INTRODUCTION TO INFERENTIAL STATISTICS

Exploratory Data Analysis helps you understand how to **discover patterns in data** using various techniques and approaches. EDA is one of the most important parts of the data analysis process. It is also the part on which data analysts spend most of their time.

However, sometimes, you may require a very large amount of data for your analysis which may need too much time and resources to acquire. In such situations, **you are forced to work with a smaller sample of the data**, instead of having the entire data to work with.

Situations like these arise all the time at big companies like Amazon. For example, say the Amazon QC department wants to know what proportion of the products in its warehouses are defective. Instead of going through all its products (which would be a lot!), the Amazon QC team can just check a small sample of 1,000 products and then find, for this sample, the defect rate (i.e., the proportion of defective products). Then, based on this sample's defect rate, the team can "infer" what the defect rate is for all the products in the warehouses.

---

*This process of **inferring** insights from sample data to the entire population is called **Inferential Statistics***

Even after using inferential statistics, you will arrive at only an estimate of the population data from the sample data, not the exact values. This is because when you don't have the exact data, you can only make reasonable estimates about it with a limited level of certainty. Therefore, when certainty is limited, we talk in terms of **probability.**

In the next topic, you will learn the basic concepts of probability and the various rules associated with it. The topics covers the following:

- Permutation and combination
- Definition of Probability and its properties
- Key terms related to Probability
- Probability rules (Addition and Multiplication)

Reference e-book:

Statistical Inference for Data Science by Brian Caffo