SWAYAM NPTEL COURSE ON

MINE AUTOMATION AND DATA ANALYTICS

By

Prof. Radhakanta Koner

Department of Mining Engineering

Indian Institute of Technology (Indian School of Mines) Dhanbad

Module 9 : Hypothesis Testing

Lecture 22A : t - test

CONCEPTS COVERED

1) Introduction to t-test (for Population unknown variance)
   - Motivation Example about t-test
   - Assumptions of t-test

2) One Sample t-test

3) Two Sample t-test.
   (i) Assuming Populations with equal Variance Case
   (ii) Assuming Populations with unequal Variance Case

Overview of t test: Mathematically

$X1, X2, X3, \ldots, Xn \sim$ iid Normal (μ, σ^2)

σ^2 is unknown

$E(X) = \mu$; $Var(X) = \sigma^2$

Testing for mean,

Null hypothesis $(H_0) : \mu = \mu0$,

Alternative hypothesis $(H_A) : \mu > \mu0$

t-test

- T-test is a statistical method used to determine if there is a significant difference between the means of two groups or between the mean of a sample and a known value.

- It helps you assess whether any observed differences between the groups are likely to have occurred by chance or if they are statistically significant.

- The t-test is based on the t-distribution, which is a mathematical distribution similar to the normal distribution but with heavier tails.

- The test calculates a t-statistic, which measures the difference between the means of the two groups in terms of the standard error of the difference.

- The larger the t-statistic, the more likely it is that the difference between the groups' means is not due to random chance.

When Should We Perform a t-test?

Let’s first understand where a t-test can be used before we dive into its different types and their implementations. The best way to learn a concept is by visualizing it through an example. So, let’s take a simple example to see where a t-test comes in handy.

Consider a telecom company that has two service centers in the city. The company wants to find out whether the average time required to service a customer is the same in both stores.

The company measures the average time taken by 50 random customers in each store. Store A takes 22 minutes, while Store B averages 25 minutes. Can we say that Store A is more efficient than Store B in terms of customer service?

It does seem that way, doesn’t it? However, we have only looked at 50 random customers out of the many people who visit the stores. Simply looking at the average sample time might not be representative of all the customers who visit both stores.

This is where the t-test comes into play. It helps us understand if the difference between two sample means is actually real or simply due to chance.

Assumptions for Performing a t-test

There are certain assumptions we need to heed before performing a t-test:

- The data should follow a continuous or ordinal scale (the IQ test scores of students, for example)
- The observations in the data should be randomly selected
- The data should resemble a bell-shaped curve when we plot it, i.e., it should be normally distributed.
- Large sample size should be taken for the data to approach a normal distribution (although a t-test is essential for small samples as their distributions are non-normal)
- Variances among the groups should be equal (for independent two-sample t-test).
- Variances among the groups should not be equal (for Welch's t-test).

One sample t-test

The one-sample t-test is a statistical method used to determine whether the mean of a single sample is statistically different from a known or hypothesized population mean.

Background:
The one-sample t-test is employed when you have a single sample and want to assess whether its mean differs significantly from a known or hypothesized population mean.

It's particularly useful in situations where you're interested in evaluating the effectiveness of a treatment, comparing sample data to a theoretical expectation, or testing a hypothesis about a population parameter.

One sample t-test

Hypotheses:

The null hypothesis (H_0) typically states that there is no difference between the sample mean and the population mean, while the alternative hypothesis (H_1) states that there is a significant difference.

- **Null Hypothesis ($H0$): $\mu = \mu_0$ (The sample mean is equal to the population mean)**

- **Alternative Hypothesis ($H1$): $\mu \neq \mu_0$ (The sample mean is not equal to the population mean)**

One sample t-test

Test Statistic:

The test statistic for the one-sample t-test is calculated as:

$$t = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Where:

$\bar{X}$ is the sample mean.

μ_0 is the population mean under the null hypothesis.

s is the sample standard deviation.

n is the sample size.

One sample t-test

Assumptions:

Random Sampling: The sample is randomly selected from the population.

Normality: The data are approximately normally distributed or the sample size is large enough for the Central Limit Theorem to apply.

