Statistics

Before we jump right into the article let's look at the list of the things we are about to discuss:

- What is Statistics?
- Types of statistics
- Why Learn Statistics?

What is Statistics?

Statistics is a mathematical science including methods of collecting, organizing and analyzing data in such a way that meaningful conclusions can be drawn from them. In general, its investigations and analyses fall into two broad categories called descriptive and inferential statistics.

Types of statistics

1. Descriptive statistics:

It deals with the processing of data without attempting to draw any inferences from it. The characteristics of the data are described in simple terms. Events that are dealt with include everyday happenings such as accidents, prices of goods, business, incomes, epidemics, sports data, population data.

Descriptive statistics may also cover graphical methods that can be used to visualize samples of data. Charts and graphics can provide a useful qualitative understanding of both the shape or distribution of observations as well as how variables may relate to each other.

2. Inferential statistics:

It is a scientific discipline that uses mathematical tools to make forecasts and projections by analysing the given data. This is of use to people employed in such fields as engineering, economics, biology, the social sciences, business, agriculture and communications.

There are many examples of inferential statistical methods given the range of hypotheses we may assume and the constraints we may impose on the data in order to increase the power or likelihood that the finding of the test is correct.

How is statistics different from mathematics?

- Mathematics deals with the exact numerical values whereas in statistics an approximation is used.
- However, you might contradict that even in mathematics, the approximation is used; it is true, but it is only used to reduce the complexity of the calculation of the numerics(rounding off a number), but the numerical elements are not approximated.

Why Learn Statistics?

Raw observations alone are data, but they are not information or knowledge.

Instead, data raises questions, such as:

- What is the most common or expected observation?
- What are the limits on the observations?
- What does the data look like?

Although they appear simple, we must answer these questions to turn raw observations into information that we can use and share.
Beyond raw data, we may design experiments to collect observations. From these experimental results, we may have more sophisticated questions, such as:

- What variables are most relevant?
- What is the difference in the outcome between two experiments?
- Are the differences real or the result of noise in the data?

Questions of this type are important. The results matter to the project, to stakeholders, and effective decision making.

Statistical methods are required to find answers to the questions that we have about data.
We can see that to both understand the data used to train a machine learning model and to interpret the results of testing different machine learning models, that statistical methods are required.

This is just the tip of the iceberg, as each step in a predictive modelling project will require the use of a statistical method.