Course Overview

Basic Statistics

Overview: We will start with getting familiar with the statistics jargon and we address the need for statistical tools and techniques. Then we discuss the statistical tools commonly used for data science while we cover descriptive statistics in detail which includes 5-point summary stats, area-under-the-curve, Uni-variate, and bi-variate analysis, etc. and finally we explore some functions and tools available in python to perform statistical analysis.

How statistics plays an important role in Data Science & Easier approach to Distributions

- Getting familiar with the terms, "Distribution", "Process", "mean", "Standard Deviation", "Normal Distribution", "Population", "Sample" and "Model"
- Need for statistical tools

Important statistical tests you need to know and their uses.

- Need for Z-scores
- Meaning of Null and Alternate Hypothesis
- Need for statistical tools and techniques

Introduction to Data Science and modeling and why it's important?

- What does it mean to model a process?

Descriptive statistics: Variance, Quartiles, and more!

- Variance/Spread of the data
- Standard deviation
- Range
- Quartiles

How to Find Area under the curve of a normally distributed data

- Area-under-the-curve and it's relation to Probability

Concept of central values - Mean, Median and Mode with examples

- The measure of central values
- Central values and spread

Distributions and characteristics of a well -defined the stable process

- Normal distribution with an example

Descriptive Statistics: Visualising your Data using Pairplot Analysis using cars-mpg dataset

- Pairplot analysis in-depth

Descriptive Statistics: Hands-on approach (Includes notebooks)

- Using pandas' .describe() function to obtain the descriptive statistics of individual columns
- Analyzing individual attributes using their respective descriptive statistics

Descriptive Statistics: Uni-Variate and Bi-Variate Analysis

- Using pair-plot() from the seaborn library to analyze bivariate and univariate distributions of the attributes

Summary of Descriptive statistics

- Need for univariate and bivariate analysis

Inferential Statistics

Overview: We will go beyond descriptive statistics and into the realm of inferential statistics where we discuss the use of probability theory and probability distributions in modeling real-world processes. We explore two of the most common probability distributions and dive into a couple of concepts within Probability theory like Joint and Conditional probabilities. Finally, we end with the Central limit theorem and a brief introduction to Hypothesis testing.

How probability plays a key role in building models?

- Using probability as a measure of certainty or uncertainty
- Need for the input data to have any of the characteristic distributions like binomial, Poisson, normal etc., when using a parametric algorithm to model the outcome

Binomial Distribution: Definition, formula, and examples.

- Characteristics of a binomial distribution
- Equation of Binomial distribution
- Understanding binomial distribution using an example 'process' where the outcome of the 'process' follows a binomial distribution

Poisson Distribution: Definition, formula, and examples

- Characteristics of Poisson distribution
- The way Poisson distribution relates to Binomial distribution
- The general structure of an event following a Poisson distribution

- Equation of Poisson distribution
- Other examples of Poisson processes

Understanding Joint and Conditional Probability with some examples.
- Joint probability using an example
- Conditional probability using the same example
- Mathematical equations for Joint and conditional probabilities

Inferential Statistics: Introduction and it is important in deciding the quality of data
- Need for inferential statistics
- The concept of Central Limit Theorem

Hypothesis Testing made easy
- Meaning of the word Hypothesis and the goal of formulating a Hypothesis
- Few examples of Hypotheses
- Need for Hypothesis testing

Hypothesis Testing: Null vs Alternate Hypotheses
- Meaning of Null Hypothesis
- Meaning of Alternate Hypothesis
- Examples of Null hypotheses and corresponding Alternate Hypotheses

Hypothesis Testing

Overview: From the need and applications of hypothesis testing, we discuss the various methods of Hypothesis testing used under different circumstances.

A deeper dive into Hypothesis Testing
- Key points about Hypothesis testing, addressing the need for hypothesis testing, null and alternate hypothesis, P-value, Confidence interval, etc.

Chi-Square test- Definition, formula, and examples
- Origin of Chi-square test
- Properties of Chi-square distribution in detail
- Key points about the Chi-square test addressing the need for the test, the outcome of the test and variants of Chi-square test
- Demonstration of Chi-square test of independence using some example data

Normal Deviate Z Test - Definition, formula and examples
- The objective of the test – Verify if the sample data belongs to the population
- The sample mean vs population mean

- Population Standard deviation vs Standard error

Code Analysis: Understanding Statistics_Tests.ipynb notebook step by step
- Performing Normal Deviate Z-test in python – Starting with formulating the Null and Alternative Hypothesis, we obtain Mu and Std of the population and finally compute the Z-statistic and obtain the P-value to either reject or accept the Null hypothesis

Inferential Statistics: Using Normal Deviate Z - test and establishing data reliability
- Need for inferential stats in process modeling and EDA
- A simple demonstration of applying a z-test to find the reliability of a dummy sample data

One sample t-test: How and when to use?
- Necessity of t-statistic
- t-statistic formula
- The outcome of a t-test

Types of t-tests: One tail & two-tail tests
- Non-Directional vs Directional hypothesis testing
- One tail vs Two tail test with regards to their respective rejection regions

Analysis of Variance (ANOVA) - A simpler approach with examples
- Need for ANOVA to measure similarity/dissimilarity of two datasets/samples
- F-statistic formula
- Interpreting the F-statistic

Making Sense of the Two-Sample T-Test
- Need for 2-sampled t-test
- Paired vs unpaired 2-sample t-test
- The equation to compute t-statistic for 2-sample t-test

Making Sense of the Two-Sample T-Test with examples (Dataset Included)
- Performing 2-sample t-test on 2 dummy sample datasets in excel
- Interpreting the final result of the test statistic