Independence: The observations in the sample are independent of each other.

One sample t-test

Decision Rule:

To make a decision about the null hypothesis,

- we compare the calculated t-value to the critical t-value from the t-distribution with $n-1$ degrees of freedom, where n is the sample size.

- Alternatively, we can use the p-value associated with the test statistic.

One sample t-test

Conclusion:

- **If the calculated t-value is greater than the critical t-value (or if the p-value is less than the significance level, commonly 0.05), we reject the null hypothesis. This indicates that there is a statistically significant difference between the sample mean and the population mean.**

- **If the calculated t-value is less than the critical t-value (or if the p-value is greater than the significance level), we fail to reject the null hypothesis. This suggests that there is insufficient evidence to conclude that there is a difference between the sample mean and the population mean.**

Interpretation:

- **If we reject the null hypothesis, it means that the observed sample mean is unlikely to have occurred by random chance alone, and there is evidence to support the alternative hypothesis, indicating a difference between the sample mean and the population mean.**

t-test critical values

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										

Example of one sample t-test

Example: Suppose you have a sample of 20 students, and you want to test if their average score is significantly different from the population mean of 70. The sample mean is 72, and the sample standard deviation is 8. Test whether the sample mean is significantly different from the population mean at a 5% significance level.

Solution:

1.Formulate Hypotheses:

1. H_0: $\mu = 70$ (No significant difference)
2. H_1: $\mu \neq 70$ (Significant difference)

2. Choose Significance Level:

1. $\alpha = 0.05$

3. Collect and Analyze Data:

1. $\overline{x} = 72$
2. $s = 8$
3. $n = 20$

4. Calculate the Test Statistic:

$$t = \frac{72 - 70}{\frac{8}{\sqrt{20}}} \approx 1.58$$

5. Determine Degrees of Freedom:

1. ***df*** **= 20−1 = 19**

6. Find Critical Value or P-value:

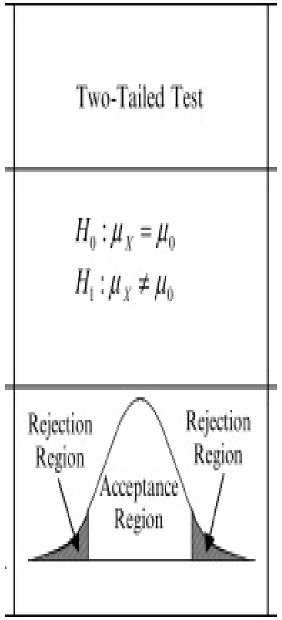
1. **At *α*/2=0.025 and *df*=19, *t (α*/2 , df) ≈ 2.093. (With the help of the t distribution table shown earlier)**

7. Make a Decision:

1. **Since |1.58| < 2.093 and the p-value is greater than 0.05, it fails to reject the null hypothesis.**

8. Interpret the Results:

1. **There is not enough evidence to suggest that the average score of the sample is significantly different from the population mean at the 5% significance level.**



Two sample t-test cases - Overview

Assumption: Is the variance for two populations equal?

Yes

$$t_{cal} = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$df = n_1 + n_2 - 2$$

No

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

$$df = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{(n_1 - 1)} + \frac{(s_2^2/n_2)^2}{(n_2 - 1)}}$$

Two sample t-test
Equal Variance Case

Hypothesis testing using a t-test for two samples is a statistical method used to determine if there is a significant difference between the means of two independent groups. It is commonly used in research and experimentation to compare means from different populations or treatments.

Here's a detailed explanation of the process:

1. Define the Hypotheses:

Null Hypothesis (H0): There is no significant difference between the means of the two groups.

$$Ho: \mu1 = \mu2$$

Alternative Hypothesis (H1 or Ha): There is a significant difference between the means of the two groups.

$$Ha: \mu1 \neq \mu2 \text{ (two-tailed test)}$$
$$Ha: \mu1 > \mu2 \text{ or } \mu1 < \mu2 \text{ (one-tailed test)}$$

Two sample t-test

Equal Variance Case

2. **Collect Data:**

Obtain data from two independent samples, each with its own set of observations.

