Statistics

z-Statistics

It is used when population means and standard deviations are known. The formula of z-statistics is given by:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

where

- $\bar{x}$ is the sample mean,
- μ represents the population mean,
- σ is the standard deviation
- and n is the size of the sample.

z-Statistics

Q1. A company manufactures light bulbs and claims that the average lifespan of their bulbs is 1000 hours. A consumer group wants to test this claim. They randomly sample 64 bulbs and find that the sample mean lifespan is 980 hours. Assume the population standard deviation is known to be 100 hours.

1. State the hypotheses:

- **Null hypothesis (H0):** The average lifespan of the bulbs is 1000 hours. ($\mu = 1000$)
- **Alternative hypothesis (Ha):** The average lifespan of the bulbs is not 1000 hours. ($\mu \neq 1000$)

2. Determine the level of significance:

- Let's choose a significance level of 0.05 (alpha = 0.05). This means we are willing to accept a 5% chance of rejecting the null hypothesis when it is actually true.

3. Calculate the test statistic:

- The formula for the z-test statistic is: $z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$
  - Where:
    - $\bar{x}$ is the sample mean (980 hours)
    - μ is the population mean under the null hypothesis (1000 hours)
    - σ is the population standard deviation (100 hours)
    - n is the sample size (64)
- Plugging in the values: $z = (980 - 1000) / (100 / \sqrt{64}) = -1.6$

4.Find the critical value:

• This is a two-tailed test (Ha: $\mu \neq 1000$)

• For a two-tailed test with alpha = 0.05, the critical values are approximately ±1.96.

5.Make a decision:

• Compare test statistic (-1.6) to the critical values (±1.96).

•Since the absolute value of the test statistic (1.6) is less than the critical value (1.96), we fail to reject the null hypothesis.

Conclusion:

At a 0.05 significance level, there is not enough evidence to conclude that the average lifespan of the light bulbs is different from 1000 hours.

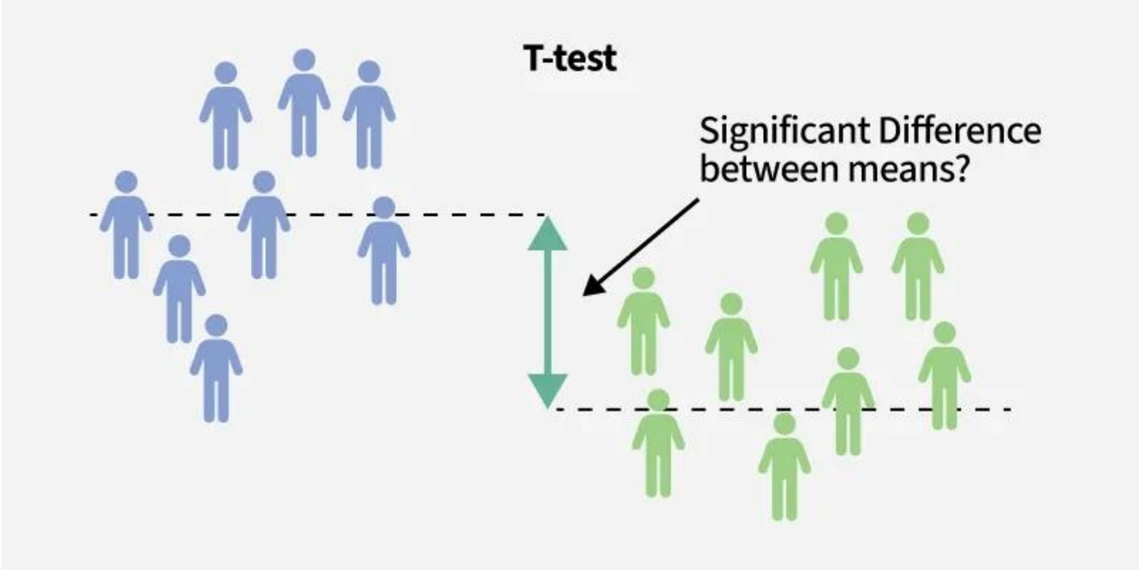
T-Statistics

T-test is used to compare the means of two datasets (e.g., experimental vs. control groups) to assess if the difference is statistically significant. It is used when n<30 t-statistic calculation is given by:

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

where

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



T-Statistics

Suppose You want to compare the test scores of two groups of students:

•Group 1: 30 students who studied with Method A.

•Group 2: 30 students who studied with Method B.

