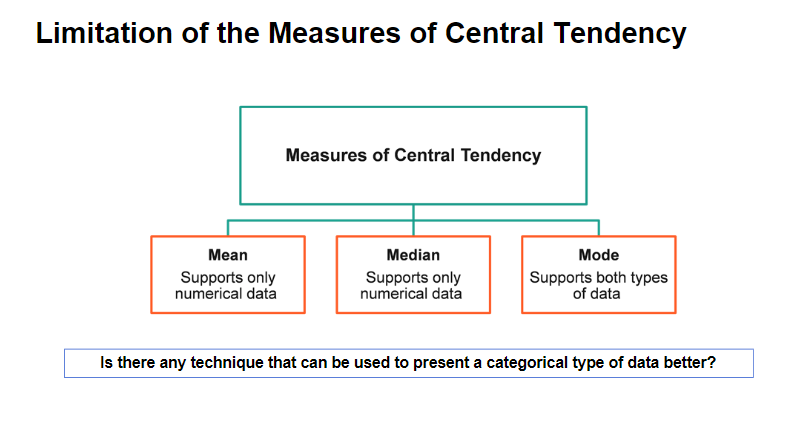
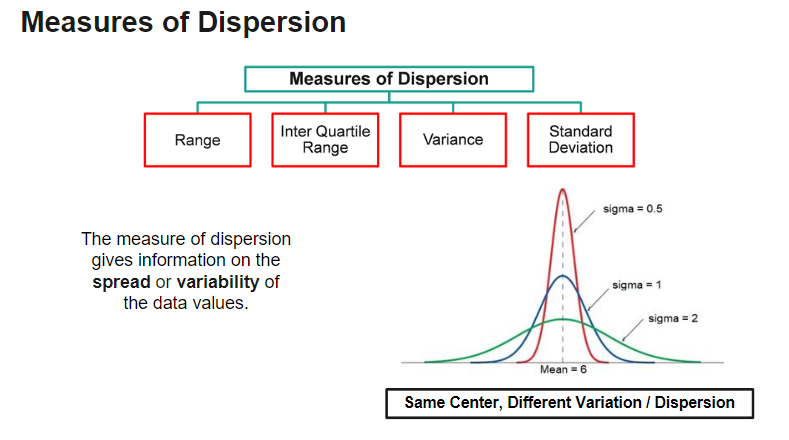


1. **Mean (Average)**: The mean is the sum of all data points divided by the number of data points.
2. **Median**: The median is the middle value when the data is ordered, or the average of the two middle values in case of an even number of data points.
3. **Mode**: The mode is the value that appears most frequently in the dataset.



