Unit No-02

T test

What is a *t*-test?

A *t*-test (also known as Student's *t*-test) is a tool for evaluating the means of one or two populations using hypothesis testing. A t-test may be used to evaluate whether a single group differs from a known value (a one-sample t-test), whether two groups differ from each other (an independent two-sample t-test), or whether there is a significant difference in paired measurements (a paired, or dependent samples t-test).

A t test is appropriate to use when we have collected a small, random sample from some statistical "population" and want to compare the mean from our sample to another value. The value for comparison could be a fixed value (e.g., 10) or the mean of a second sample.

For example, if our variable of interest is the average height of sixth graders in our region, then we might measure the height of 25 or 30 randomly-selected sixth graders. A t test could be used to answer questions such as, "Is the average height greater than four feet?"

t- Test assumptions

While *t*-tests are relatively robust to deviations from assumptions, *t*-tests do assume that:

- The data are continuous.
- The sample data have been randomly sampled from a population.

- There is homogeneity of variance (i.e., the variability of the data in each group is similar).
- The distribution is approximately normal.

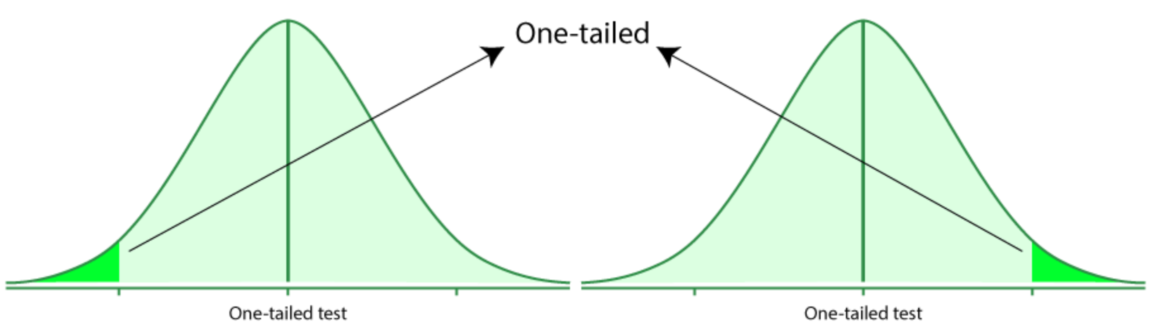
For two-sample *t*-tests, we must have independent samples. If the samples are not independent, then a paired *t*-test may be appropriate.

Hypothesis Testing:

One and Two-Tailed Tests are ways to identify the relationship between the statistical variables. For checking the relationship between variables in a **single direction** (Left or Right direction), we use a one-tailed test. A two-tailed test is used to check whether the relations between variables are in any direction or not.

One-Tailed Test

A one-tailed test is based on a uni-directional hypothesis where the area of rejection is on only one side of the sampling distribution. It determines whether a particular population parameter is larger or smaller than the predefined parameter. It uses one single critical value to test the data.



Difference Between One-Tailed and Two-Tailed Tests

Null Hypothesis (*H*0): where represents a parameter (e.g., population mean) and *θ*0 is a specific value.

Alternative Hypothesis (*H*1):

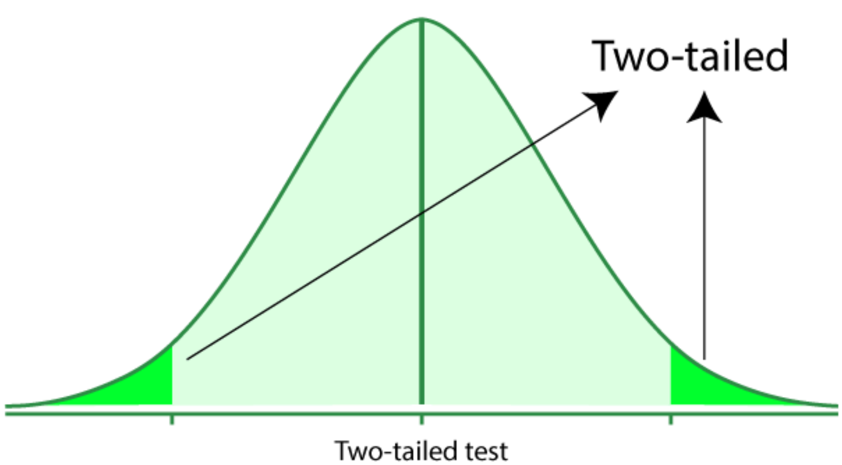
- For a right-tailed test: $\theta > \theta 0$
- For a left-tailed test: $\theta < \theta 0$

