|  |  |
| --- | --- |
| **Descriptive Statistics** | **Inferential Statistics** |
| Focuses on organizing and summarizing the data | Focuses on making predictions, drawing conclusions about a population based on a sample |
| Measures of Central Tendency  (Mean, Median, Mode)  Measures of Central Dispersion  (Range, Variance, Std Dev)  Graphical Representations  (Histograms, Boxplots, Pie-Charts) | Hypothesis Testing  Confidence Intervals  Regression Analysis |
| Provides a snapshot of the data. (summarization) | Provides Insights that can be generalized to the population |
| Uses sample/ entire population data | Works with a sample to infer about a larger population |
| Example: find the mean test scores of a class | Predicting the mean test scores of all students based on a sample |

A **hypothesis** is an educated guess about a relationship between variables or a characteristic of a population based on sample data.

**Hypothesis Testing** is to determine whether there is enough statistical evidence to support or reject a specific claim.

=======================================================

A **confidence interval** helps estimate a population parameter (like an average or proportion) falls within a range of possible values based on your sample data with a certain level of confidence.

The 95% represents the **confidence level**, meaning if you repeated the sampling process many times, about 95% of the calculated intervals would contain the true population parameter.

**For example**, if you survey 100 people and calculate a

confidence interval for the average age to be 95%, and

provide the confidence interval : [ 25 to 30 years].

This means you're 95% confident that the true average age of the entire population falls somewhere between 25 and 30 years.

**Regression analysis** works by fitting a mathematical model to your data, estimating the relationship between the variables, and using that model to make predictions or draw insights.

It's about identifying how one variable is influenced by other variables.

Y=β0​+β1​X+ϵ

Here, Y is the dependent variable (the outcome),

X is the independent variable (the predictor),

β0 ​ is the intercept (where the line crosses the Y-axis),

β1​ is the slope (how much Y changes for each unit change in X).

ϵ (epsilon) represents the error or residual (the difference between the predicted and actual values).