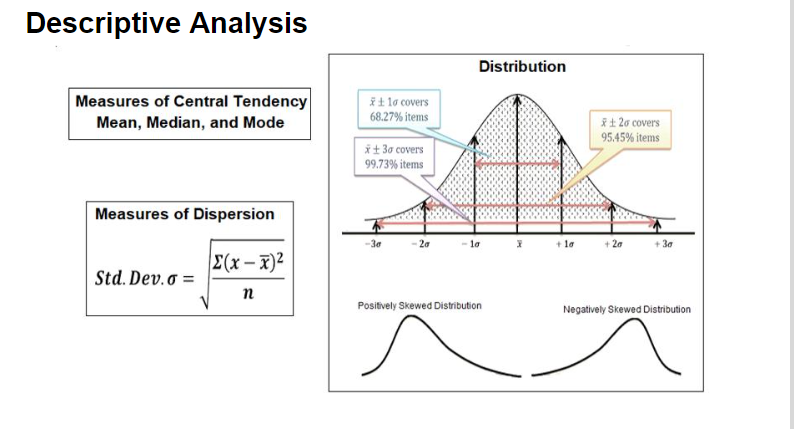


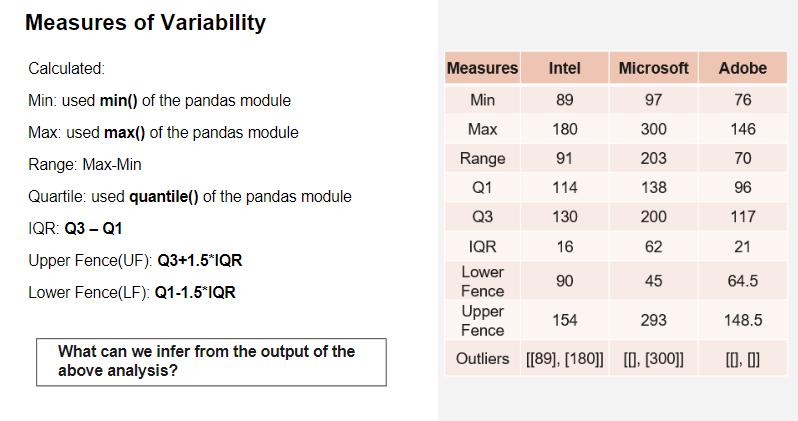
1. **Range**:
   * The range is the simplest measure of dispersion.
   * It is calculated as the difference between the maximum and minimum values in a dataset.
   * Range = Maximum Value - Minimum Value
2. **Variance**:
   * Variance measures how far each data point in the dataset is from the mean.
   * It is the average of the squared differences between each data point and the mean.
   * Variance is commonly denoted as �2*σ*2 for a population and �2*s*2 for a sample.
3. **Standard Deviation**:
   * The standard deviation is the square root of the variance.
   * It provides a measure of the average distance between each data point and the mean.
   * Standard Deviation = √(Variance)
4. **Interquartile Range (IQR)**:
   * The IQR is a measure of the spread of the middle 50% of the data.
   * It is the difference between the third quartile (Q3) and the first quartile (Q1) of the data.
   * IQR = Q3 - Q1

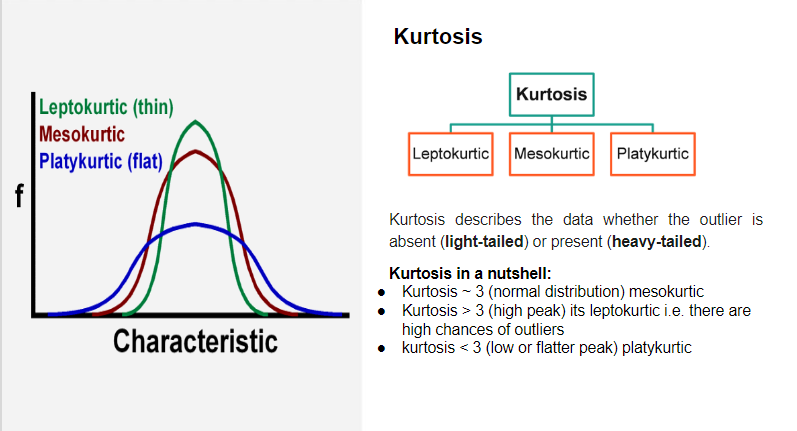
**Coefficient of Variation (CV)**:

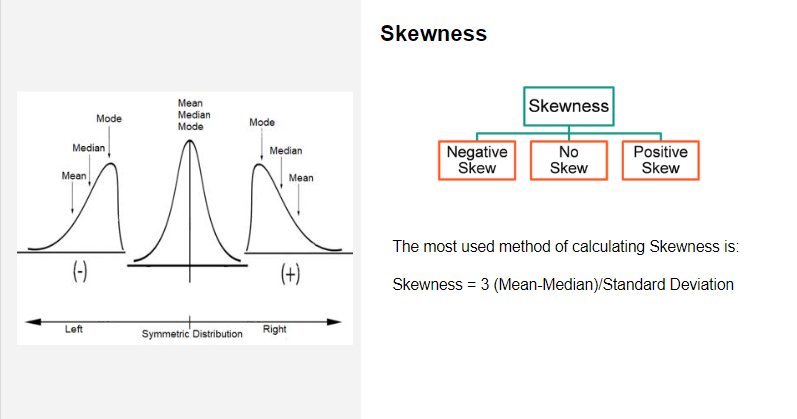
* The coefficient of variation measures the relative variability of data, especially when comparing datasets with different units or scales.
* It is calculated as the standard deviation divided by the mean, expressed as a percentage.
* CV = (Standard Deviation / Mean) \* 100