Test Statistic: Depending on the type of test and the distribution, the test statistic is computed (for example *t*-score for normal distribution).
Decision Rule: If the test statistic falls in the critical region, reject the null hypothesis in favor of the alternative hypothesis.
Example: Effect of participants of students in coding competition on their fear level.

- H0: There is no important effect of students in coding competition on their fear level.

The main intention is to check the **decreased** fear level when students participate in a coding competition.

Two-Tailed Test
A two-tailed test is also called a non directional hypothesis. For checking whether the sample is greater or less than a range of values, we use the two-tailed. It is used for null hypothesis testing.



Difference Between One-Tailed and Two-Tailed Tests

Null Hypothesis (*H*0): where represents a parameter (e.g., population mean) and $\theta 0$ is a specific value.

Alternative Hypothesis ($H1$): $\theta \neq \theta 0$

Test Statistic: Compute the test statistic as appropriate for the distribution (for example t-score for normal distribution).

Decision Rule: If the test statistic falls in either tail of the distribution's critical region, reject the null hypothesis in favor of the alternative hypothesis.

Example: Effect of new bill pass on the loan of farmers.

- H0: There is no significant effect of the new bill passed on loans of farmers.

New bill passes can affect in both ways either **increase or decrease** the loan of farmers.

Difference Between One and Two-Tailed Test:

One-Tailed Test	Two-Tailed Test
A test of any statistical hypothesis, where the alternative hypothesis is **one-tailed** either right-tailed or left-tailed.	A test of a statistical hypothesis, where the alternative hypothesis is **two-tailed**.
For one-tailed, we use either > or < sign for the alternative hypothesis.	For two-tailed, we use ≠ sign for the alternative hypothesis.
When the alternative hypothesis specifies a direction then we use a one-tailed test.	If no direction is given then we will use a two-tailed test.
Critical region lies entirely on either the right side or left side of the sampling distribution.	Critical region is given by the portion of the area lying in both the tails of the probability curve of the test statistic.

One-Tailed Test	Two-Tailed Test
Here, the Entire **level of significance (α)** i.e. 5% has either in the left tail or right tail.	It splits the **level of significance (α)** into half.
Rejection region is either from the left side or right side of the sampling distribution.	Rejection region is from both sides i.e. left and right of the sampling distribution.
It checks the relation between the variable in a singles direction.	It checks the relation between the variables in any direction.
It is used to check whether the one mean is different from another mean or not.	It is used to check whether the two mean different from one another or not.

Types of T-Test

There are three types of t-test

- One Sample T-test
- Independent Samples T-test
- Paired Samples T-test

One Sample T-Test

As the name implies, this test is used when we have one data set for a sample and we need to determine whether this data set belongs to a particular population or not. The mean value for the population data must be known in this case. The formula to determine T-value, in this case, is as follows:

$t = (\bar{x} - \mu) / (\sigma / \sqrt{n})$

Where,

- ***t*** *is the t-value,*
- ***$\bar{x}$*** *is the Sample mean,*
- ***μ*** *is the Population mean,*
- ***σ*** *is the Sample standard deviation, and*
- ***n*** *is the Sample size.*

Steps to Calculate T Value One Sample T-Test

To perform the One Sample T-test, the steps listed below are generally followed:

Step 1: *State a null hypothesis and an alternative hypothesis. The null hypothesis assumes that the sample mean and the known population mean (μ) are equal, while the other assumes that the sample mean is different from the population mean.*

Step 2: *Define values for the level of significance (α) and the degree of freedom (df). The degree of freedom equals (n – 1) for this case.*

Step 3: *Calculate the t-value using the formula stated above by putting all the known values of the sample mean ($\bar{x}$), sample standard deviation (σ), the population mean (μ), and the sample size (n).*

