1) Define T-Test?

Statistical method for the comparison of the mean of the two groups of the normally distributed sample(s).

It is used when:

- Population parameter (mean and standard deviation) is not known
- Sample size (number of observations) < 30

2) T-Test: (Explanation)

Type of t-test.

The T-test is mainly classified into 3 parts:

- One sample
- Independent sample
- Paired sample

1. One Sample

In one sample t-test, we compare the sample mean with the population mean.

Mathematical Formula:

$$t = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\overline{X} = \text{Sample mean}$$
$$\mu = \text{Population mean}$$
$$\sigma = \text{sample standard deviation}$$
$$n = \text{sample size}$$

- Region of rejection lies either on extreme left or extreme right of the distribution.
- In z-test, we use population standard deviation instead of sample standard deviation.

Example:

Problem Statement:

Marks of student are 10.5, 9, 7, 12, 8.5, 7.5, 6.5, 8, 11 and 9.5.

Mean population score is 12 and standard deviation is 1.80.

Is the mean value for student significantly differ from the mean population value.

Solution:

Firstly, we will calculate the mean of 10 students:

$$\overline{X} = \frac{10.5+9+7+12+8.5+7.5+6.5+8+11+9.5}{10} = 8.95$$

Step-1: State Null and Alternate Hypothesis
Null Hypothesis:

$$H_0: \overline{X} = 8.95$$

Alternate Hypothesis:

$$H_a: \overline{X} > 8.95$$

Step-2: Set the significance level (alpha level)
Let alpha-value is 0.05, so corresponding t-value is 2.262

Step-3: Find the t-value

$$t = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{8.95-12}{\frac{1.80}{\sqrt{10}}} = -5.352$$

Step-4: Comparison with the significance level
From step-3, we have
|-5.352| > 2.262
So, we have to reject the null hypothesis.
i.e. there is significantly difference between mean of sample and population.

2. **Independent (two-sample t-test):**

In this test, we compare the means of two different samples.
Mathematical Formula:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2}}}$$

$\overline{X}_1, \overline{X}_2$: *Sample Mean*

n_1, n_2: *Sample Size*

s^2: *estimator of common variance such that*

$$s^2 = \frac{\Sigma(x - \overline{X}_1)^2 + \Sigma(x - \overline{X}_2)^2}{(n_1 - 1) + (n_2 - 1)}, \text{ where}$$

$(n_1 - 1) + (n_2 - 1)$: degree of freedom

Degree of Freedom:

Degree of freedom is defined as the number of independent variables.

It is given by:

$$df = \Sigma(n_i - 1),$$

Where

df = *degree of freedom*

n_i = *sample size*

Note: There are two regions of rejection, one in either directions towards tail of each distribution.

Let's understand two-sample t-test by an example:

Problem Statement:

The marks of boys and girls are given:

Boys: 12, 14, 10, 8, 16, 5, 3, 9, and 11

Girls: 21, 18, 14, 20, 11, 19, 8, 12, 13, and 15

Is there any significant differnece between marks of males and females i.e. population means are different.

Solution:

Firstly, we will calculate mean, standard deviation and degree of freedom for marks of boys and girls:
Boys:

$$N_1 = 9,$$
$$df = (9 - 1) = 8$$
$$\overline{X}_1 = 9.778,\ s_1 = 4.1164$$

Girls:

$$N_2 = 10,$$
$$df = (10 - 1) = 9$$
$$\overline{X}_2 = 15.1,\ s_2 = 4.2805$$

Step-1: State Null and Alternate Hypothesis:

Null Hypothesis:

$$H_0:\ \mu_1 = \mu_2$$

Alternate Hypothesis:

$$H_0:\ \mu_1 \neq \mu_2$$

Step-2: Set the significance level (alpha level)
Let the alpha-value is 0.05, and
since the degree of freedom is 9+8=17
So, t-value is 2.11

Step-3: Find the t-value

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{9.778 - 15.1}{\sqrt{\frac{(4.1164)^2}{9} + \frac{(4.2805)^2}{10}}} = \frac{-5.322}{1.93} = -\ 2.758$$

Step-4: Comparison with the significance level
From step-3, we have
|-2.758| > 2.11
So, we have to reject the null hypothesis.
i.e. population means are different

Firstly, we will calculate mean, standard deviation and degree of freedom for marks of boys and girls:

3. **Paired t-test:**

   In this test, we compare the means of two related or same group at two different time.
   Mathematical Formula:

$$t = \frac{m}{\frac{s}{\sqrt{n}}}$$

m: *mean of difference between each pair of values*
s: *standard deviation of difference between each pair of values*
n: *sample size*

Note: Degree of freedom is n-1.