Emperical Law 68.27% data comes under the 1st Standard Deviation

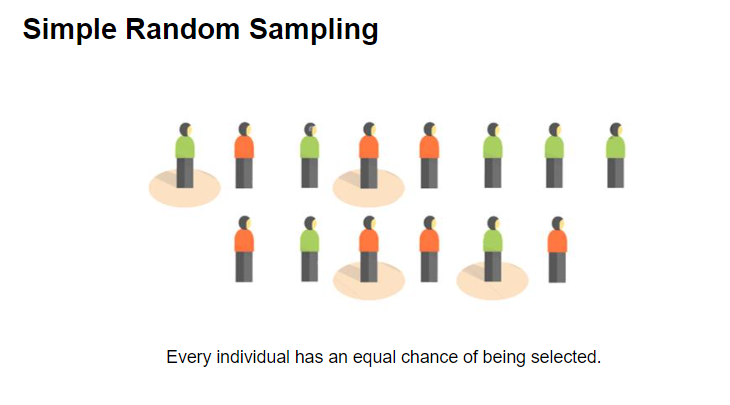


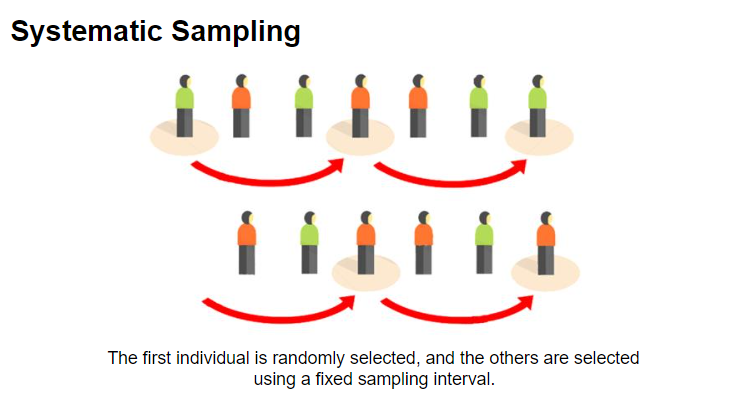


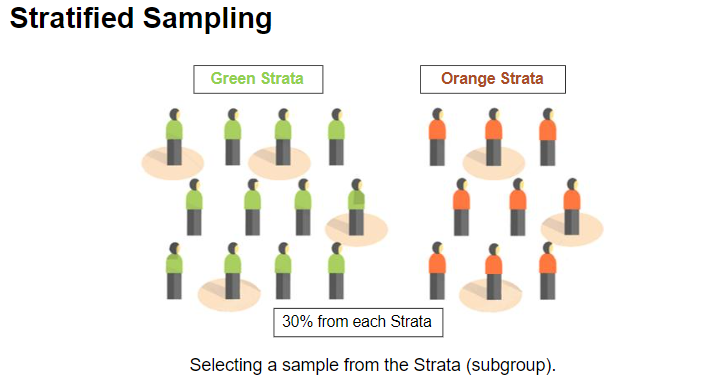


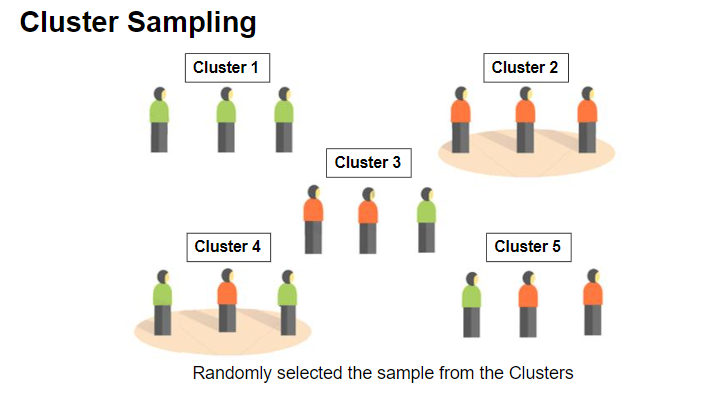


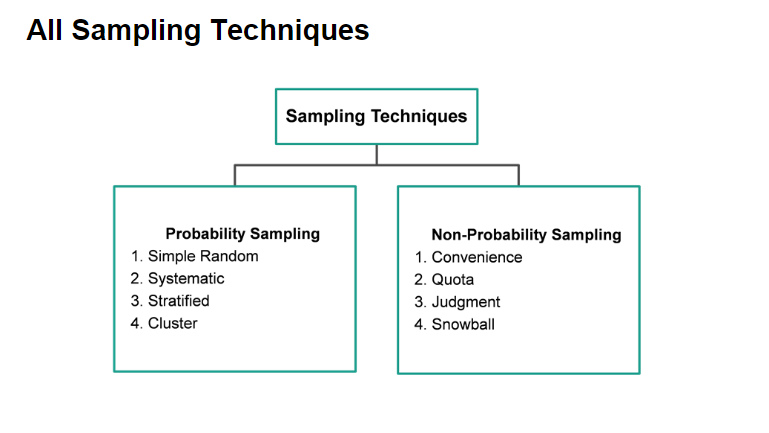


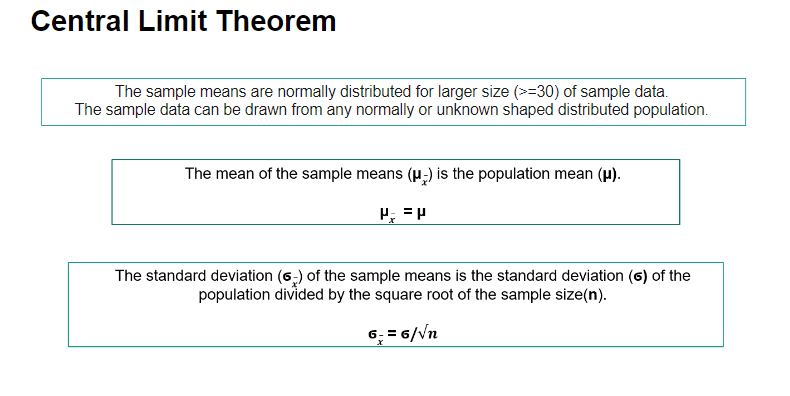


