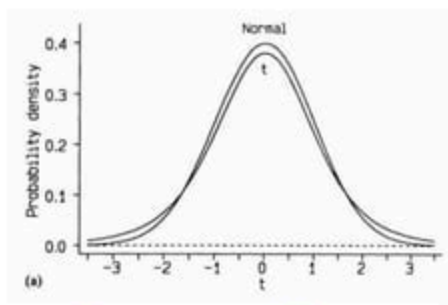


T Distribution, Paired and Unpaired T test

**Mean Comparison
between 2 groups**



Prepared by:

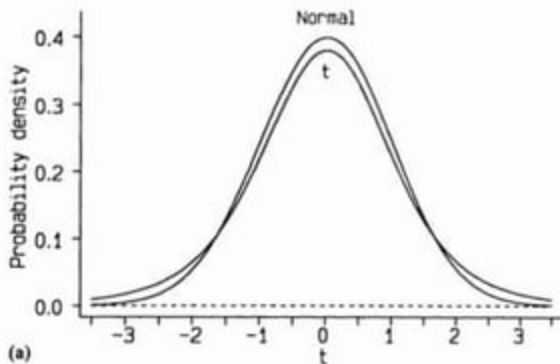
Bikram Adhikari (136)

A bit of History

- **W.A. Gassit** (1905) first published a t-test.
- He worked at the **Guinness Brewery in Dublin** and published under the name Student.
- The test was called Student Test → later shortened to t test

T Distribution

- Aka **Student's T distribution**
- A probability distribution that is used to estimate population parameters when the sample size is small and/or when the population variance is unknown.



(a)

T Distribution

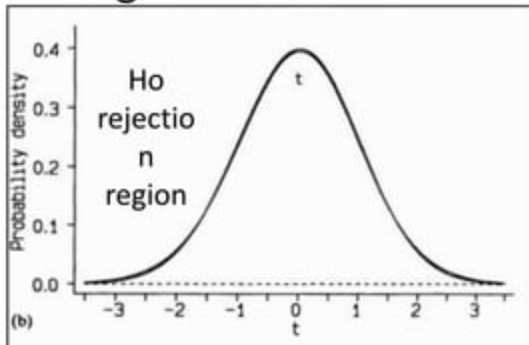
- If X is normally distributed and a sample of size n is randomly chosen from this underlying population, the probability distribution of the random variable is known as **Student's t distribution** with $n-1$ degrees of freedom.

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

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}$$

Properties of t Distribution

- The mean of distribution is ZERO.
- It is symmetrical about the mean.
- In general, it has a variance greater than 1, but the variance approaches 1 as the sample size becomes large.
- The variable t ranges from $-\infty$ to $+\infty$



Similarities with Normal Distribution

- Both are uni-modal
- Symmetric around its mean of zero
- Total area of the curve is One.

Difference between t and normal distribution

- Compared to the normal distribution, the t distribution is less peaked in the center and has thicker tails.

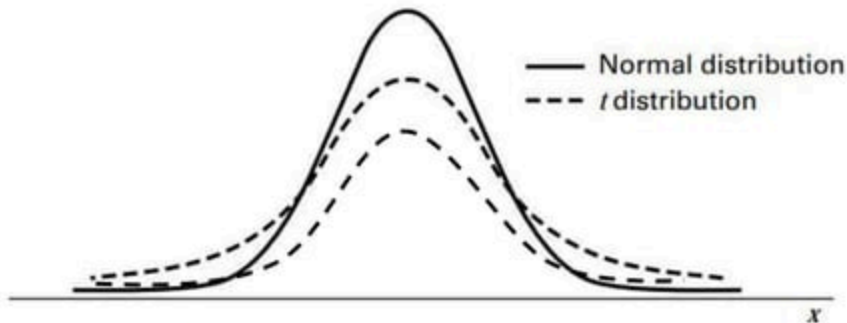


FIGURE 6.3.2 Comparison of normal distribution and t distribution.

