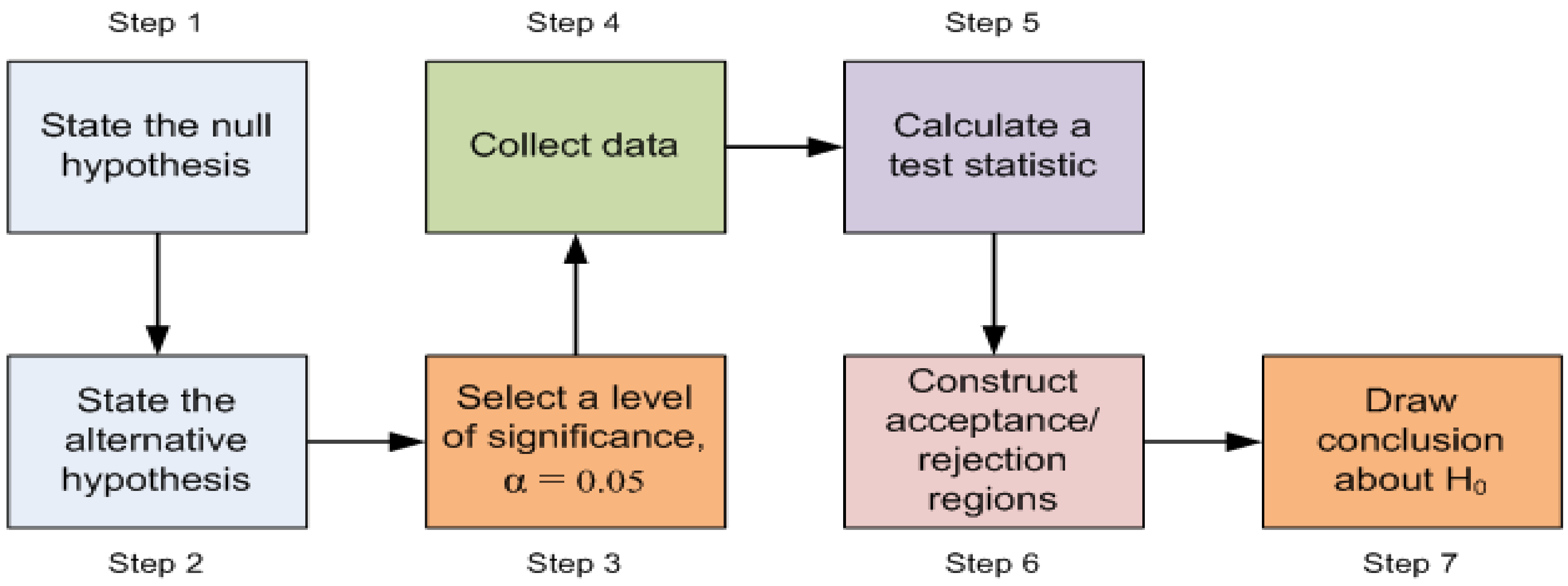
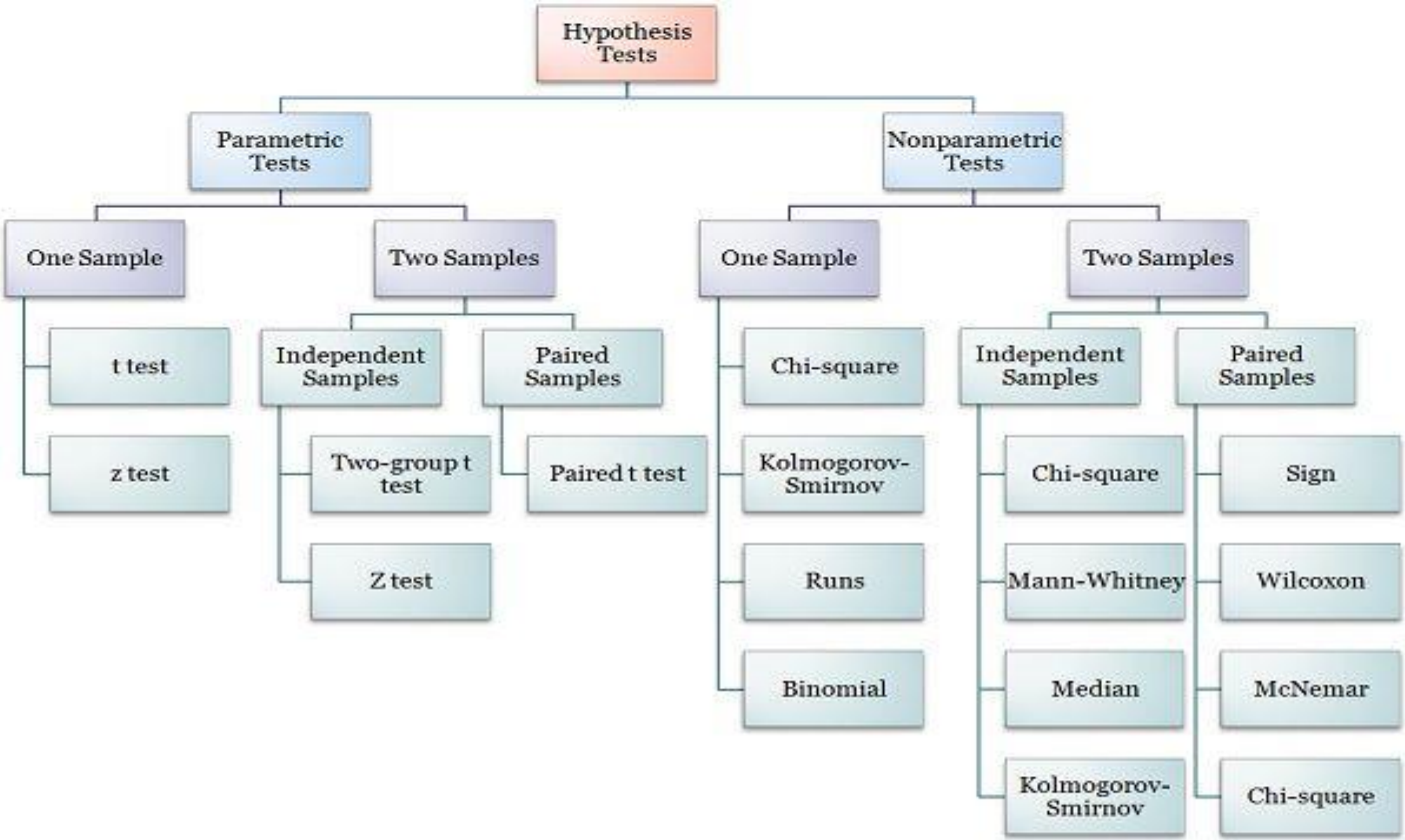
Hypothesis testing