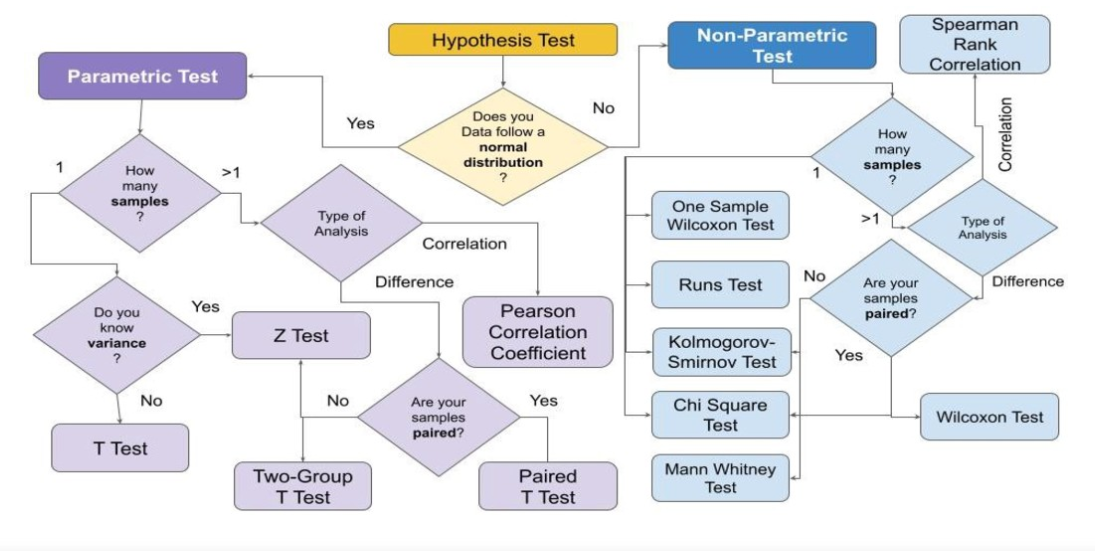












**There are two types of errors that can occur when executing a Hypothesis Test:**

• **Type-I Error:** Rejecting the Null Hypothesis when it is actually true is a Type-I Error.

• **Type-II Error:** Accepting the Null Hypothesis when it is false is a Type-II Error.