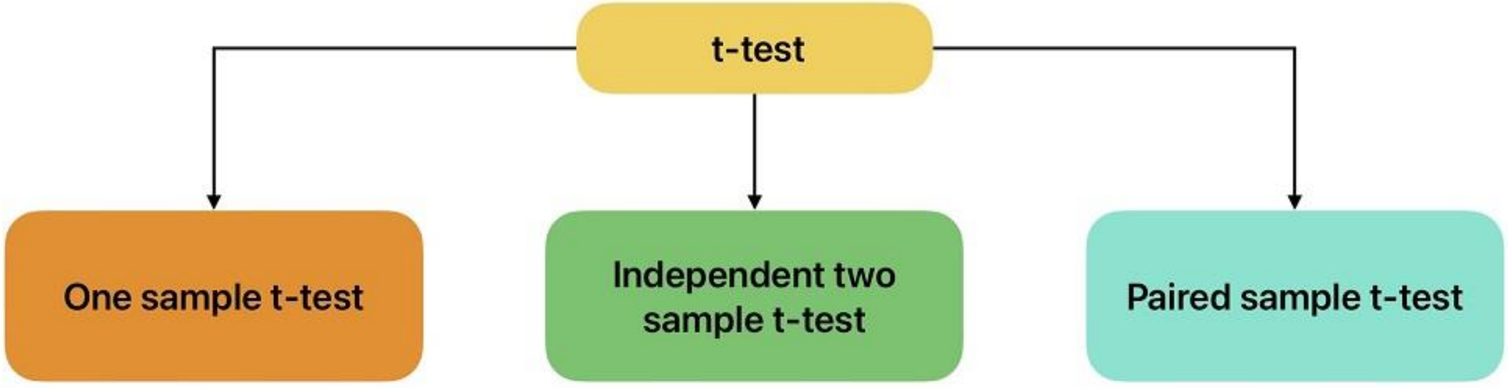
You use a **t-test** to check if there is a significant difference in the average test scores between the two.

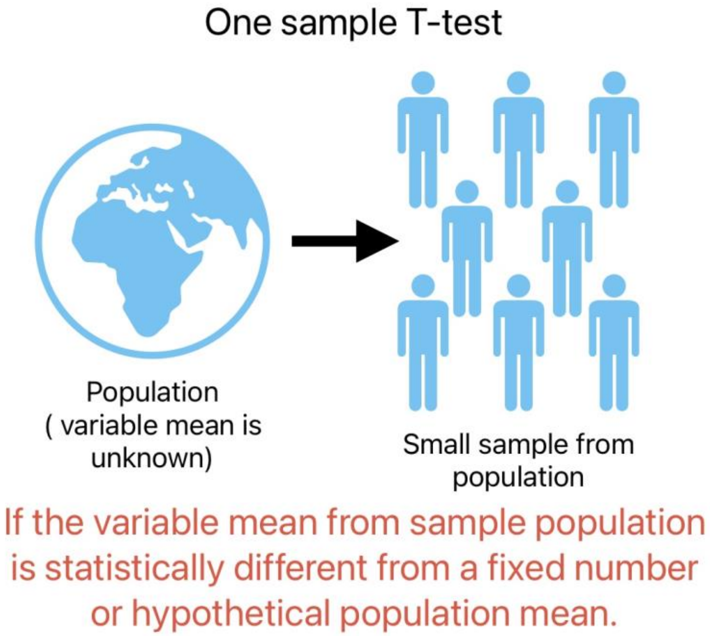
The t-test is part of **hypothesis testing** where you start with an assumption the null hypothesis that the two-group means are the same. Then the test helps you decide if there's enough evidence to reject that assumption and conclude that the groups are different.

T-Statistics

- **Degree of freedom (df):** The degree of freedom tells us the number of independent variables used for calculating the estimate between 2 sample groups.
  In a t-test the degree of freedom is calculated as the total sample size minus 1 i.e. $df=\sum ns-1$, where "n_s" is the number of observations in the sample. Suppose, we have 2 samples A and B. The df would be calculated as $df=(nA-1)+(nB-1)$
- **Significance Level:** The significance level is the predetermined threshold that is used to decide whether to reject the null hypothesis. Commonly used significance levels are 0.05, 0.01, or 0.10.
- **T-statistic**: The t-statistic is a measure of the difference between the means of two groups. It is calculated as the difference between the sample means divided by the standard error of the difference. It is also known as the t-value or t-score.
  - If the t-value is large => the two groups belong to different groups.
  - If the t-value is small => the two groups belong to the same group.