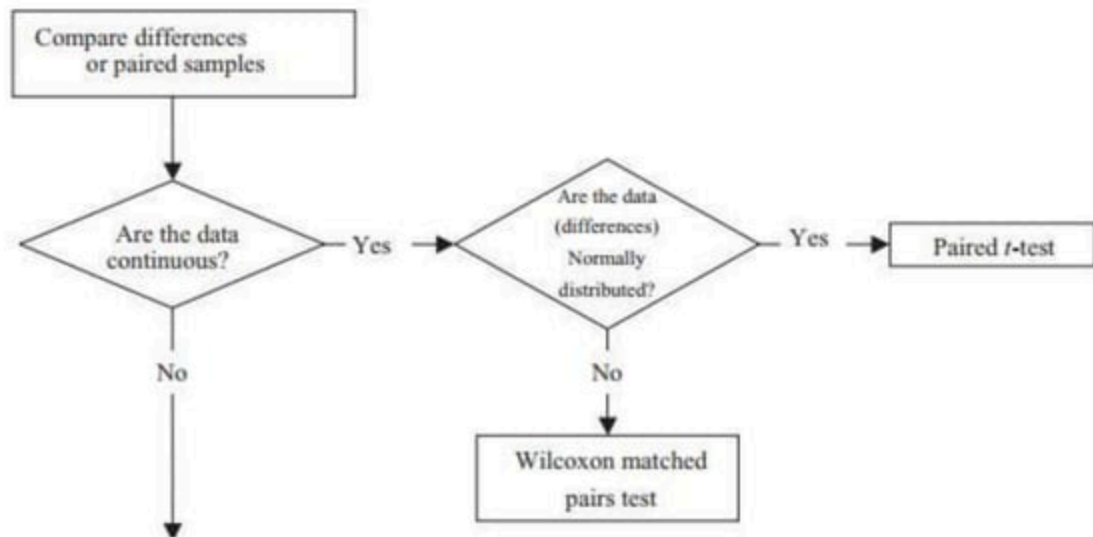
When to Use the t Distribution?

- The t distribution can be used with any statistic having a bell-shaped distribution
 - Quantative data, random samples
 - The population distribution is normal.
 - The population distribution is symmetric, unimodal, without outliers, and the sample size is at most 30.
 - The population distribution is moderately skewed, unimodal, without outliers, and the sample size is at most 40.

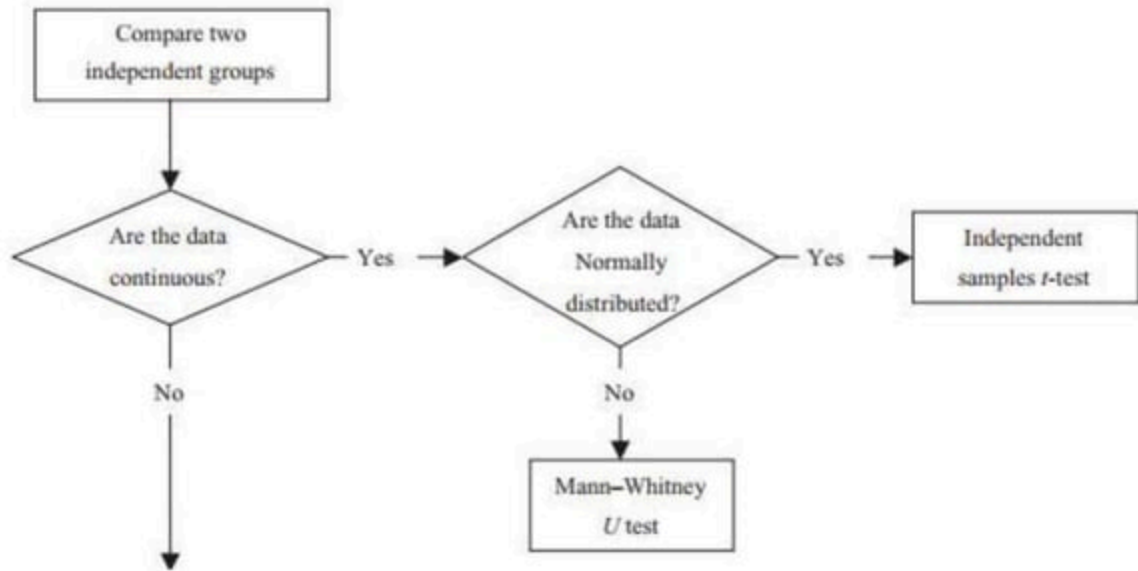
Application of t test

- T test can be applied if:
 - Samples are **randomly selected** from population
 - There is **homogeneity of variance** in sample
- It is applied to find the **significance of difference between two means** as:
 - Unpaired *t*-test
 - Paired *t*-test.