Let's understand two-sample t-test by an example:

Problem Statement:
Blood pressure of 8 patients are before and after are recorded:
Before: 180, 200, 230, 240, 170, 190, 200, and 165
After: 140, 145, 150, 155, 120, 130, 140, and 130
Is there any significant difference between BP reading before and after.

Solution:

Firstly, we will find the mean and standard deviation of difference between each pair of values

Before	After	d (= Before - After)	d^2
180	140	40	1600
200	145	55	3025
230	150	80	6400
240	155	85	7225
170	120	50	2500
190	130	60	3600
200	140	60	3600
165	130	35	1225
		$\Sigma d = 465$	$\Sigma d^2 = 29175$

$$Mean\ (m) = \frac{\Sigma d}{8} = \frac{465}{8} = 58.125$$

$$s = \sqrt{\frac{\Sigma d^2 - \frac{(\Sigma d)^2}{n}}{n-1}} = \sqrt{\frac{(29175) - \frac{(465)^2}{8}}{8-1}} = 17.51$$

Step-1: State Null and Alternate Hypothesis:

Null Hypothesis:

H_0: *there is no significant difference between BP before and after*

Alternate Hypothesis:

H_a: *there is significant difference between BP before and after*

Step-2: Set the significance level (alpha level)

Let the alpha-value is 0.05, and
since the degree of freedom is 8-1=7
So, t-value is 2.36

Step-3: Find the t-value

$$\boxed{t = \frac{m}{\frac{s}{\sqrt{n}}} = \frac{58.125}{\frac{17.51}{\sqrt{8}}} = \frac{58.125}{6.191} = 9.38}$$

Step-4: Comparison with the significance level

From step-3, we have
9.38 > 2.36
So, we have to reject the null hypothesis.
i.e. there is significant difference between BP reading before and after.

Conclusion:

T-test is a statistically significant test for the hypothesis testing (null and alternative hypotheses) when the sample size is small and the population parameter (mean and variance) is unknown.

3) Define F-Test?

An F-test is any statistical test in which the test statistic has an F-distribution under the null hypothesis. It is most often used when comparing statistical models that have been fitted to a data set, in order to identify the model that best fits the population from which the data were sampled.

4) F-Test: (Explanation)

F-Test is any test that utilizes the F-Distribution table to fulfil its purpose (for eg: ANOVA). It compares the ratio of the variances of two populations and determines if they are statistically similar or not.

We can use this test when:

- The population is normally distributed.
- The samples are taken at random and are independent samples.

Formulas Used

$$F_{calc} = \frac{\sigma_1^2}{\sigma_2^2}$$

where,

F_{calc} = Critical F-value.
σ_1^2 & σ_2^2 = variance of the two samples.

$$df = n_S - 1$$

where,

df = Degrees of freedom of the sample.
n_s = Sample size.

Steps involved:
Step 1: Use Standard deviation (σ) and find variance (σ2) of the data. (if not already given)
Step 2: Determine the null and alternate hypothesis.

- H0 -> no difference in variances.
- Ha -> difference in variances.

Step 3: Find Fcalc using Eq-1.

***NOTE :** While calculating **F_{calc}**, divide the larger variance with small variance as it makes calculations easier.*
Step 4: Find the degrees of freedom of the two samples.
Step 5: Find Ftable value using d1 and d2 obtained in Step-4 from the F-distribution table. (link here). Take learning rate, α = 0.05 (if not given)
Looking up the F-distribution table:
In the F-Distribution table, refer the table as per the given value of α in the question.

- d1 (Across) = df of the sample with numerator variance. (larger)
- d2 (Below) = df of the sample with denominator variance. (smaller)

Consider the F-Distribution table given below,

While performing One-Tailed F-Test.
GIVEN :
α = 0.05

$d_1 = 2$
$d_2 = 3$

d_2 / d_1	1	2	..
1	161.4	199.5	..
2	18.51	19.00	..
3	10.13	9.55	..
:	:	:	..

Then, $\mathbf{F_{table} = 9.55}$
Step 6: Interpret the results using *Fcalc* and *Ftable*.
Interpreting the results:

```
If Fcalc < Ftable :
    Cannot reject null hypothesis.
    ∴ Variance of two populations are similar.

If Fcalc > Ftable :
    Reject null hypothesis.
    ∴ Variance of two populations are not similar.
```

Example Problem (Step by Step)
Consider the following example,

Conduct a two-tailed F-Test on the following samples:

	Sample 1	Sample 2
σ	10.47	8.12
n	41	21

Step 1:
- $\boldsymbol{\sigma_1^2} = (10.47)^2 = 109.63$
- $\boldsymbol{\sigma_2^2} = (8.12)^2 = 65.99$

Step 2:

- H_0: no difference in variances.
- H_a: difference in variances.

