Applied Statistics: Basic Principles and Application

Alisa Bilal Zorić

Polytechnic Baltazar Zaprešić, Zaprešić, Croatia

Abstract: Statistics is an old scientific discipline, but its application has never been more topical. Statistics is a branch of mathematics that deals with the collection, analysis, interpretation and presentation of data. Increased computing power had a huge impact on the popularisation of the practice of statistical science. With new technologies such as the internet of things, we start to collect data from various sources like web server logs, online transaction records, tweet streams, social media, data from all kinds of sensors. With increased access to big data, there is a need for professionals with applied statistics knowledge who can visualize and analyze data, make sense of it, and use it to solve real complex problems. Applied statistics is the root of data analysis, and the practice of applied statistics involves analyzing data to help define and determine organizational needs. Today we can find applied statistics in various fields such as medicine, information technology, engineering, finance, marketing, accounting, business, etc. The goal of this paper is to clarify the applied statistics, its principles and to present its application in various fields.

Keywords: Applied Statistics, Statistics, Big Data

1. Introduction

Math is a big part of all of our lives especially statistics. Statistics is the branch of mathematics that deals with the collection, analysis, interpretation and representation of data. We use statistics to analyze what is happening in the world around us. It is a combination of methods which permit us to make reasonable optimal decisions in cases of uncertainty (Sachs, 2012). Statistics compare data through mean, median and mode. We live in the information age and most of the daily information is determined mathematically using statistics. With the help of statistics we can know what happened in the past and what may occur in the future. Properly analyzed data could be used to prevent disease, to collect important demographic information and test potential life-saving pharmaceutical products. It could be also used to increase the efficiency and profitability of an organization.

The goals of the paper are: (i) to present definition and principles of applied statistics; (ii) to elaborate similarities and differences between Descriptive and Interferential statistics, (iii) to give an overview of the most used statistical tools, (iv) to outline the benefits of proper statistical interpretation of data and (v) to point out various applications of statistics.

2. Statistics

Statistics is a brench of science which deals with: collection of data, organizing and sumerizing the data, analyses of data and making inferences, or decissions and predictions (Montgomery and Runger, 2010).
There are two major types of statistics: Descriptive Statistics and Inferential Statistics. Descriptive statistics utilize numerical and graphical methods to look for patterns in a data set, to summarize the information revealed in a data set, and to present the information in a

convenient form that individuals can use to make decisions. The main goal of descriptive statistics is to describe a data set. Thus, the class of descriptive statistics includes both numerical measures (e.g. the mean or the median) and graphical displays of data (e.g. pie charts or bar graphs).

Inferential statistics utilizes sample data to make estimates, decisions, predictions, or other generalizations about a larger set of data. The main goal of inferential statistics is to make a conclusion about a population based off of a sample of data from that population. One of the most commonly used inferential techniques is hypothesis testing. Statistical hypothesis is an educated guess about the relationship between two (or more) variables.

Table 1. Simililarities and difference between Descriptive and Inferential statistical problem (Singpurwalla, 2018)

Elements of a Descriptive Statistical Problem	Elements of an Inferential Statistical Problem
Define the population of interest	Define the population of interest
Select the variables that are going to be investigated	Select the variables that are going to be investigated
Select the tables, graphs, or numerical summary tools.	Select a sample of the population units
Identify patterns in the data.	Run the statistical test on the sample.
	Generalize the result to your population and draw conclusions.

Here are two most common examples of application of interferential statistics.

In the pharmaceutical industry, it is impossible to test a new medicine on every single person that may require it. Statistics can help them to create a sample of individuals and administer the medicine to them. After that, the statistician can analyze the effects of the drug on the sample and generalize their findings to the population.

In financial institutions, in order to measure credit card risk, managers often recruit statisticians to build statistical models that predict the chances a person will default on paying their bill. The manager can then apply the model to potential customers to determine their risk and that information can be used to decide whether or not to offer them a financial product.

We can also divide statistical analysis in three major parts: descriptive, moderate and advanced. Descriptive statistical analysis uses specific tools to describe data. These are relatively simple calculations that give a basic picture of what the data looks like overall. Descriptive tools include: frequency, percentages and measures of central tendency.