Statistical methods for differences or paired samples



Statistical methods for comparing two independent groups or samples



Standard Error of Mean

- Difference between sample estimates of statistics and population parameter is measured by standard error
- Standard error is a **measure of chance variation and it does not mean** → error or mistake
- Calculation of standard error of mean:

$$s\bar{X} = \sqrt{\frac{(\bar{X} - \mu)^2}{n - 1}}$$

$$SE \bar{X} \text{ or } (s\bar{X}) = \frac{SD}{\sqrt{n}} = \frac{s}{\sqrt{n}}$$

Standard Error of difference between Means

- Frequency distribution of difference give a normal curve.
- The standard deviation of a distribution of differences is known as **standard error of difference between two means**.
- Calculation of standard error of difference between two means:

$$s(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}} \text{ or } \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$SE = SD \times \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \text{ because } SD = \sqrt{SD^2}$$

Degree of Freedom (df)

- Degrees of freedom depends on two factors:
 - No. of groups we wish to compare
 - No. of parameters we need to estimate to calculate the standard deviation of the contrast of interest.
- For paired t test: **$df=n-1$**
- For one sample t test: **$df=n-1$**
- For two sample t test: **$df=2n-2$**

Paired t test

- Two groups of paired observations, $x_{11}, x_{12}, \dots, x_{1n}$ in Group 1 and $x_{21}, x_{22}, \dots, x_{2n}$ in Group 2 such that x_{1i} is paired with x_{2i} and the difference between them, $d_i = x_{1i} - x_{2i}$.
- *Assumptions*
 - The d_i 's are plausibly Normally distributed. It is not essential for the original observations to be Normally distributed
 - The d_i 's are independent of each other.

Paired t test

- *Steps*

- Hypothesis generation:
 - Null Hypothesis: mean difference is zero
 - Alternative Hypothesis: mean difference in the population is not zero.
- Calculate the differences $d_i = x_{1i} - x_{2i}$, $i = 1$ to n .
- Calculate the mean difference (d) and standard deviation, (Sd) of the differences d_i .
- Calculate the standard error of the mean difference

$$SE(\bar{d}) = s_d / \sqrt{n}$$

- Calculate the test statistic

$$t = \frac{\bar{d}}{SE(\bar{d})}$$

Paired t test

- *Steps*

- Find the degrees of freedom.
 - $df=n-1$
- Refer 't' table and find the probability of the calculated 't' corresponding to $n - 1$ degrees of freedom.
- Result

Paired t test

Problem

- Systolic blood pressure (SBP) of 9 normal individuals, who had been recumbent for 5 minutes was taken. Then 2 ml of 0.5% solution of hypotensive drug was given and blood pressure recorded again. **Did the injection of drug lower the blood pressure?**

Paired t test

- Hypothesis:
 - Ho: no mean difference
 - Ha: mean difference
- Calculate differences
- Calculation of mean difference:

$$\bar{X} = \frac{\Sigma \bar{X}}{n} = 27 \div 9 = 3$$

- Calculate Sd of mean difference

$$SD = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}} = \sqrt{\frac{105 - \frac{27 \times 27}{9}}{9-1}} = \sqrt{\frac{24}{8}} = \sqrt{3} = 1.73$$

Serial No.	BP before injection	BP after injection	Difference	Squares
	X_1	X_2	$X_1 - X_2 = x$	x^2
1	122	120	2	4
2	121	118	3	9
3	120	115	5	25
4	115	110	5	25
5	126	122	4	16
6	130	130	0	0
7	120	116	4	16
8	125	124	1	1
9	128	125	3	9
Total			27	105

Paired t test

- Calculate standard error of mean difference:

$$SE \text{ difference} = \frac{SD}{\sqrt{n}} = \frac{1.73}{\sqrt{9}} = 0.58$$

- Calculate test statistic

$$t_8 = \frac{\bar{X}}{SE} = \frac{3}{0.58} = 5.17$$

- Degree of freedom: $n-1=9-1=8$

Paired t test

- Find the probability of the calculated ' t ' corresponding to 8 degrees of freedom and 5% significance

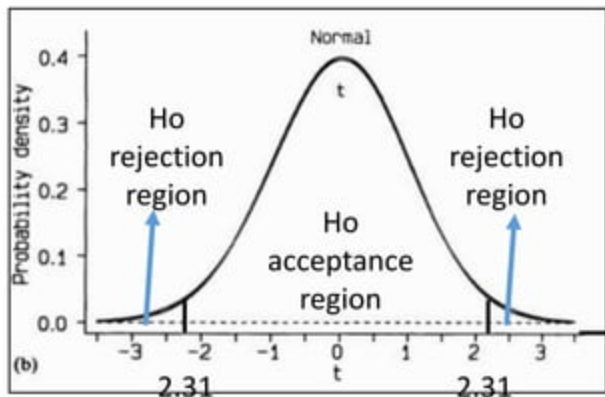
Table of ' t '
Probability of Larger Value of ' t '

