A Brief Description of Data Analysis in Statistics

George Grill*

Department of Applied Mathematics and Computer Science, State University of Campinas, Sao Paulo, Brazil

Commentary

*For Correspondence:

George Grill,
Department of Applied Mathematics and Computer Science, Technical State University of Campinas, Sao Paulo, Brazil
E-mail: Grillgeo@ufabc.edu.br

ABOUT THE STUDY

Statistics is the branch of mathematics that deals with the collecting, organising, analysing, interpreting, and presenting of data. When using statistics to a scientific, industrial, or social problem, it is typical to proceed with a statistical population or model to be studied. "All persons living in a country" or "every atom in a crystal" are examples of populations, and so are "all individuals living in a country." Statistics is concerned with all elements of data, including survey and experiment design and data collecting planning.

When census data isn't accessible, statisticians create their own experimental designs and survey samples. Inferences and conclusions derived from a sample can be generalised to the full population with representative sampling. An experimental study is collecting measurements of the system under investigation, altering it, and then taking other measurements using the same process to see if the alteration has changed the measurements' results. An observational research, on the other hand, does not require any experimental changes.

In data analysis, there are two primary statistical methods: descriptive statistics, which use indices like the mean or standard deviation to describe data from a sample, and inferential statistics, which derive inferences from data that is susceptible to random fluctuation (e.g., observational errors, sampling variation). Descriptive statistics are most often concerned with two properties of a distribution (sample or population): central tendency (or location) seeks to

characterise the distribution's central or typical value, whereas dispersion (or variability) seeks to characterize the extent to which members of the distribution depart from its centre and each other. Probability theory, which deals with the examination of random processes, is used to make inferences based on mathematical statistics.

A standard statistical technique is collecting information in order to assess the link between two statistical data sets, or between a data set and synthetic data produced from an idealised model. The statistical link between the two data sets is hypothesised, and it is contrasted to an idealised null hypothesis of no association between the two data sets. Statistical tests that quantify the sense in which the null hypothesis may be proven inaccurate based upon the data presented in the test are used to reject or disprove the null hypothesis.

When working with a null hypothesis, there are two types of errors to be aware of: Type I errors (incorrectly rejecting the null hypothesis, resulting in a "false positive") and Type II errors (incorrectly rejecting the null hypothesis, resulting in a "false negative") are two types of errors (null hypothesis fails to be rejected and an actual relationship between populations is missed giving a "false negative").This paradigm has been linked to a number of issues, ranging from getting a suitable sample size to defining an appropriate null hypothesis. Errors can also be found in the measurement techniques that create statistical data. Many of these errors are characterized as random (noise) or systematic (bias), although other sorts of errors (for example, blunder, when an analyst reports erroneous units) can also occur. Limited data or censoring can cause estimates to be skewed, and particular strategies have been developed to overcome these issues.

Statistics is a mathematical study or a mathematical body of knowledge concerned with the collecting, analysis, interpretation, and presentation of data. Statistics is sometimes considered a separate mathematical discipline rather than a subject of mathematics. While data is used in many scientific projects, statistics is concerned with using data in the context of uncertainty and making decisions in the face of ambiguity. It is usual practice to start with a population or process to be examined when applying statistics to a problem. Populations can mean to anything from "all individuals living in a country" to "every atom that makes up a crystal." In an ideal world, statisticians would collect data on the entire population (an operation called census). Federal statistical institutions may be in charge of this. To summarize population data, descriptive statistics might be employed. Numerical variables such as mean and standard deviation are ideal for continuous data (such as income), whereas frequency and percentage are preferable for categorical data.

When a census is not possible, a sample of the population is examined. Data is collected for the sample members in an observational or experimental environment once a sample that is representative of the population has been selected. To describe the sample data, descriptive statistics might be employed once again. However, because of sample is drawn at random, the numerical characteristics derived from it are also subject to inaccuracy.