3. **Verify Assumptions:**

Both samples are independent.
Both populations follow a normal distribution.
Homogeneity of variances (the variances of the two populations are equal).

4. **Calculate the Test Statistic:**

The test statistic formula for a two-sample t-test is calculated using the difference between the sample means divided by the standard error of the difference between the means. Here's the formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{pooled} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Two sample t-test
Equal Variance Case

Where:
- $\overline{x_1}$ and $\overline{x_2}$ are the sample means of the two groups.
- s_{pooled} is the pooled standard deviation, calculated as:

$$s_{pooled} = \sqrt{\frac{(n_1-1)\times s_1^2+(n_2-1)\times s_2^2}{n_1+n_2-2}}$$

- *s*1 and *s*2 are the sample standard deviations of the two groups.
- *n*1 and *n*2 are the sample sizes of the two groups.

5. **Determine the Degrees of Freedom:**

The test statistic *t* follows a t-distribution with degrees of freedom.
Degrees of freedom (df) for the t-test is calculated using the formula:

df = (n1 - 1) + (n2 - 1)

Example: Two sample t-test (1/2)

Problem: Suppose we want to compare the exam scores of two different classes (Class A and Class B) to see if there's a significant difference between their mean scores. Assuming equal variance for the population.

Class A:Sample size n_1 = 30 , Mean Score $\overline{x_1}$ = 75 , Standard deviation s_1 = 8
Class B :Sample size n_2 = 25 , Mean Score $\overline{x_2}$ = 72 , Standard deviation s_2 = 7

Solution:

Null Hypothesis (H0): There is no significant difference between the mean scores of Class A and Class B.

$$H0 : \mu_1 = \mu_2$$

Alternative Hypothesis (Ha): There is a significant difference between the mean scores of Class A and Class B.

$$Ha : \mu_1 \neq \mu_2$$

We will use a significance level of α=0.05.

Example: Two sample t-test (1/2)

Calculate the Test Statistic:

$$s_{pooled} = \sqrt{\frac{(n_1-1)\times s_1^2+(n_2-1)\times s_2^2}{n_1+n_2-2}}$$

$$s_{pooled} = \sqrt{\frac{(30-1)\times 8^2+(25-1)\times 7^2}{30+25-2}}$$

$$s_{pooled} \approx 7.40$$

$$t = \frac{\bar{x}_1-\bar{x}_2}{s_{pooled}\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$t = \frac{75-72}{7.40\times\sqrt{\frac{1}{30}+\frac{1}{25}}}$$

$$t \approx \frac{3}{1.99} \approx 1.51$$

t-test critical values

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										

Example: Two sample t-test (1/2)

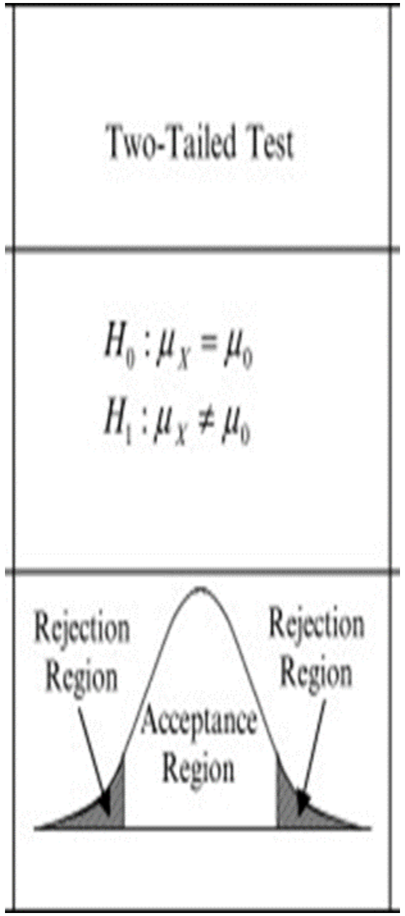
Degrees of Freedom: $df = 30+25-2=53$

Critical Values:

From the t-distribution table, with df=53 and α=0.05,

$t_{critical}$ is approximately ±2.004.