- $T=2.31$

<i>DF</i>	0.10	0.05	0.02	0.01	0.001
1	6.31	12.71	3.82	63.66	636.62
2	2.92	4.30	6.97	9.93	31.60
3	2.35	3.18	4.54	5.84	12.92
4	2.13	2.78	3.75	4.60	8.61
5	2.02	2.57	3.37	4.03	6.87
6	1.94	2.45	3.14	3.71	5.96
7	1.90	2.37	3.00	3.50	5.41
8	1.86	2.31	2.90	3.36	5.04
9	1.83	2.26	2.82	3.25	4.78
10	1.81	2.23	2.76	3.17	4.59
11	1.80	2.20	2.72	3.11	4.44
12	1.78	2.18	2.68	3.06	4.32
13	1.77	2.16	2.65	3.01	4.27

Paired t test

- Test statistic lies in rejection region.
- H_0 is rejected
- It means there is significant difference between means at 5% cl.



Unpaired t test

- One Sample t test
- Independent two Sample t test

Independent sample t test

- The independent samples t -test is used to test for a difference in the mean value of a continuous variable between two groups.
- **Assumptions**
 - The groups are independent.
 - The variables of interest are continuous
 - The data in both groups have similar standard deviations
 - **Homogeneity of variance**
 - The data is Normally distributed in both groups.

Homogeneity of variance test

- **Levene's test**

- This tests whether the variances of two samples are approximately equal or not.
- We want Levene's test to be not significant
 - H_0 : Variances are equal
 - H_a : Variances are not equal
- **Note:** As long as $N > 30$, n_1 nearly = to n_2 , there is robust to violations of homogeneity of variance

Independent sample t test

- **Steps**

- Generate hypothesis
 - Ho: no difference in two means
 - Ha: there is difference in two means
- Calculate mean difference between groups
- Calculate pooled standard deviations

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

- Calculate the standard error of the difference between two means

$$SE(\bar{x}_1 - \bar{x}_2) = SD_{\text{pooled}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Independent sample t test

- Calculate the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}$$

- Calculate degree of freedom:

$$df = n_1 + n_2 - 2$$

- Refer 't' table and find the probability of the calculated 't' corresponding to $2n - 2$ degrees of freedom.
- Result interpretation

Independent sample t test

A random sample is selected from each population. The sample of $n_1 = 9$ healthy children has mean serum iron level $\bar{x}_1 = 18.9 \mu\text{mol/l}$ and standard deviation $s_1 = 5.9 \mu\text{mol/l}$; the sample of $n_2 = 13$ children with cystic fibrosis has mean iron level $\bar{x}_2 = 11.9 \mu\text{mol/l}$ and standard deviation $s_2 = 6.3 \mu\text{mol/l}$ [5]. Is it likely that the observed difference in sample means—18.9 versus 11.9 $\mu\text{mol/l}$ —is the result of chance variation, or should we conclude that the discrepancy is due to a true difference in population means?

$$\begin{aligned}s_p^2 &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \\ &= \frac{(9 - 1)(5.9)^2 + (13 - 1)(6.3)^2}{9 + 13 - 2} \\ &= \frac{(8)(34.81) + (12)(39.69)}{20} \\ &= 37.74.\end{aligned}$$

$$\begin{aligned}t &= \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 [(1/n_1) + (1/n_2)]}} \\ &= \frac{(18.9 - 11.9) - 0}{\sqrt{(37.74)[(1/9) + (1/13)]}} \\ &= 2.63.\end{aligned}$$

Independent sample t test

- Degree of freedom = $9+13-2 = 20$
- $T_{20}=2.086$ (from table)
- Calculated t statistic = 2.63

What if the variances in two groups are not Equal?

- Use of modification of Independent sample t test.
- Instead of using sp^2 as an estimate of the common variance σ^2 , we substitute s_1^2 for σ^2 and s_2^2 for σ_2^2 . Therefore, the appropriate test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}}$$

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



Degree of freedom

Suppose that we are interested in investigating the effects of an antihypertensive drug treatment on persons over the age of 60 who suffer from isolated systolic hypertension. By definition, individuals with this condition have a systolic blood pressure greater than 160 mm Hg while their diastolic blood pressure is below 90 mm Hg. Before the beginning of the study, subjects who had been randomly selected to take the active drug and those chosen to receive a placebo were comparable with respect to systolic blood pressure. After one year of participating in the study, the mean systolic blood pressure for patients receiving the drug is denoted by μ_1 and the mean for those receiving the placebo by μ_2 . The standard deviations of the two populations are unknown, and we do not feel justified in assuming that they are equal. Since we would like to determine whether the mean systolic blood pressures of the patients in these two different groups remain the same,