Step 4: *Determine the associated t-value using a t-distribution table.*

Step 5: *Compare the t-value to the calculate t-value. If the calculated t-value is greater than the tabulate t value , reject the null hypothesis and conclude that the sample mean is significantly different from the population mean. Otherwise, conclude that there is no significant difference between the sample mean and the population mean.*

Independent Samples T-Test

As the name suggests, an Independent samples T-test is used when we need to compare the statistical means of two independent samples or groups. It helps us determine whether there is a significant difference between the means of the two groups. If there is a significant difference, it suggests that the groups likely have different population means; otherwise, they have the same population means.

T Test for Independent Samples

The steps listed below are generally followed to perform this test:

Step 1: *State a null hypothesis and an alternate hypothesis. The null hypothesis assumes that the means of the two groups are equal ($\bar{x}1 = \bar{x}2$), while the other assumes that the means of the two groups are significantly different ($\bar{x}1 \neq \bar{x}2$).*

Step 2: *Define the values for the level of significance (α) and the degrees of freedom (df). The degree of freedom equals ($n1 + n2 – 2$) in this case.*

Step 3: *Calculate the t-value from the formula after obtaining the required data related to each group.*

$$t =((\bar{x}1 – \bar{x}2)/ s)* \sqrt{}n1n2/n1+n2$$

Step 4: *Find the critical t-value from a t-distribution table with the corresponding degrees of freedom and level of significance.*

Step 5: *If the calculated t-value is greater than the critical t-value, then reject the null hypothesis. This indicates that there is a significant difference between the means of the two groups. Otherwise, the null hypothesis is not rejected. And, this suggests that there is no significant difference between the means of the two groups.*

Paired Samples T-Test

The Paired samples t-test is used when we want to compare the means of two related groups or samples. For example, we may use this test to compare the average scores of the players of an athletics team before and after a training program. To calculate the t-value in this case, the following formula is used,

D = Differences between two paired samples

di = The ith observation in D

n = The sample size

$\overline{d}$ = The sample mean of the differences

σ^ = The sample standard deviation of the differences

T =The critical value of a t-distribution with (n − 1) degrees of freedom

t = The t-statistic (t-test statistic) for a paired sample t-test

The four steps are listed below:

1. Calculate the sample mean.

   $\overline{d} = d1 + d2 + \cdots + dn$

2. Calculate the sample standard deviation.

   $S=\sqrt{\sum(d- \overline{d})^2 / n-1}$

3. Calculate t value

$t = (\bar{d} / s)\sqrt{n}$

Steps for Paired Samples T-Test

Following are the steps to perform this type of T-test:

Step 1: *State the null hypothesis which assumes that there is no significant difference between the statistical means of the paired observations (μd = 0) while the alternative hypothesis assumes that there is a significant difference between the statistical means of the paired observations (μd ≠ 0).*

Step 2: *Match each observation in one group with a corresponding observation in the other group.*

Step 3: *Calculate the differences between each paired observation and then, calculate the mean of the differences and the sample standard deviation of the differences Furthermore, calculate the t-value from the formula.*

Step 4: *Obtain the critical t-value from a t-distribution table corresponding to the chosen level of significance (α) and degree of freedom (df). The degree of freedom (df) equals (n – 1) in this case.*

Step 5: *If the calculated t-value is greater than the critical t-value, then reject the null hypothesis. This indicates a significant difference in the sample before and after the intervention. Otherwise, it can be concluded that there is no significant difference in the sample before and after the intervention.*

Degree of Freedom

The degrees of freedom that are mathematical concepts to statistical calculation represents the number of variables that have the freedom to vary in a calculation. Calculating degrees of freedom can help ensure the validity of t-tests, and highly f-tests, among other tests. These tests are often used to compare data that has been detected with data that would be expected if a particular hypothesis were true.

The fact that the statistical degrees of freedom indicating the number of values in the final calculation is allowed to vary means that they can contribute to the validity of the result. Although the number of observations and parameters to be measured depends on the size of the sample, or the number of observations, and the parameters to be measured, the degree of freedom in the calculations is usually equal to the value of the observations minus the number of parameters. This means that for larger sample size, there are degrees of freedom available.