Step 3:

F_{calc} = (109.63 / 65.99) = **1.66**

Step 4:

$d_1 = (n_1 – 1)$ = (41 – 1) = 40

$d_2 = (n_2 — 1)$ = (21 – 1) = 20

Step 5 - Using d_1 = **40** and d_2 = **20** in the F-Distribution table.
Take **α = 0.05** as it's not given.
Since it is a two-tailed F-test,
α = 0.05/2
= 0.025
Therefore, F_{table} = **2.287**

Step 6 - Since $F_{calc} < F_{table}$ **(1.66 < 2.287)**:
We cannot reject null hypothesis.
∴ Variance of two populations are similar to each other.

F-Test is the most often used when comparing statistical models that have been fitted to a data set to identify the model that best fits the population. Researchers usually use it when they want to test whether two independent samples have been drawn from a normal population with the same variability.

5) What is analysis of variance?

Analysis of variance is a collection of statistical models and their associated estimation procedures used to analyze the differences among means. ANOVA was developed by the statistician Ronald Fisher

6) Define effect size estimation.

Effect size estimates provide important information about the impact of a treatment on the outcome of interest or on the association between variables. • Effect size estimates provide a common metric to compare the direction and strength of the relationship between variables across studies.

7) What is mean by multiple comparisons, multiplicity or multiple testing.

The multiple comparisons, multiplicity or multiple testing problem occurs when one considers a set of statistical inferences simultaneously or infers a subset of parameters selected based on the observed values. The more inferences are made, the more likely erroneous inferences become

8) Define ANOVA.

Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not. Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.

9) The Formula for ANOVA is:

$$F = \frac{MST}{MSE}$$

where:

F = ANOVA coefficient

MST = Mean sum of squares due to treatment

MSE = Mean sum of squares due to error

10) One-Way ANOVA vs. Two-Way ANOVA:

There are two main types of analysis of variance: one-way (or unidirectional) and two-way (bidirectional). One-way or two-way refers to the number of independent variables in your analysis of variance test. A one-way ANOVA evaluates the impact of a sole factor on a sole response variable. It determines whether the observed differences between the means of independent (unrelated) groups are explainable by chance alone, or whether there are any statistically significant differences between groups.

A two-way ANOVA is an extension of the one-way ANOVA. With a one-way, you have one independent variable affecting a dependent variable. With a two-way ANOVA, there are two independents. For example, a two-way ANOVA allows a company to compare worker productivity based on two independent variables, such as department and gender. It is utilized to observe the interaction between the two factors. It tests the effect of two factors at the same time.

A three-way ANOVA, also known as three-factor ANOVA, is a statistical means of determining the effect of three factors on an outcome.

11) Mention a two-factor factorial design.

A two-factor factorial design is an experimental design in which data is collected for all possible combinations of the levels of the two factors of interest. If equal sample sizes are taken for each of the possible factor combinations then the design is a balanced two-factor factorial design.

12) Define statistical test in F-test.

An F-test is any statistical test in which the test statistic has an F-distribution under the null hypothesis. It is most often used when comparing statistical models that have been fitted to a data set, in order to identify the model that best fits the population from which the data were sampled.

13) What are the two- way analyses of variance?

The two-way analysis of variance is an extension of the one-way ANOVA that examines the influence of two different categorical independent variables on one continuous dependent variable.

14) What are the types of ANOVA?

There are two main types of ANOVA: one-way (or unidirectional) and two-way. There also variations of ANOVA. For example, MANOVA (multivariate ANOVA) differs from ANOVA as the former tests for multiple dependent variables simultaneously while the latter assesses only one dependent variable at a time.

15) Define chi-square test.

The Chi-Square test is a statistical procedure used by researchers to examine the differences between categorical variables in the same population.

For example, imagine that a research group is interested in whether or not education level and marital status are related for all people in the U.S.

16) What Does the Analysis of Variance Reveal?

The ANOVA test is the initial step in analyzing factors that affect a given data set. Once the test is finished, an analyst performs additional testing on the methodical factors that measurably contribute to the data set's inconsistency. The analyst utilizes the ANOVA test results in an f- test to generate additional data that aligns with the proposed regression models.

