

Segment: Descriptive Statistics

Topic: Introduction to Data Analysis

Table of Contents

1. Responsibilities of a Data Analyst	4
2. Statistics Domain.....	4
2.1 Descriptive Statistics	4
2.2 Inferential Statistics	5
2.2.1 Types of Inferences	6
3. Summary	7
4. Glossary.....	7



Introduction

Data analysis and making decisions is inevitable for all business entities. Data analysis provides a range of statistical tools for organising, presenting, analysing, and interpreting data. There is an upsurge in demand for data analysts throughout the world. Data analysts are now able to collect tremendous amounts of data and data analysis had become relatively easier with the usage of new world technology.



Learning Objectives

At the end of this topic, you will be able to:

- List out the common responsibilities of a data analyst
- Distinguish between descriptive statistics and inferential statistics
- Explain the types of inferences

1. Responsibilities of a Data Analyst

Data analyst acts as a facilitator between an organisation and its stakeholders in summarising, analysing and presenting the data in a more meaningful manner. Every industry from retail to healthcare takes insights from the data analysts. In an increasingly competitive business environment, data analysts with an understanding of statistical analysis techniques can make better business decisions and possibly gain a competitive advantage.

A data analyst may need to perform data analysis for many different purposes. The most common ones are to:

- describe a data set using the results of an analysis
- make inferences from the results of an analysis
- estimate unknown quantities from appropriate sample data
- test hypotheses about unknown quantities
- quantify relationships among variables
- make decisions based on statistical results.

2. Classification of Statistical Analysis

Data Analysts from various domains make use of various tools to describe data as well as to draw relationships among the variables. All these methods and tools fall into two domains under Statistics, namely, Descriptive Statistics and Inferential Statistics.

1.1 Descriptive Statistics

It deals with describing the data using graphs, tables, and various summary measures such as measures of central tendency and measures of variation. Descriptive statistics provide general insights about a problem or a situation at hand by describing the data. Irrespective of the size of the data, data analysts can summarise and present the data to the clients or audience. Measures of central tendency comprise tools such as mean, median, mode, etc., whereas, the measures of variation involve range, variance, standard deviation, etc.

For example, assume that a manager of a retail chain wants to have insights about the sales data of various products under his organisation. He can make use of various charts and measures

Introduction to Data Analysis

such as mean and standard deviation to get better inputs about the products sold by the organisation.

1.2 Inferential Statistics

It speaks about drawing some meaningful inferences from the data after performing suitable statistical tests. In other words, inferential statistics allow the data analyst to make meaningful predictions from the data.

Suppose we wish to learn about Fortune 500 companies. These 500 companies become the population that we need to collect data about. To collect the data, we could look at published data in libraries or on the web, mail questionnaires to company managers, or conduct face to face or telephone interviews.

Suppose we wish to learn about all the companies in Europe. This population is too large to collect data about. We may then resort to sampling. The population is a set of all the elements under a study of interest or research, whereas the sample is a subset derived from a population.

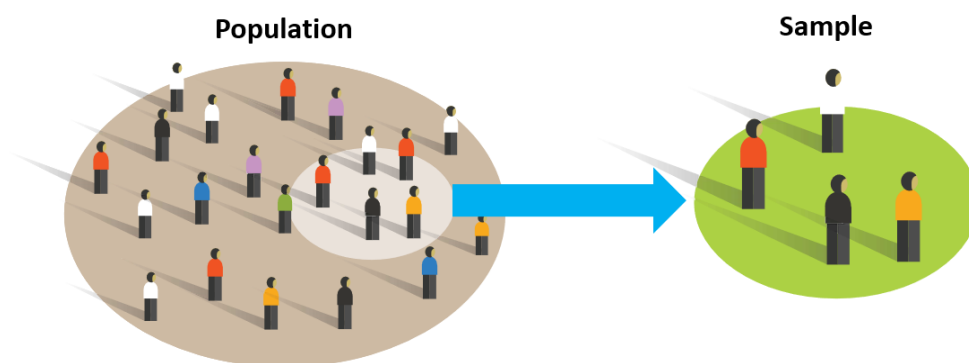


Fig. 1: Population and Sample

Using the appropriate computed results from the sample data it is possible to infer the parameters of the population. By designing the sampling experiment suitably, we can make the inference as reliable as needed.

Sometimes it may be difficult to define the population. For example, a store manager may be interested in knowing about the characteristics of all potential customers. The population should then include anyone who is a potential customer. However, it is hard to know who exactly those people are. In such cases, one has to make some assumptions and proceed. The store manager

Introduction to Data Analysis

may randomly pick customers entering the store and treat them as a representative sample of the population.

In another example, an inspector may be interested in the average weight of all the parts to be produced by a machine. Since all the parts have not been produced yet, the population is not there yet. The inspector may randomly pick parts being produced by the machine and treat them as a representative sample of the population.

1.2.1 Types of Inferences

It is possible to make two types of inferences from such data, namely, estimation and hypothesis.

Estimation

When the inferences we make about the population parameters are about their exact values, then the process is called estimation.

For example, we might estimate the sales revenue of a company to be US\$3,284,500 for the coming year.

Hypothesis

Some inferences made regarding population parameters are not about their exact values.

For example, we may be interested in simply testing whether a parameter is not less than 100, because someone has claimed that it is not less than 100 and a few others have challenged that claim. This process is called hypothesis testing.

Sometimes a hypothesis might be made about two or more population parameters. Such a hypothesis can also be tested using sample data.

2. Summary

Here is a quick recap of what we have learnt so far:

- Data analysis provides a range of statistical tools for organising, presenting, analysing, and interpreting data.
- Descriptive statistics deals with describing the data using graphs, tables, and various summary measures such as measures of central tendency and measures of variation.
- Inferential statistics speaks about drawing some meaningful inferences from the data after performing suitable statistical tests.
- The population is a set of all the elements under a study of interest or research whereas the sample is a subset derived from a population.
- It is possible to make two types of inferences from data, namely, estimation and hypothesis.
- When inferences are made about the exact values of the population parameters, then this process is called estimation.
- In hypothesis testing, the inferences are made about the population parameters which are not about their exact values.

3. Glossary

Sampling	The process of selecting a subset of population with a view to study the characteristics of that population.
Estimation	The process of inferring the value of an unknown parameter using sampling.
Hypothesis testing	The process of testing whether a statement regarding a population parameter is true using sampling.
Descriptive statistics	Numerical values are computed and used to describe the values in a data set.
Inferential Statistics	Drawing meaningful predictions from the data by using estimation and hypothesis testing.