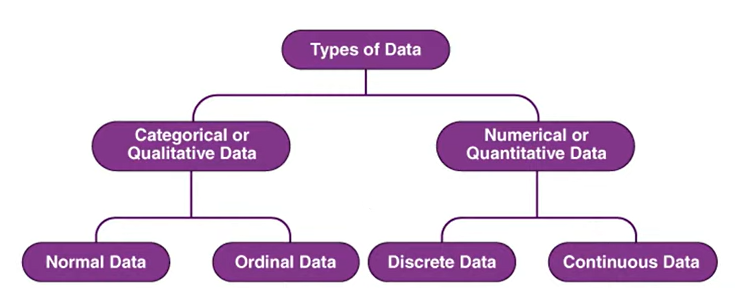
Statistics

1. Descriptive: For Exploratory Data Analysis on existing data
2. Inferential: For Future Prediction
   1. Hypothesis testing
   2. Confidence intervals
   3. Analysis of variance (ANOVA)
   4. Regression Analysis
   5. Chi-square test
   6. Sampling technique
   7. Bayesian statistics

Population: Example all citizens of India

Sample: When we predict average age, income then we take data from some people from different sections of people. This set is called sample.

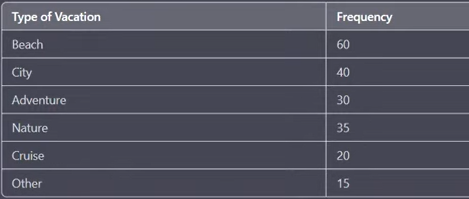


|  |  |  |  |
| --- | --- | --- | --- |
| States of India  Religions | Feedback  Degrees | Integer Only  Roll number | May be decimal  Temperature  Height |

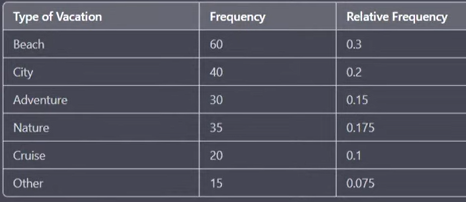
|  |  |
| --- | --- |
| **Mean:** *Highly sensitive/influential/****prone to outlier*** |  |
| **Median:** *It is the middle value of data set when arranged in order.*  *It is* ***NOT prone*** *to* ***outliers*** | |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | 3 | 4 | 1 | 2 | 5 | 9 | 6000 |   Arrange->   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | 1 | 2 | 3 | 4 | 5 | 7 | 6000 |   for even set of numbers, Median= Middle number=4,  for even set of numbers, Median= (sum of middle two numbers)/2 |
| **Mode:** *It is the value which appears most frequently.*  *It is used mainly in categorical or discrete numerical values.* |  |
| **Weighted Mean:** | = (weight1\*value1+ weight2\*value2+….+ weightn\*valuen) / weight1 + weight2 + …. +weightn |
| **Trimmed Mean:** *First trim the outliers with IQR and the take mean.* |  |
| **Measure of Dispersion:** *It is a statistical measure that describes the spread or visibility of dataset.* |  |
| **Measure of Central Tendency:** *It is the measure of central position of data.* |  |
| **Range:** *It is the difference between min and max.* |  |
| **Variance:** *It is the average of the square of the difference between each data point and the mean.*  *It is* ***very much prone to outliers*** *because of square.*  *Change the unit of base data* |  |
| **Standard Deviation:** *it is the square root of Variance* |  |
| **Coefficient of Variation(CV):** *is used to compare spread of data between two different value* | **(Standard Deviation /Mean)\*100** |

Types of charts for different types:

1. Categorical Column:
   1. Frequency distribution table: The number of times a category repeats in sample.
      1. **Bar Chart**



* 1. Relative frequency: Percentage of a category in a dataset or sample.
     1. **Pie Chart**



* 1. Cumulative frequency: Running total of frequency or relative frequency of a category in the sample.
     1. **Line Plot**