For Example

Mention that we have seven shirts that we can wear for a week, and we decide to wear each shirt only once a week.

On Sunday, Consider choosing 1 of the 7 shirts. Wear any of the 7 shirts. On the second day, the shirt worn on the first day cannot be selected, and we should choose from the remaining shirts. The pattern continues as follows:

1. Sunday: 7 shirts to choose from
2. Monday: 6 shirts to choose from
3. Tuesday: 5 shirts to choose from
4. Wednesday: 4 shirts to choose from
5. Thursday: 3 shirts to choose from
6. Friday: 2 shirts to choose from
7. Saturday: 1 shirt to choose from

On the last day, Saturday, there is only one shirt to choose from, which means, in fact, there is no choice. Put it in different names, we are forced on Saturday by our choice of which shirt to wear. In this one week, we have to choose one shirt a day, we have six free days to choose a shirt. It is the same as saying that our choice of shirt is restricted for one day. So, this week, there are six levels of freedom.

Understanding the Degrees of Freedom

An easy way to understand the degrees of mental freedom is by using an example:

- Consider a sample of data that combines, in order to simplify, five positive numbers. Values can be any number that does not have a known relationship between them. This data sample, theoretically, can have up to five degrees of freedom.
- The four numbers in the sample are {3, 8, 5, and 4} and the total number of data samples is expressed as 6.
- This should mean that the fifth number should be 10. It can't be any other. It does not have the freedom to be different.
- So the freedom degrees of this data sample are 4.

The free degree formula is equal to the size of a sample of data except one:
$Df=N-1$

Where as;
Df=Degrees of Freedom

N= Actual Sample size
Degrees of freedom are often discussed in relation to various methods of hypothesis testing in mathematics, such as t-test. It is important to calculate

degrees of freedom when trying to understand the importance of the t test arithmetic and the validity of the null hypothesis.

Degrees of Freedom Formula

The statistical formula to find out how many degrees of freedom are there is quite simple. It implies that degrees of freedom is equivalent to the number of values in a data set minus 1, and appears like this:

$df=N-1$

Where N represents the number of values in the data set (sample size).

That being said, let's have a look at the sample calculation.
If there is a data set of 6, (N=6).
Call the data set X and make a list with the values for each data.
For this example data, set X of the sample size includes: 10, 30, 15, 25, 45, and 55

This data set has a mean, or average of 30. Find out the mean by adding the values and dividing by N:

$(10 + 30 + 15 + 25 + 45 + 55)/6= 30$

Using the formula, the degrees of freedom will be computed as d_f = N-1:

In this example, it appears, $d_f = 6-1 = 5$

This further implies that, in this data set (sample size), five numbers contain the freedom to vary as long as the mean remains 30.

Application of the Degree of Freedom

Although the level of freedom is a vague and often overlooked concept in mathematics, it is very effective in the real world.

For example, business owners who want to hire employees to produce a product face two changes - function and effect. Additionally, the relationship between employees and output (i.e., the amount of product that an employee can produce) is a liability.

In such a case, the business owners may determine the amount of product to be produced, which may determine the number of employees to be employed, or the number of employees, which may be sufficient for the product to be produced. So, in terms of output and staff, owners have one level of freedom.

Normal Distribution

We define Normal Distribution as the probability density function of any continuous random variable for any given system. Now for defining Normal

Distribution suppose we take f(x) as the probability density function for any random variable X

Normal Distribution Curve

In any Normal Distribution, random variables are those variables that take unknown values related to the distribution and are generally bound by a range. An example of the random variable is , suppose take a distribution of the height of students in a class then the random variable can take any value in this case but is bound by a boundary of 2 ft to 6 ft, as it is generally forced physically.

Range of any normal distribution can be infinite in this case we say that normal distribution is not bothered by its range. In this case, range is extended from –∞ to + ∞.

Bell Curve still exist, in that case, all the variables in that range are called Continuous variable and their distribution is called Normal Distribution as all the values are generally closed aligned to the mean value. The graph or the curve for the same is called the **Normal Distribution Curve** Or Normal Distribution Graph.