Moderate statistical analysis tools deal with the relationships between variables, what is the nature of these relationships and are they significant. These include correlation and regression. A correlation describes the relationship and strength of the relationship between two variables. Regression deals with predicting the impact of one variable on another. Like correlation, regression does not show causation.

Advanced analyzes include calculations of variance. They can help the researcher determine what diversity exists in the data, as well as the positive outcomes in the research. To calculate variance, the researcher must use the standard deviation. Standard deviation measures the degree to which an individual value varies from mean or average. Once the standard deviation is known, an analysis of variance can be performed. Analysis of variance or ANOVA is used to

compare the difference in the mean or average of the variable groups. This will show whether the outcome of one group is statistically different from the outcome of the other group. Covariance analysis or ANACOVA is a tool that can be used for experimental research designs. ANACOVA will detect discrepancies between the data before and after the test.

3. Principles of Applied Statitics

It is difficult to give a simple statement of the main principles of applied statistics because of its variety of applications. Statistical analysis of data is not a highly specific particular field of study. In the book Principles of applied statistics (2011), the authors pointed out the following principles as the most important:

- formulating and explaining specific research questions relevant to the subject
- creating solutions that provide a secure answer and open up new possibilities
- development of efficient and reliable measurement procedures
- development of analytical methods with suitable software following the primary research problem
- effective presentation and visualization of conclusions
- structured analysis to facilitate their interpretation in terms of subjects and their relationship to the knowledge base of the field

The main goal of applied statistics is to develop suitable concepts and methods that will help in solving the tasks listed above.

There are four main phases in statistical analysis. First, there is data collection and data manipulation in a form suitable for detailed analyses. There are a variety of software programs available that can help in organisation of the data well.Then, there is a preliminary analysis whose goal is to clarify the general form of the data and to suggest a direction in which analysis may go. Simple graphs and tables are used in this phase. The third phase is definitive analysis in which the basis for the conclusions is provided. The last phase is an accurate and concise presentation of the conclusion in a clear form (Doane and Seward, 2011).
Depending on the problem being solved, some of the phases can be omitted, and some can be repeated. For example, preliminary analyses can give clear results so that they can be considered definitive. Similarly, some analysis that was supposed to be definitive, may have huge unexpected discrepancies that require a reconsideration of the whole basis of the analysis (Rasch, Vardooren and Pilz, 2020).

Errors in analytical work are inevitable,so a complex precautionary system must be put in place to prevent errors and pitfalls for detecting them. An important aspect of quality control is the detection of random and systematic errors. This can be done by critically observing the performance of the analysis as a whole, as well as the instruments and operators involved in the process. For the detection itself, as well as for the quantification of errors, requires the processing of statistical data, as well as an experienced statistician.

4. Statistical tools

There are various tools that can analyze statistical data, some of them are simple, some complicated, and often very specific for certain purposes. Basic analyses can be easily computed, while more advanced methods require a solid understanding of advanced statistics as well as specialized computer software (FAO, 2021).

Statistical tools are useful in collecting, presentation and analysis of statistical data to make decisions. We can say that importance of statistical tools lies in the following reasons:

- helps to understand various economic and other problems and to find the right policy to solve such problems

- helps in converting raw data into summarized form so that it becomes easy to understand.
- helps in forecasting the future change based on the bast studies.
- makes decision making process easier by providing relevant information.

No matter which method of statistical analysis is chosen, it is important to be aware of each potential downside. The choice of method depends on the type of collected data as well as the type of problem to be solved (Calvello, 2020)

There are widely available computer packages for implementing statistical techniques in a relatively painless way. Statistical Softwares can look at current and past data to find trends and predict patterns, uncover hidden relationships between variables, visualize data interactions and identify important factors to answer even the most challenging questions and problems (Murray, 2019). Below are just some of the most well-known statistics software packages.

SPSS Statistics is the most widely used statistics software package for human behavior research. SPSS offers a robust set of features that allows easily compile descriptive statistics, parametric and non-parametric analyses, as well as graphical depictions of results through the graphical user interface (GUI). SPSS Statistics allows understanding large and complex data sets quickly and advanced statistical procedures ensure high accuracy and quality decision making (IBM, 2021).