We begin by selecting a random sample from each of the two groups. The sample of $n_1 = 2308$ individuals receiving the active drug treatment has mean systolic blood pressure $\bar{x}_1 = 142.5$ mm Hg and standard deviation $s_1 = 15.7$ mm Hg; the sample of $n_2 = 2293$ persons receiving the placebo has mean $\bar{x}_2 = 156.5$ mm Hg and standard deviation $s_2 = 17.3$ mm Hg [7]. We are interested in detecting differences that could occur in either direction, and thus conduct a two-sided test at the $\alpha = 0.05$ level of significance. The alternative hypothesis is

$$\begin{aligned}t &= \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{(s_1^2/n_1) + (s_2^2/n_2)}} \\&= \frac{(142.5 - 156.5) - 0}{\sqrt{[(15.7)^2/2308] + [(17.3)^2/2293]}} \\&= -28.74.\end{aligned}$$

$$\begin{aligned}v &= \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{[(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)]} \\&= \frac{[(246.49/2308) + (299.29/2293)]^2}{[(246.49/2308)^2/(2308 - 1) + (299.29/2293)^2/(2293 - 1)]} \\&= 4550.5.\end{aligned}$$

One Sample t test

- It is used for test of the null hypothesis that our data are a sample from a population with a specific 'hypothesized' mean.
- Test statistic(t):

$$t = \frac{\text{sample mean} - \text{hypothesized mean}}{\text{standard error of sample mean}}$$

The mean and standard deviation of the dietary intake data were 6753.6 and 1142.1 kJ, and the hypothetical value of interest was the recommended intake of 7725 kJ. We can thus calculate the value of t as

Ho: there is no difference in dietary intake from recommended level.

Ha: there is difference in dietary intake from recommended level.

The mean and standard deviation of the dietary intake data were 6753.6 and 1142.1 kJ, and the hypothetical value of interest was the recommended intake of 7725 kJ. We can thus calculate the value of t as

$$t = \frac{6753.6 - 7725}{1142.1/\sqrt{11}}$$
$$= - 2.821.$$

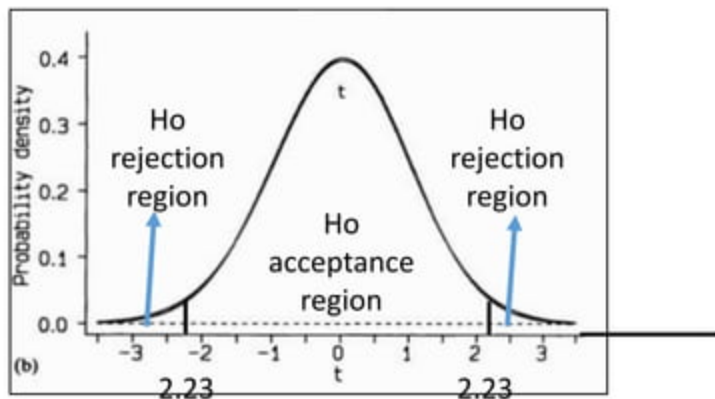
Table of ' t '
Probability of Larger Value of ' t '

<i>DF</i>	0.10	0.05	0.02	0.01	0.001
1	6.31	12.71	3.82	63.66	636.62
2	2.92	4.30	6.97	9.93	31.60
3	2.35	3.18	4.54	5.84	12.92
4	2.13	2.78	3.75	4.60	8.61
5	2.02	2.57	3.37	4.03	6.87
6	1.94	2.45	3.14	3.71	5.96
7	1.90	2.37	3.00	3.50	5.41
8	1.86	2.31	2.90	3.36	5.04
9	1.83	2.26	2.82	3.25	4.78
10	1.81	2.23	2.76	3.17	4.59
11	1.80	2.20	2.72	3.11	4.44
12	1.78	2.18	2.68	3.06	4.32
13	1.77	2.16	2.65	3.01	4.22
14	1.76	2.15	2.62	2.98	4.14

- 95% CI
- $df=10$
- $t=2.23$

One Sample t test

- Result interpretation
 - The dietary intake of female is significantly lower than recommended level



References

- Altman, A practical statistics for medical research, 8th edition, 1999
- Mahajan's Methods in Biostatistics for medical students and research works. 8th edition, 2016
- Medical statistics-Text book for health sciences, 4th edition, 2007
- Principles of Biostatistics, Marcello Pagano, 2nd Edition
- Few internet sites.

Any Queries????