Since |1.51| < 2.004, we fail to reject the null hypothesis.



Example: Two sample t-test (2/2)

Problem: Suppose a school district is considering implementing a new teaching method (Method A) for teaching mathematics. To evaluate its effectiveness, they conduct a study comparing it to the traditional teaching method (Method B). Assuming the equal variance for the population. They randomly select two groups of students from the same grade level.

Group A: Students taught using Method A
Sample size n_1 = 35 , Mean Score $\overline{x_1}$ = 85 , Standard deviation s_1 = 10
Group A: Students taught using Method B
Sample size n_2 = 40 , Mean Score $\overline{x_2}$ = 80 , Standard deviation s_2 = 8

Solution:

Null Hypothesis (H0): There is no significant difference between the mean scores of students taught using Method A and Method B.

$$H0: \mu 1=\mu 2$$

Alternative Hypothesis (Ha): There is a significant difference between the mean scores of students taught using Method A and Method B.

$$Ha: \mu 1\neq\mu 2$$

Example: Two sample t-test (2/2)

We will use a significance level of α=0.05.

Calculate the Test Statistic:

$$s_{pooled} = \sqrt{\frac{(35-1)\times 10^2 + (40-1)\times 8^2}{35+40-2}}$$

$$s_{pooled} \approx 9.04$$

$$t = \frac{85-80}{9.04\times\sqrt{\frac{1}{35}+\frac{1}{40}}}$$

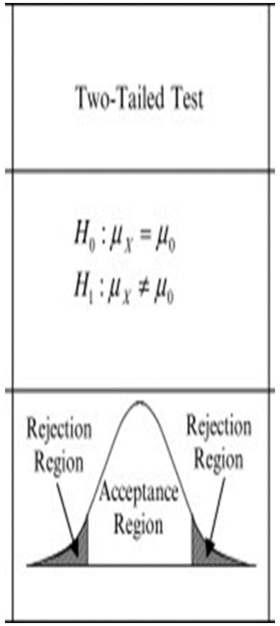
$$t \approx \frac{5}{2.303} \approx 2.17$$

Degrees of Freedom: ***df*** **=35+40−2=73**

Critical Values:

From the t-distribution table, with *df* = 73 and *α*=0.05, $t_{critical}$ is approximately ±1.994.

Since |2.17|>1.994, we reject the null hypothesis.



t-test critical values

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										

Two sample t-test

Unequal Variance Case

the hypothesis testing process for a two-sample t-test when the population variances are not assumed to be equal, also known as the Welch's t-test.

Here's a detailed explanation of the process:

1. Define the Hypotheses:

Null Hypothesis (H0): There is no significant difference between the means of the two groups.

$$Ho: \mu 1 = \mu 2$$

Alternative Hypothesis (H1 or Ha): There is a significant difference between the means of the two groups.

$$Ha: \mu 1 \neq \mu 2 \text{ (two-tailed test)}$$

$$Ha: \mu 1 > \mu 2 \text{ or } \mu 1 < \mu 2 \text{ (one-tailed test)}$$

Two sample t-test

Unequal Variance Case

2. **Collect Data:**
Obtain data from two independent samples, each with its own set of observations.

3. **Verify Assumptions:**
Both samples are independent.
Both populations follow a normal distribution.
The populations do not need to have equal variances.

4. **Calculate the Test Statistic:**
Use the Welch's t-test formula to calculate the test statistic:

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

Two sample t-test

Unequal Variance Case

Where:

- $\overline{x_1}$ and $\overline{x_2}$ are the sample means of the two groups.
- s_1^2 and s_2^2 are the sample standard deviations of the two groups.
- $n1$ and $n2$ are the sample sizes of the two groups.

5. Determine the Degrees of Freedom:

The test statistic *t* follows a t-distribution with degrees of freedom.
Degrees of freedom (df) for this t-test is calculated using the formula:

$$\mathrm{df} \approx \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

This equation is used to approximate the degrees of freedom for the t-distribution when the sample sizes and variances of the two groups being compared are different. It is used in situations where the assumption of equal variances (as assumed in the traditional Student's t-test) does not hold.