R is an Open Source statistical software package that is widely used in various research fields. Toolboxes are available for a great range of applications, which can simplify various aspects of data processing. R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, …) and graphical techniques, and is highly extensible. It is a very powerful software, but it requires a certain degree of coding (R, 2021).

MatLab is a programming and numeric computing platform that is widely used by engineers and scientists to analyze complex data sets from diverse fields, develop algorithms and create models. It offers a massive amount of flexibility thru plentiful amount of toolboxes. As R, it requires a certain degree in coding (MatLab, 2021)

Microsoft Excel offers a wide variety of tools for data visualization and simple statistics. It's simple to generate summary metrics and customizable graphics and figures, making it a usable tool for many who want to see the basics of their data. There is a powerful and flexible Excel data analysis add-on (XLSTAT) that allows users to analyze, customize and share results within Microsoft Excel (XLSTAT, 2021).

SAS (Statistical Analysis Software) is a statistical analysis platform that offers options to use either the GUI, or to create scripts for more advanced analyses. It provides a comprehensive set of tools for both specialized and enterprisewide statistical needs – from analysis of variance and linear regression to Bayesian inference and high-performance model selection for massive data. Like R and MatLab, the coding can be a difficult adjustment for those not used to this approach (SAS/STAT, 2021).

The Minitab is powerful statistical software that offers a huge range of basic and fairly advanced statistical tools for data analysis. Commands can be executed through the GUI and scripted commands, making it accessible to novices as well as users looking to carry out more complex analyses (Minitab, 2021).

5. Application of statistics

Statistics is used in every aspect of life and statistical analyses are needed for almost every project. It is used to solve problems in different fields and we will mention some of them:

Health care: Statisticians participate in all stages of developing new medicine: discovery, testing, approval and marketing. Statistics is a tool that can be used to measure the spread of certain diseases and the pressure created by it on the health care system. It can be used to predict the number of future cases and how many beds, doctors, hospitals, etc., will be needed to control the situation. It can also predict the possible timeline of vaccine manufacturing and the number of deaths that can be incurred in the meantime. Statistical graphs can also be used to develop a strategy of containing the viruses by feeding real data in models for future projections

Government: Many national policies are decided using statistical methods, and administrative decisions are made on the basis of their data. Statistics provides the most accurate data, which helps the government to make budgets and estimate expenditures and revenues. It is also used to revise the pay scale of employees in case the cost of living is rising. Statistic surveys are very important becouse they can open big questions like salary discrepancies, disease clusters. Statistic is used to evaluate the effectiveness of fraud and crime strategies and tactics.

Finance: Statistics play a major role in the financial industry, especially in banking and investment. Banks use statistics to reduce risk in their loans, analyze the financial market and predict the impact of the economic crisis. Investors use statistics to understand the risks and potential of certain stocks.
Marketing: Another big application of statistics is in marketing especialy in social media and advertising. In social media, they analiye data to increase number of followers and in advertising, they analiye data to optimize campaign's performance.

Economics: There are so many concepts of economics that are completely dependent on statistics. All the data collected to find out the national income, employment, inflation, the theory of demand and supply, the relationship between exports and imports etc., are interpreted through it.

Politics: Statistics are crucial in a political campaign. It helps the politicians to have an idea about how many chances they have to win an election in a particular area. Statistic is used to target specific voter demographics and predict the winner of the election.

Data Science: A data scientist uses different statistical techniques to study the collected data, such as Classification, Hypothesis testing, Regression, Time series analysis, and much more. Statistics is one of the helpful measures for data scientists to obtain the relevant outputs of the sample space.

Robotics: Statistics is an important parameter that is used in robotics. The robot always senses the present state by estimating the probability density function value. With the help of new input sensories, the robots continuously update themselves and give priority to the current actions. With the application of statistics in computer science and machine learning, algorithms' efficiency can be increased significantly.

It is difficult to measure the mass, size, large distances (the distance between sun and earth, or moon and earth), density of objects in the universe without any error, but statistics formulas and methods do this with the best probability.

Research: Statistics is essential for all sections of science, as it is amazingly beneficial for decision making and examining the correctness of the choices that one has made. The uses of statistics play an essential role in the work of researchers. It is highly used to design surveys and present results in research papers.

There are many other uses of statistics in different fields related to daily human activities such as weather forecasting and sports.

