the apparent relationships are meaningful and not simply chance occurrences.

Author details

Türkmen Göksel

Address all correspondence to: turkmen.goksel@ankara.edu.tr

Ankara University, Ankara, Turkey

References

[1] Pearson K. The Grammar of Science. London: Walter Scott/Dover Publications; 1892, 2004. ISBN: 0-486-49581-7

[2] Dodge Y. The Oxford Dictionary of Statistical Terms. New York, USA: Oxford University Press Inc.; 2006

IntechOpen