Types of Hypothesis Testing

* **Hypothesis tests are divided into two categories:**

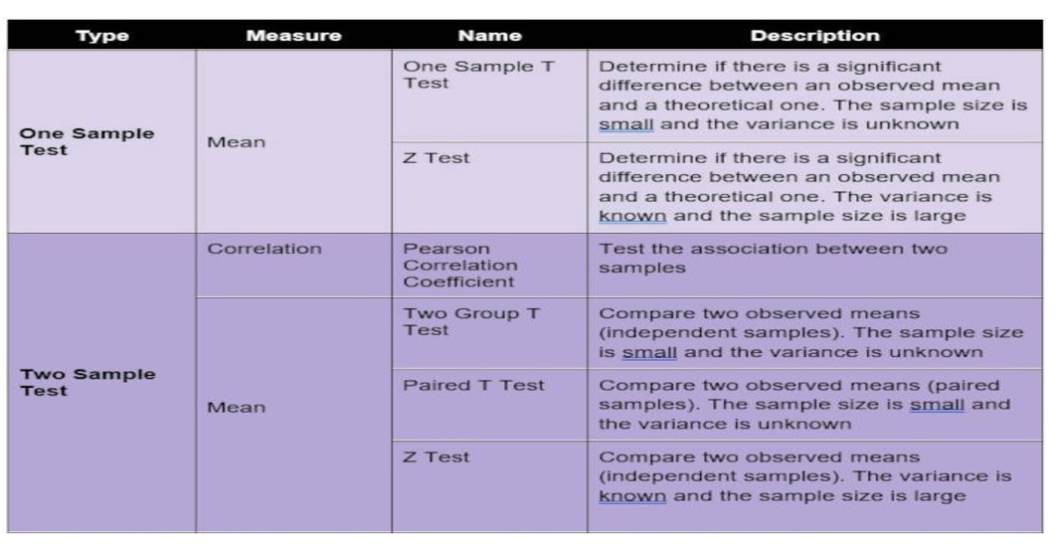
**1) Parametric tests –** are used when the samples have a normal distribution. In general, samples with a mean of 0 and a variance of 1 follow a normal distribution.

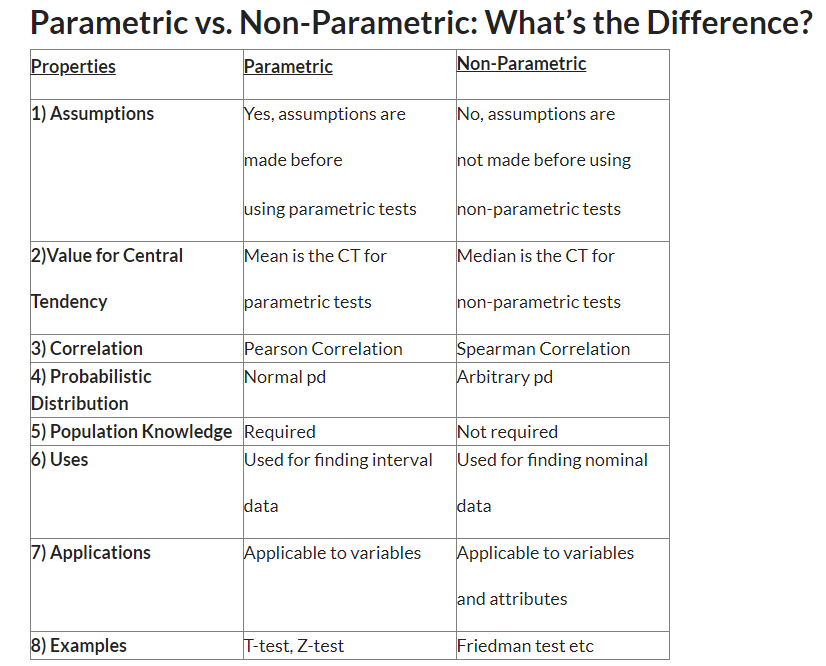
**2) Non-Parametric tests –** If the samples do not follow a normal distribution, non-parametric tests are used.

* **Two types of Hypothesis Testing can be created depending on the number of samples to be compared:**

• **One Sample –** If there is only one sample that must be compared to a specific value, it is called a single sample.

