

Statistics Roadmap Overview

- This roadmap outlines how to prepare for statistics, covering the entire syllabus for data analysts and data scientists.
- The playlist will include approximately 30-40 videos and will be entirely in Hindi.
- This learning path is beneficial for aspiring data analysts, data scientists, and Business Intelligence (BI) developers.

Types of Statistics

- Statistics is broadly divided into two main fields: **Descriptive Statistics** and **Inferential Statistics**.
- Data analysts primarily work with both types of statistics.

Feature	Descriptive Statistics	Inferential Statistics
Main Goal	Summarize and visualize data	Draw conclusions about a population from sample data
Focus	Describing characteristics of the observed data	Making predictions or inferences about a larger group
Key Activities	Calculating measures of central tendency/dispersion, creating charts	Hypothesis testing, estimation, prediction
Example Topics	Mean, Median, Mode, Variance, Histograms, Box Plots	Z-test, T-test, Chi-square, ANOVA, Hypothesis Testing

Descriptive Statistics


- **Descriptive Statistics** focuses on summarizing and visualizing data.
- Key topics include:
 - **Measures of Central Tendency:** Mean, Median, and Mode.
 - **Measures of Dispersion:** Variance and Standard Deviation.
 - **Correlation:** Understanding how one data point relates to another, including Spearman and Pearson correlation, and Covariance.
- Common visualizations and functions covered are:
 - **Histograms** and **Probability Density Function (PDF)**.
 - **Probability Mass Functions (PMF)** and **Cumulative Distribution Functions (CDF)**.
 - **Kernel Density Estimator**.
 - **Box Plots** (also known as Whisker Plots) are specifically used for finding outliers.
 - Other plots like bar graphs and scatter plots are also used, depending on variables and data types.
- It also covers **Univariate** and **Bivariate analysis** to analyze single or two features and their relationships.

Inferential Statistics

- **Inferential Statistics** involves drawing conclusions about a population based on sample data through experiments and tests.
- A crucial topic in inferential statistics is **Hypothesis Testing**.
 - It involves defining a **Null Hypothesis** and an **Alternate Hypothesis**.
 - This process helps data scientists predict outcomes and make conclusions about population datasets.
- Examples of statistical tests used in inferential statistics include:
 - **Z-test** and **T-test**.
 - **Chi-square test**.
 - **ANOVA** (Analysis of Variance), also known as **F-test**.
- For instance, a data analyst might use hypothesis testing to determine if a new ATM should be opened in a specific location by analyzing transaction data from nearby ATMs.

Here is a flowchart illustrating the general process of hypothesis testing:

Importance of Statistics for Decision Making

- Statistical tools are essential for understanding data and making informed decisions.
-  **Key Insight:** "Data never lies." If something is present in the data, it will be revealed through these tools.
- Sufficient data and comprehensive analysis enhance decision-making skills.

Related Business Intelligence Tools

- The roadmap also touches upon Business Intelligence (BI) tools like **Power BI** and **Tableau**.
- These tools are used to find Key Performance Indicators (KPIs) and create reports and various visualizations.