Normal Distribution Standard Deviation

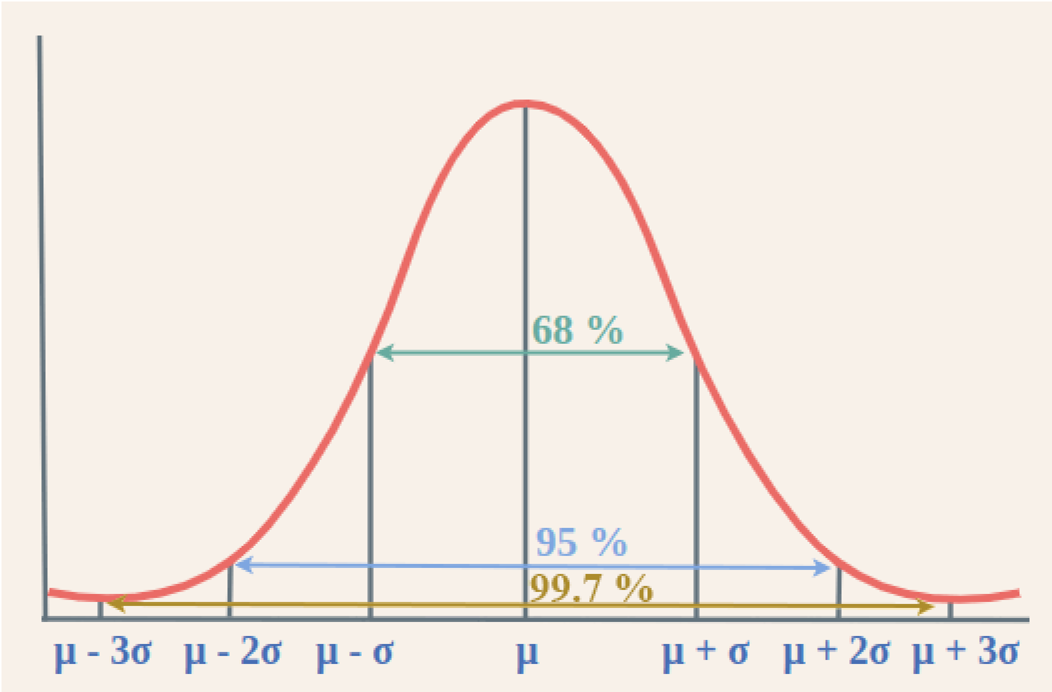
We know that mean of any data spread out as a graph helps us to find the line of the symmetry of the graph whereas, Standard Deviation tells us how far the data is spread out from the mean value on either side. For smaller values of the standard deviation, the values in the graph come closer and the graph becomes narrower. While for higher values of the standard deviation the values in the graph are dispersed more and the graph becomes wider.

Empirical Rule of Standard Deviation

Generally, the normal distribution has a positive standard deviation and the standard deviation divides the area of the normal curve into smaller parts and each part defines the percentage of data that falls into a specific region This is called the Empirical Rule of Standard Deviation in Normal Distribution.

Empirical Rule states that,

- 68% of the data approximately fall within one standard deviation of the mean, i.e. it falls between {Mean – One Standard Deviation, and Mean + One Standard Deviation}
- 95% of the data approximately fall within two standard deviations of the mean, i.e. it falls between {Mean – Two Standard Deviation, and Mean + Two Standard Deviation}
- 99.7% of the data approximately fall within a third standard deviation of the mean, i.e. it falls between {Mean – Third Standard Deviation, and Mean + Third Standard Deviation}



Studying the graph it is clear that using Empirical Rule we distribute data broadly in three parts. And thus, **empirical rule** is also called **"68 – 95 – 99.7" rule.**.

Properties of Normal Distribution

Some important properties of normal distribution are,

- For normal distribution of data, mean, median, and mode are equal, (i.e., **Mean = Median = Mode).**
- Total area under the normal distribution curve is equal to 1.
- Normally distributed curve is symmetric at the center along the mean.
- In a normally distributed curve, there is exactly half value to the right of the central and exactly half value to the right side of the central value.
- Normal distribution is defined using the values of the mean and standard deviation.
- Normal distribution curve is a Unimodal Curve, i.e. a curve with only one peak.

