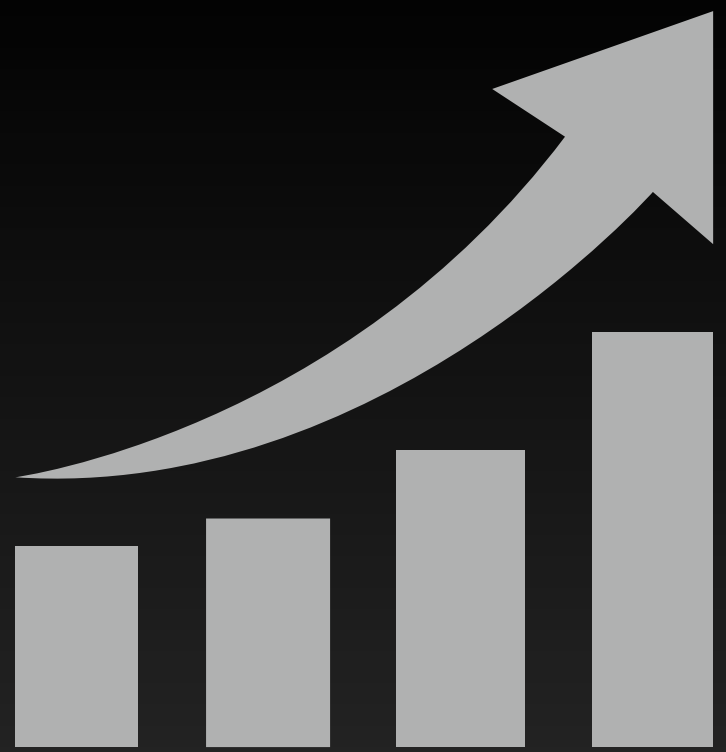
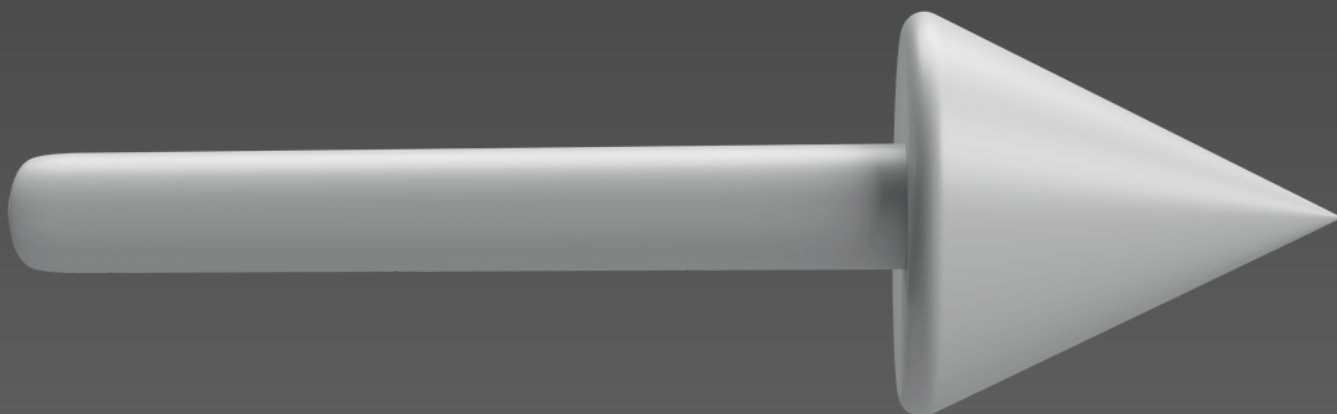


@LEARNEVERYTHINGAI



WHY STATISTICS MATTER IN DATA SCIENCE



SHIVAM MODI

@learneverythingai

Statistics is the backbone of **data science**, helping us make sense of **raw data** and draw meaningful insights.

From **hypothesis** testing to **machine learning algorithms**, statistics plays a crucial role in every step of the data analysis process.

DESCRIPTIVE STATISTICS

- Descriptive statistics summarize and describe the main features of a dataset.

- **Key Formulas:**

- **Mean:** $\mu = (\Sigma x) / n$
- **Median:** Me = Middle value
- **Mode:** Value that appears most frequently
- **Standard Deviation:** $\sigma = \sqrt{((\Sigma (xi - \mu)^2) / n)}$



INFERENCEAL STATISTICS

- Inferential statistics draw conclusions about a population from a sample of data.
- **Key Formulas:**

- **Confidence Interval:** $CI = \bar{x} \pm (Z * (\sigma/\sqrt{n}))$
- **Z-Score:** $Z = (x - \mu) / \sigma$
- **Margin of Error:** $ME = Z * (\sigma/\sqrt{n})$



PROBABILITY DISTRIBUTIONS

- Probability distributions model the likelihood of different outcomes in a dataset.
- **Key Formulas:**

- **Normal Distribution:** $N(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-(x - \mu)^2 / (2\sigma^2)}$
- **Binomial Distribution:** $P(X=k) = \binom{n}{k} * p^k * (1-p)^{n-k}$
- **Poisson Distribution:** $P(X=k) = \frac{e^{-\lambda} * \lambda^k}{k!}$



CORRELATION AND REGRESSION

- Correlation measures the relationship between two variables, while regression predicts one variable based on another.
- **Key Formulas:**

- **Correlation Coefficient (Pearson):** $r = \frac{\sum((x_i - \bar{x}) * (y_i - \bar{y}))}{n * s_x * s_y}$
- **Linear Regression:** $y = mx + b$
- **Coefficient of Determination (R^2):** $R^2 = \frac{\text{Explained Variance}}{\text{Total Variance}}$



HYPOTHESIS TESTING

- Hypothesis testing helps us make decisions about a population based on sample data.
- **Key Formulas:**

- **Null Hypothesis (H_0) vs. Alternative Hypothesis (H_1)**
- **p-value: Probability of obtaining observed results under H_0**
- **Significance Level (α): Threshold for p-value acceptance**



HYPOTHESIS TESTING

- Hypothesis testing helps us make decisions about a population based on sample data.
- **Key Formulas:**

- **Null Hypothesis (H_0) vs. Alternative Hypothesis (H_1)**
- **p-value: Probability of obtaining observed results under H_0**
- **Significance Level (α): Threshold for p-value acceptance**



SAMPLING AND ESTIMATION

- Sampling techniques and estimation methods allow us to draw conclusions about a population from a smaller sample.
- **Key Formulas:**

- **Simple Random Sampling:** Each member has an equal chance of selection.
- **Point Estimation:** Using sample data to estimate population parameters.
- **Margin of Error:** $ME = (Z * (\sigma/\sqrt{n}))$ for estimating population mean.



@learneverythingai

Like this Post?

- Follow Me
- Share with your friends
- Check out my previous posts



SHIVAM MODI
@learneverythingai



SAVE THIS

Follow

 **SHARE**

www.learneverythingai.com