We can sumerise that applied statistic is crusial to solve problems in varous fields such as engineering, law, medicine, finance, busines, etc. It has to develop statistical model using different mathematical and statistical theories and techniques, alalyze and interpret data and report conclusions. It seems plain and simple, but every word in previous sentence requires a lot of knowledge and experience.

The importance of data, data analysis, and applied statistics is relevant to nearly every area of our lives. As the field of applied statistics continues to evolve, those qualified to lead organizations and governments with insight gleaned from data will make a significant impact on the lives of generations to come (Michigan Tech, 2021).

6. Conclusion

In the information age, there is no shortage of data - there are too much data. The key is to filter the vast amount of data available and interpret its implications correctly. To achieve this, appropriate statistical data analysis tools are needed, as well as expertise in statistics.

Statistic is the science of data, it involves collecting, classifying, summarizing, organizing, analyzing, and interpreting numerical information. Applied statistics is a collection of applicable statistical methods and the application of these methods. Without the use of statistics it would be very difficult to make decisions based on any data collected. Like any other tool, statistics can be used or misused. It is important to know some of the basic statistical concepts, in order to be in a better position to evaluate the information you have been given.

Statistics are important for us as a society as it allows us to make well-balanced decisions based on researched facts. It allows us as consumers, politicians and corporations to plan our future direction and stops us from making major mistakes (Dilard, 2019).

The growing status of statisticians and analysts is the result of the explosive growth of information that needs scientific processing and analysis, so statistics will continue to be an area of active research, for example on the problem of how to analyze big data.

In this paper, we wanted to point out the importance of statistics and its application in almost every aspect of our lives. Although, there are numerous statistical software packages for analysing data, we need to know the basics of statistics to be able to correctly interpret the data we encounter on a daily basis.

References

- Cox, D. R. and Donnelly, C. A. (2011). Principles of applied statistics. Cambridge University Press. *Crossref*
- Doane, D. P. and Seward, L. W. (2011). Applied statistics in business and economics. New York, NY: McGraw-Hill/Irwin,.
- Dillard, J. (2019). 5 Most Important Methods For Statistical Data Analysis. https://www.bigskyassociates.com/blog/bid/356764/5-Most-Important-Methods-For-Statistical-Data-Analysis
- Rasch, D., Verdooren, R. and Pilz, J. (2020). Applied statistics. John Wiley.
- Glaz, B., & Yeater, K. M. (2020). Applied statistics in agricultural, biological, and environmental sciences (Vol. 172). John Wiley & Sons.
- Sachs, L. (2012). Applied statistics: a handbook of techniques. Springer Science & Business Media.
- Singpurwalla, D. (2018). A handbook of statistics: An overview of statistical methods.
- IBM, 2021, IBM SPSS Statistics, https://www.ibm.com/products/spss-statistics

- R, 2021, The R Project for Statistical Computing, https://www.r-project.org/
- MatLab, 2021, Math. Graphics. Programming. https://ch.mathworks.com/products/matlab.html?s_tid=hp_products_matlab
- XLSTAT, 2021, The Leading Data Analysis And Statistical Solution For Microsoft Excel, https://www.xlstat.com/en/
- SAS/STAT, 2021, State-of-the-art statistical analysis software for making sound decisions. https://www.sas.com/en_us/software/stat.html
- Minitab, 2021, Minitab, https://www.minitab.com/en-us/products/minitab/
- Murray, B. (2019) Statistical Analysis - Which Test To Use https://imotions.com/blog/statistical-analysis/.
- Montgomery, D. C., and Runger, G. C. (2010). Applied statistics and probability for engineers. John Wiley & Sons.
- Motulsky, H. J. (2015). Common misconceptions about data analysis and statistics. Pharmacology research & perspectives, 3(1),
- FAO, 2021, Basic Statistical Tools. http://www.fao.org/3/W7295E/w7295e08.htm#TopOfPage
- Calvello, M. (2020). Statistical Analysis Methods That Take Data to the Next Level. https://learn.g2.com/statistical-analysis-methods
- Michigan Tech, (2021). Importance of applied statistic in daily lives. https://onlinedegrees.mtu.edu/news/every-number-counts-importance-applied-statistics-our-daily-lives-infographi