Student's t-distribution in Statistics

As we know normal distribution assumes two important characteristics about the dataset: a large sample size and knowledge of the population standard deviation. However, if we do not meet these two criteria, and we have a small sample size or an unknown population standard deviation, then we use the t-distribution.

What is t-distribution?

Student's t-distribution, also known as the t-distribution, is a probability distribution that is used in statistics for making inferences about the population mean when the sample size is small or when the population standard deviation is unknown. It is similar to the standard normal distribution (Z-distribution), but it has heavier tails. Theoretical work on t-distribution was done by **W.S. Gosset**; he has published his findings under the pen name "**Student**". That's why it is called a **Student's t-test**. The t-score represents the number of standard deviations the sample mean is away from the population mean.

T-Score

The T-score, also known as the t-value or t-statistic, is a standardized score that quantifies how many standard deviations a data point or sample mean is from the population mean. It is commonly used in statistical hypothesis testing, particularly in scenarios where the sample size is small or the population standard deviation is unknown.

The formula for calculating the T-score in the context of a t-distribution is given by:

$t= (\bar{x}-\mu / s)* \sqrt{n}$

- t = t-score,
- $\bar{x}$ = sample mean
- μ = population mean,
- s = standard deviation of the sample,
- n = sample size

As we know, we use t-distribution when the standard deviation of the population is unknown and the sample size is small. The formula for the t-distribution looks very similar to the normal distribution; the only difference is that instead of the standard deviation of the population, we will use the standard deviation of the sample.

When to Use the t-Distribution?

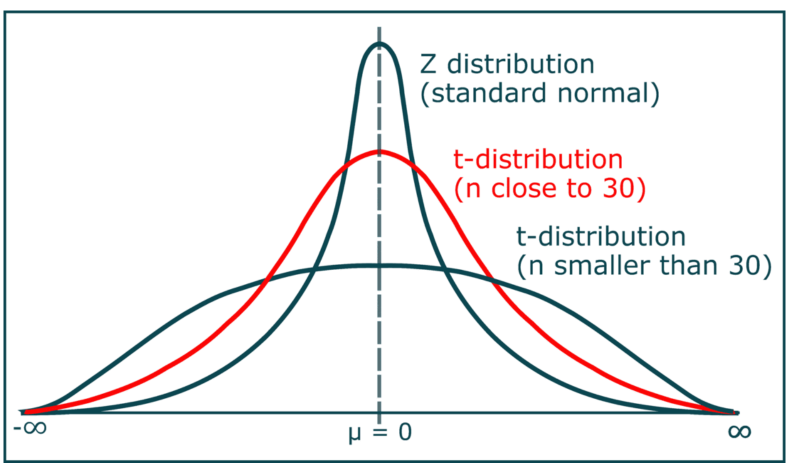
Student's t Distribution is used when :

- The sample size is 30 or less than 30.
- The population standard deviation(σ) is unknown.
- The population distribution must be unimodal and skewed.

Properties of the t-Distribution

- The variable in t-distribution ranges from $-\infty$ to $+\infty$ ($-\infty < t < +\infty$).

- t- distribution will be symmetric like the normal distribution if the power of t is even in the probability density function(pdf).
- For large values of ν(i.e. increased sample size n); the t-distribution tends to a standard normal distribution. This implies that for different ν values, the shape of t-distribution also differs.
- The t-distribution is less peaked than the normal distribution at the center and higher peaked in the tails. From the above diagram, one can observe that the red and green curves are less peaked at the center but higher peaked at the tails than the blue curve.
- The value of y(peak height) attains highest at $\mu = 0$ as one can observe the same in the above diagram.
- The mean of the distribution is equal to 0 for $\nu > 1$ where ν = degrees of freedom, otherwise undefined.
- The median and mode of the distribution is equal to 0.
- The variance is equal to $\nu / \nu\text{-}2$ for $\nu > 2$ and ∞ for $2 < \nu \leq 4$ otherwise undefined.



Degrees of freedom refer to the number of independent observations in a set of data. When estimating a mean score or a proportion from a single sample, the number of independent observations is equal to the sample size minus one.