• **Two Samples –** if you’re comparing two or more samples. Correlation and sample difference are two tests that could be used in this situation. Samples can be paired or not in both circumstances. Dependent samples are sometimes known as paired samples, while independent samples are known as unpaired samples. Natural or matched couplings occur in paired samples.





**T-Test:**

**1.** It is a parametric test of hypothesis testing based on **Student’s T distribution**.

**2.** It is essentially, testing the significance of the difference of the mean values when the sample size is small (i.e, less than 30) and when the population standard deviation is not available.

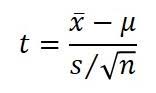
**3.** Assumptions of this test:

* Population distribution is normal, and
* Samples are random and independent
* The sample size is small.
* Population standard deviation is not known.

**4.** Mann-Whitney ‘U’ test is a non-parametric counterpart of the T-test.

**A T-test can be a:**

**One Sample T-test:**To compare a sample mean with that of the population mean.



**where,**

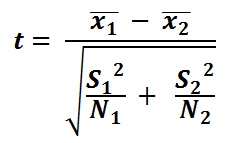
**x̄** is the sample mean

**s** is the sample standard deviation

**n** is the sample size

**μ** is the population mean

**Two-Sample T-test:** To compare the means of two different samples.



where,

**x̄1** is the sample mean of the first group

**x̄2** is the sample mean of the second group

**S1** is the sample-1 standard deviation

**S2**is the sample-2 standard deviation

**n** is the sample size

**Conclusion:**

* If the value of the test statistic is greater than the table value -> **Rejects the null hypothesis**.
* If the value of the test statistic is less than the table value ->**Do not reject the null hypothesis**.

**Z-Test:**

**1.** It is a parametric test of hypothesis testing.

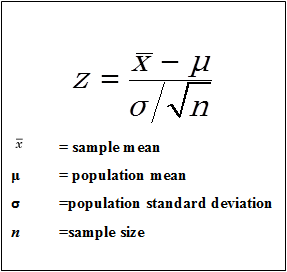
**2.** It is used to determine whether the means are different when the population variance is known and the sample size is large (i.e, greater than 30).

**3.**Assumptions of this test:

* Population distribution is normal
* Samples are random and independent.
* The sample size is large.
* Population standard deviation is known.

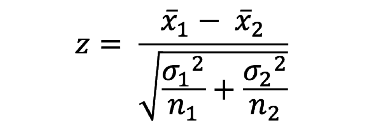
**A Z-test can be:**

**One Sample Z-test:**To compare a sample mean with that of the population mean.



***Image Source: Google Images***

**Two Sample Z-test:** To compare the means of two different samples.



**where,**

**x̄1** is the sample mean of 1st group

**x̄2** is the sample mean of 2nd group

**σ1** is the population-1 standard deviation

**σ2**is the population-2 standard deviation

**n** is the sample size

**F-Test:**

**1.** It is a parametric test of hypothesis testing based on **Snedecor F-distribution**.

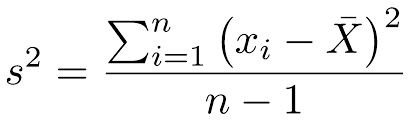
**2.** It is a test for the null hypothesis that two normal populations have the same variance.

**3.** An F-test is regarded as a comparison of equality of sample variances.

**4.** F-statistic is simply a ratio of two variances.

**5.** It is calculated as:

F = s12/s22



**6.** By changing the variance in the ratio, F-test has become a very flexible test. It can then be used to:

* Test the overall significance for a regression model.
* To compare the fits of different models and
* To test the equality of means.

**7.** Assumptions of this test:

* Population distribution is normal, and
* Samples are drawn randomly and independently.

**ANOVA :**

**1.** Also called as **Analysis of variance**, it is a parametric test of hypothesis testing.

**2.** It is an extension of the T-Test and Z-test.

**3.** It is used to test the significance of the differences in the mean values among more than two sample groups.

**4.** It uses F-test to statistically test the equality of means and the relative variance between them.

**5.** Assumptions of this test:

* Population distribution is normal, and
* Samples are random and independent.
* Homogeneity of sample variance.

**6.** One-way ANOVA and Two-way ANOVA are is types.

**7. F-statistic = variance between the sample means/variance within the sample**

ANOVA is always one-tailed test with rejection region in right tail.