Two sample t-test

Unequal Variance Case

6. **Find Critical Values or P-value:**
Look up the critical value of t from the t-distribution table based on the significance level (α) and degrees of freedom.
Alternatively, calculate the p-value using software or statistical tools.

7. **Make a Decision:**
If the calculated t-value is greater than the critical value (or if the p-value is less than α), reject the null hypothesis.
If the calculated t-value is less than the critical value (or if the p-value is greater than α), fail to reject the null hypothesis.

Two sample t-test

Unequal Variance Case

8. **Interpretation:**

If the null hypothesis is rejected, it indicates that there is a significant difference between the means of the two groups.
If the null hypothesis is not rejected, it suggests that there is insufficient evidence to conclude a significant difference between the means of the two groups.

This procedure allows for hypothesis testing when the assumption of equal population variances is violated, making it applicable in a wider range of scenarios where the population variances may differ.

Example

Question 1:

Suppose you want to investigate whether there is a significant difference in the average scores between two teaching methods. You have two groups of students: Group A, taught using Method 1, and Group B, taught using Method 2. Assume Variance of Populations are not equal.

You collect the following data:

- **Group A (Method 1): Sample size (n_1) = 25, Sample mean ($\bar{x}_1$) = 78, Sample standard deviation (s_1) = 7**
- **Group B (Method 2): Sample size (n_2) = 30, Sample mean ($\bar{x}_2$) = 82, Sample standard deviation (s_2) = 9**

Test whether there is a significant difference in the average scores between the two teaching methods at a 5% significance level.

Solution:

1. Formulate Hypotheses:

1. H_0: $\mu_1 = \mu_2$ (No significant difference between teaching methods)
2. H_1: $\mu_1 \neq \mu_2$ (Significant difference between teaching methods)

2. Choose Significance Level:

1. $\alpha = 0.05$

3. Collect and analyze Data:

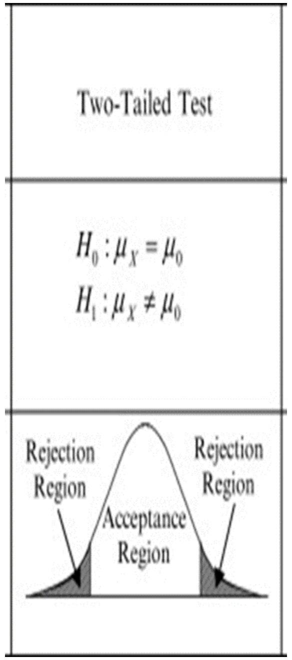
1. Group A: $\bar{x}_1 = 78$, $s_1 = 7$, $n_1 = 25$
2. Group B: $\bar{x}_2 = 82$, $s_2 = 9$, $n_2 = 30$

Example

4. Calculate the Test Statistic:

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

On Substituting values: **t value : -2.06**



5. Determine Degrees of Freedom:

Degrees of freedom (*df*) are calculated using the formula:

$$\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

***df* ≈ 51.69 (rounded down to the nearest whole number, *df* = 51)**

t-test critical values

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										

Example

6. Find Critical Value or P-value:
    1. At $\alpha/2$=0.025 and *df* = 51, $t_{\alpha/2,df}$ is approximately ±2.009 (using a t-table or statistical software).

7. Make a Decision:
    1. Since −2.06<−2.009, reject the null hypothesis.

8. Interpret the Results:
    1. There is enough evidence to suggest that there is a significant difference in the average scores between the two teaching methods at the 5% significance level.

REFERENCES

- Introduction to Probability and Statistics for Engineers and Scientists, Sixth Edition, Sheldon M. Ross
- Statistical Methods Combined Edition (Volume I& II), N G Das

CONCLUSION

- We discussed the t-test (for Population unknown variance)
    - Motivation Example about t-test
    - Assumptions of t-test

- We discussed one Sample t-test along with solved examples.

- We discussed about Two Sample t-test along with solved examples.
    (i) Assuming Populations with equal Variance
    (ii) Assuming Populations with unequal Variance


THANK YOU