Hence, the distribution of the t statistic from samples of size 10 would be described by a t distribution having 10 – 1 or 9 degrees of freedom Similarly, a t- distribution having 15 degrees of freedom would be used with a sample of size 16.

t-Distribution Table

t-Distribution table gives the t-value for a different level of significance and different degrees of freedom. The calculated t-value will be compared with the tabulated t-value. For example, if one is performing a student's t-test and for that performance, he has taken a 5% level of significance and he got or calculated t-value and he has taken his tabulated t-value and if the calculated t-value is higher than the tabulated t-value, in that case, it will say that there is a significant difference between the population mean and the sample means at 5% level of significance and if vice versa then, in that case, it will say that there is no significant difference between the population means and the sample means at 5% level of significance.

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

t-scores and p-values

t-scores :

- It represents the deviation of a data point from the mean in a t-distribution, expressed in terms of standard deviations. Particularly useful for small sample sizes or cases with unknown population standard deviations.
- We can obtain them from a t-table or through online tools, providing a numerical measure of how a typical a data point is within the distribution.

- t-score is important in determining confidence intervals, aiding in estimating the range within which the true population parameter is likely to fall. The critical value of t is integral in confidence interval calculations, guiding the determination of upper and lower bounds.

p-value:

The p-value (probability value) is a statistical measure that helps assess the evidence against a null hypothesis.

- p-value describes the likelihood of data occurring if the null hypothesis were true.
- We can use statistical software to directly obtain the p-value associated with the calculated t-score or we can use the t-table, which provides critical values for different levels of significance and degrees of freedom. First, find the row corresponding to our degrees of freedom and the column corresponding to our t-score to get the p-value.

Limitations of Using a T-Distribution

- **Sensitivity to Departure from Normality:** The t-distribution assumes normality in the underlying population. When data deviates significantly from a normal distribution, reliance on the t-distribution may introduce inaccuracies in statistical inferences.
- **Limited Applicability for Large Samples:** As sample sizes increase, the t-distribution converges to the normal distribution. Therefore, for sufficiently large samples and known population standard deviation, the normal distribution is more appropriate, and using the t-distribution may not offer additional benefits.

- **Impact of Outliers and Small Sample Sizes:** The t-distribution can be sensitive to outliers, and its tails can be influenced by small sample sizes. Outliers may distort results, and in cases where the sample size is very small, the t-distribution may have heavier tails, affecting the accuracy of inferences.
- **Requires Random Sampling:** The assumptions underlying the t-distribution, such as random sampling and independence of observations, need to be met for valid results. If these assumptions are violated, the accuracy of inferences drawn from the t-distribution may be compromised.

T- Distribution Applications

1. **Testing for the Hypothesis of the Population Mean:** T-distributions are commonly used in hypothesis tests regarding the population mean. This involves assessing whether a sample mean is significantly different from a hypothesized population mean.
2. **Testing for the Hypothesis of the Difference Between Two Means:** T-tests can be employed to examine if there is a significant difference between the means of two independent samples. This can be done under the assumption of equal variances or when variances are unequal.In scenarios where samples are not independent, such as paired or dependent samples, t-tests can be used to assess the significance of the mean difference between related observations.
3. **Testing for the Hypothesis about the Coefficient of Correlation:** T-distributions play a role in hypothesis testing related to correlation coefficients. This includes situations where the population correlation coefficient is assumed to be zero (ρ=0) or when testing for a non-zero correlation coefficient ($\rho \neq 0$).

Difference Between T-Distribution and Normal Distribution

T-Distribution	**Normal Distribution**
T-Distribution is defined by its degree of freedom which itself depends upon the sample size	Normal distribution is defined by its mean and standard deviation
T- distribution is used when the sample size is small	Normal distribution is used when we have large no data points in the dataset
It has a heavier tail than normal distribution which means more data points are away from the mean of the distribution	Normal distribution has a lighter tail than T-distribution which means more data points lie near the mean of the distribution
We use T-distribution in hypothesis testing when the standard variation of the population is unknown	Normal distribution is used when the standard deviation is known
T-Distribution has a larger range of critical values as compared to the normal distribution as this distribution has heavier tails	Normal distribution has a smaller range as compared to t-distribution