T-Statistics





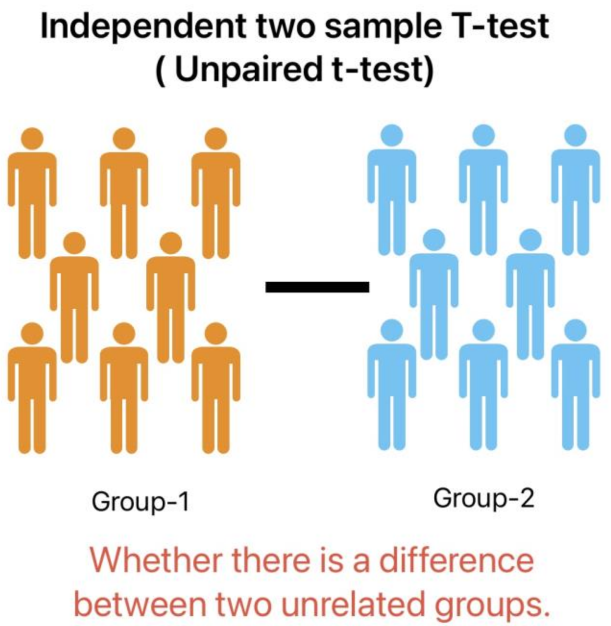
One Sample T-Test

The value against which we are comparing is a single value, i.e. we compare the mean of a sample with a single value to check how much the mean deviates from that single value.

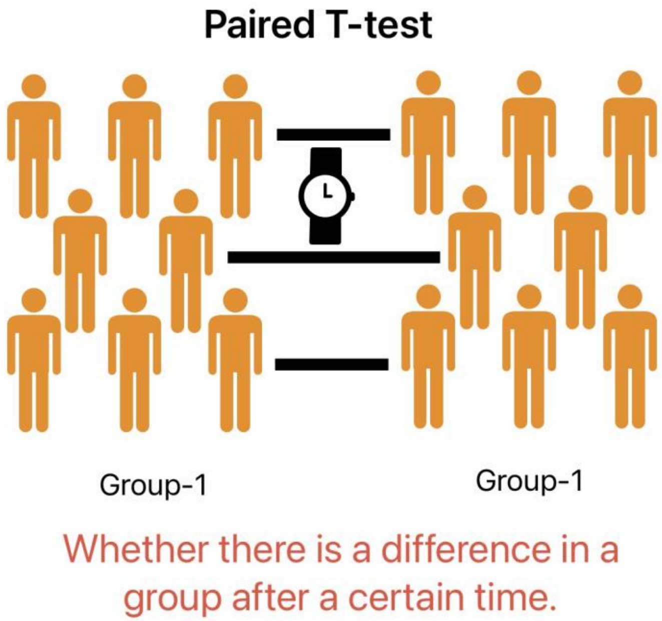
Two Sample T-Test

We compare the means and variances of two samples, we assess how much they differ.

T-Statistics



In an **independent two-sample t-test** (unpaired t-test), the samples in the two groups being compared are unrelated. The samples are drawn from two different populations or groups of subjects, and the difference between the means of the two groups is calculated using the means and variances of the two separate samples.



In a **dependent two-sample t-test** (also known as a **paired t-test**), the samples in the two groups being compared are related in some way. For example, the samples may be pairs of measurements taken on the same subjects. In this case, the difference between the means of the two groups is calculated by taking the differences between the pairs of measurements and treating these differences as a single sample.

T-Statistics

One-Sample T-Test

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$\bar{x}$ = obersved mean of the sample
μ = assumed mean
s = standard deviation
n = sample size

Two-Sample T-Test

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\bar{x}_1$ = observed mean of 1st sample
$\bar{x}_2$ = observed mean of 2nd sample
s_1 = standard deviation of 1st sample
s_2 = standard deviation of 2nd sample
n_1 = sample size of 1st sample
n_2 = sample size of 2nd sample

Paired sample T-test

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

where: sd = sqrt[Σ(di - d̄)^2 / (n - 1)]

- $\bar{d}$ is the mean of the difference scores
- s_d is the standard deviation of the difference scores
- n is the number of pairs of observations

di: The difference between the paired measurements for the i-th participant (After - Before for the i-th participant)
d̄: The mean difference (average of all the di's)

T-Statistics

Q1. A manufacturer claims that the average weight of their product is 50 grams. You want to test this claim. You randomly sample 25 products and find the following weights (in grams):

48, 52, 49, 51, 50, 47, 53, 50, 49, 52, 48, 51, 50, 49, 50, 51, 48, 52, 49, 50, 51, 47, 53, 50, 49