The ANOVA test allows a comparison of more than two groups at the same time to determinewhether a relationship exists between them. The result of the ANOVA formula, the F statistic (also called the F-ratio), allows for the analysis of multiple groups of data to determine the variability between samples and within samples.

If no real difference exists between the tested groups, which is called the null hypothesis, the result of the ANOVA's F-ratio statistic will be close to 1. The distribution of all possible values of the F statistic is the F-distribution. This is actually a group of distribution functions, with two characteristic numbers, called the numerator degrees of freedom and the denominator degrees offreedom.

17) How to Use ANOVA?

A researcher might, for example, test students from multiple colleges to see if students from one of the colleges consistently outperform students from the other colleges. In a business application, an R&D researcher might test two different processes of creating a product to see if one process is better than the other in terms of cost efficiency.

The type of ANOVA test used depends on a number of factors. It is applied when data needs to be experimental. Analysis of variance is employed if there is no access to statistical software resulting in computing ANOVA by hand. It is simple to use and best suited for small samples. With many experimental designs, the sample sizes have to be the same for the various factor level combinations.

ANOVA is helpful for testing three or more variables. It is similar to multiple two-sample t-tests. However, it results in fewer type I errors and is appropriate for a range of issues. ANOVA groups differences by comparing the means of each group and includes spreading out the variance into diverse sources. It is employed with subjects, test groups, between groups and within groups.

18) What is the Analysis of Variance in Other Applications

In addition to its applications in the finance industry, ANOVA is also used in a wide variety of contexts and applications to test hypotheses in reviewing clinical trial data.For example, to compare the effects of different treatment protocols on patient outcomes; in social science research (for instance to assess the effects of gender and class on specified variables), in software engineering (for instance to evaluate database management systems), in manufacturing (to assess product and process quality metrics), and industrial design among other fields.

19) What is a Test?

In technical analysis and trading, a test is when a stock's price approaches an established support or resistance level set by the market. If the stock stays within the support and resistance levels, the test passes. However, if the stock price reaches new lows and/or new highs, the test fails. In other words, for technical analysis, price levels are tested to see if patterns or signals are accurate.

A test may also refer to one or more statistical techniques used to evaluate differences or similarities between estimated values from models or variables found in data. Examples includethe t-test and z-test

20) Define Range-Bound Market Test.

When a stock is range-bound, price frequently tests the trading range's upper and lower boundaries. If traders are using a strategy that buys support and sells resistance, they should wait for several tests of these boundaries to confirm price respects them before entering a trade.

Once in a position, traders should place a stop-loss order in case the next test of support or resistance fails.

21) What is the Trending Market Test?

In an up-trending market, previous resistance becomes support, while in a down-trending market, past support becomes resistance. Once price breaks out to a new high or low, it often retraces to test these levels before resuming in the direction of the trend. Momentum traders can use the test of a previous swing high or swing low to enter a position at a more favorable price than if they would have chased the initial breakout.

A stop-loss order should be placed directly below the test area to close the trade if the trend unexpectedly reverses.

22) Define Statistical Tests.

Inferential statistics uses the properties of data to test hypotheses and draw conclusions. Hypothesis testing allows one to test an idea using a data sample with regard to a population parameter. The methodology employed by the analyst depends on the nature of the data used and the reason for the analysis. In particular, one seeks to reject the null hypothesis, or the notion that one or more random variables have no effect on another. If this can be rejected, the variables are likely to be associated with one another

23) What is Alpha Risk?

Alpha risk is the risk that in a statistical test a null hypothesis will be rejected when it is actuallytrue. This is also known as a type I error, or a false positive. The term "risk" refers to the chance or likelihood of making an incorrect decision. The primary determinant of the amount of alpha risk is the sample size used for the test. Specifically, the larger the sample tested, the lower the alpha risk becomes.

Alpha risk can be contrasted with beta risk, or the risk of committing a type II error (i.e., a falsenegative).

Alpha risk, in this context, is unrelated to the investment risk associated with an actively managed portfolio that seeks alpha, or excess returns above the market.

24) What is Range-Bound Trading?

Range-bound trading is a trading strategy that seeks to identify and capitalize on securities, like stocks, trading in price channels. After finding major support and resistance levels and connecting them with horizontal trendlines, a trader can buy a security at the lower trendline support (bottom of the channel) and sell it at the upper trendline resistance (top of the channel).

25) What is a One-Tailed Test?

A one-tailed test is a statistical test in which the critical area of a distribution is one-sided so that it is either greater than or less than a certain value, but not both. If the sample being tested falls into the one-sided critical area, the alternative hypothesis will be accepted instead of the null hypothesis.