1. **State the hypotheses:**
   - Null hypothesis (H0): The population mean is equal to 50 grams ($\mu = 50$).
   - Alternative hypothesis (Ha): The population mean is not equal to 50 grams ($\mu \neq 50$).
2. **Calculate the sample mean ($\bar{x}$) and sample standard deviation (s):**
   - $\bar{x}$ = (sum of all weights) / (number of samples) = 1245 / 25 = 49.8 grams
   - s = (square root of [sum of (each weight - $\bar{x}$)^2] / (number of samples - 1)) ≈ 1.83 grams
3. **Calculate the t-statistic:**
   - $t = (\bar{x} - \mu) / (s / \sqrt{n}) = (49.8 - 50) / (1.83 / \sqrt{25}) \approx -0.55$
4. **Determine the degrees of freedom (df):**
   - df = n - 1 = 25 - 1 = 24
5. **Find the critical value:**
   - You need to choose a significance level (alpha). Let's say $\alpha = 0.05$.
   - Consult a t-distribution table or use a calculator to find the critical value for a two-tailed test with df = 24 and $\alpha = 0.05$. The critical value is approximately ±2.064.

Since our calculated t-statistic (-0.55) falls within the range of -2.064 to +2.064, we fail to reject the null hypothesis.

Conclusion: There is not enough evidence to conclude that the average weight of the products is different from 50 grams.

t-test 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-Statistics

Q2. A researcher wants to know whether there is a significant difference in the average test scores of students who are taught using two different methods. The researcher randomly assigns 20 students to one of two groups. Group A is taught using method A, and group B is taught using method B. After the students have completed the course, they are given a test. The test scores are shown below:

Group A: 85, 90, 92, 88, 89, 91, 93, 87, 86, 94

Group B: 78, 82, 80, 85, 83, 81, 79, 84, 77, 86

Solution:

- t-test is used to determine whether there is a significant difference in the average test scores of the two groups.
- The null hypothesis is that there is no significant difference in the average test scores of the two groups.
- The alternative hypothesis is that there is a significant difference in the average test scores of the two groups.
- Calculates the mean and standard deviation of the test scores for each group.

Group A: mean = 89.5, standard deviation = 2.87

Group B: mean = 81.5, standard deviation = 2.87

T-Statistics

- The null hypothesis is that there is no significant difference in the average test scores of the two groups.
- The alternative hypothesis is that there is a significant difference in the average test scores of the two groups.
- Calculates the mean and standard deviation of the test scores for each group.

Group A: mean = 89.5, standard deviation = 2.87 **Group B:** mean = 81.5, standard deviation = 2.87

- Calculate t-score using two sample t-test formula:

t = (89.5 - 81.5) / sqrt((2.87^2 / 10) + (2.87^2 / 10)) = 5.57

- Degree of freedom df = =(nA−1)+(nB−1) = 18

Let, the significance level is 0.05, the critical value is obtained from the t-table as 2.101.

Since the t-statistic (5.57) is greater than the critical value (2.101), the researcher rejects the null hypothesis. This means that there is a significant difference in the average test scores of the two groups.

t-test 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-Statistics

Q3. A researcher wants to study the effectiveness of a new memory-enhancing drug. They recruit 20 participants and administer a memory test before and after taking the drug for a month. Is there a statistically significant difference in memory test scores before and after taking the drug?

Solution: this is a paired t-test question, as we have two measurements (before and after) for each participant.

1.Calculate the Differences: Subtract the "Before Drug" score from the "After Drug" score for each participant.

2.Calculate the Mean Difference ($\bar{d}$): Sum the differences and divide by the number of participants (n = 20).

$\bar{d}$ = (7+7+5+6+5+2+7+5+5+5+7+5+6+6+4+4+7+5+5+6) / 20 = 5.35

3.Calculate the Standard Deviation of the Differences (sd):

sd = sqrt[$\Sigma(di - \bar{d})^2$ / (n - 1)] **= 1.35**

4. Calculate the t-statistic: t = ($\bar{d}$) / (sd / $\sqrt{n}$) = 5.35 / (1.35 / $\sqrt{20}$) ≈ 17.7

5. Determine the Degrees of Freedom: df = n - 1 = 20 - 1 = 19

6. Find the Critical Value: for alpha 0.05 and df =19, critical value is ±2.093. Since the absolute value of our t-statistic is greater than the critical value, we reject the null hypothesis.

Participant	Before Drug	After Drug	Difference (After - Before)
1	75	82	7
2	68	75	7
3	80	85	5
4	72	78	6
5	85	90	5
6	70	72	2
7	78	85	7
8	65	70	5
9	90	95	5
10	75	80	5
11	72	79	7
12	68	73	5
13	82	88	6
14	77	83	6
15	88	92	4
16	71	75	4
17	79	86	7
18	66	71	5
19	84	89	5
20	73	79	6