Seven-Step Procedure for Testing a Hypothesis




Hypothesis Tests
Parametric Tests
Nonparametric Tests
One Sample
Two Samples
One Sample
Two Samples
t test
z test
Independent Samples
Paired Samples
Two-group t test
Z test
Paired t test
Chi-square
Kolmogorov-Smirnov
Runs
Binomial
Independent Samples
Paired Samples
Chi-square
Mann-Whitney
Median
Kolmogorov-Smirnov
Sign
Wilcoxon
McNemar
Chi-square

Difference between Parametric and Non-parametric tests

Non-parametric	Parametric
• Can be used on ordinal and nominal scale data (although also on interval and ratio scale).	• Used mainly on interval and ratio scale data.
• Can be used on small samples.	• Tend to need larger samples.
• Can be used on data that are not normally distributed.	• Data should fit a particular distribution; the data can be transformed to that distribution.
• Can be used where the samples are not selected randomly.	• Samples should be drawn randomly from the population.
• Have less power than the equivalent parametric test.	• More powerful than non-parametric equivalent.

Z test is a statistical test that is conducted on data that approximately follows a normal distribution. The z-test can be performed on one sample, two samples, or on proportions for hypothesis testing. It checks if the means of two large samples are different or not when the population variance is known.

Z-test can further be classified into left-tailed, right-tailed, and two-tailed hypothesis tests depending upon the parameters of the data. (for two-tailed test level of significance is $\alpha/2$ and for one tailed test it is $(0.5 - \alpha)$.

Rejection Region	*Level of Significance, α per cent*			
	$\alpha = 0.10$	$\alpha = 0.05$	0.01	$\alpha = 0.005$
One-tailed region	± 1.28	± 1.645	± 2.33	± 2.58
Two-tailed region	± 1.645	± 1.96	± 2.58	± 2.81

An ambulance service claims that it takes, on average 8.9 minutes to reach its destination in emergency calls. To check on this claim, the agency which licenses ambulance services has then timed 50 emergency calls, getting a mean of 9.3 minutes with a standard deviation of 1.8 minutes. Does this constitute evidence that the figure claimed is too low at the 1 percent significance level?

Solution: Let us consider the null hypothesis H_0 that 'the claim is same as observed' and alternative hypothesis is 'claim is different than observed'. These two hypotheses are written as:

$$H_0 : \mu = 8.9 \text{ and } H_1 : \mu \neq 8.9$$

Given $n = 50$, $\bar{x} = 9.3$, and $s = 1.8$. Using the z-test statistic, we get

$$z = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{9.3 - 8.9}{1.8/\sqrt{50}} = \frac{0.4}{0.254} = 1.574$$

Since $z_{cal} = 1.574$ in less than its critical value $z_{\alpha/2} = \pm 2.58$, at $\alpha = 0.01$, the null hypothesis is accepted. Thus, there is no difference between the average time observed and claimed.

A continuous manufacturing process of steel rods is said to be in a ‘state of control’ and produces acceptable rods if the mean diameter of all rods procuced is 2 inches. Although the process standard deviation exhibits stability over time with standard deviation, $\sigma = 0.01$ inch. The process means may vary due to operator error or problems of process adjustment. Periodically, random samples of 100 rods are selected to determine whether the process is producing acceptable rods. If the result of a test indicates that the process is out of control, it is stopped and the source of trouble is sought. Otherwise, it is allowed to continue operating. A random sample of 100 rods is selected resulting in a mean of 2.1 inches. Test the hypothesis to determine whether the process be continued.

Solution: Since rods that are either too narrow or too wide are unacceptable, the low values and high values of the sample mean lead to the rejection of the null hypothesis. Consider the null hypothesis H_0, that the process may be allowed to continue when diameter is 2 inches. Consequencly, rejection region is on both tails of the sampling distribution. The null and alternative hypotheses are stated as follows:

$$H_0 : \mu = 2 \text{ inches, (continue process)}$$

$$H_1 : \mu \neq 2 \text{ inches, (stop the process)}$$

Given $n = 100$, $\bar{x} = 2.1$, $\sigma = 0.01$, $\alpha = 0.01$. Using the z-test statistic

$$z = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{2.1-2}{0.01/\sqrt{100}} = \frac{0.1}{0.001} = 100$$

Since $z_{cal}=100$ value is more than its critical value $z_{\alpha/2} = 2.58$ at $\alpha = 0.01$, the null hypothesis, H_0 is rejected. Thus stop the process in order to determine the source of trouble.

The mean lifetime of a sample of 400 fluorescent light bulbs produced by a company is found to be 1600 hours with a standard deviation of 150 hours. Test the hypothesis that the mean lifetime of the bulbs produced in general is higher than the mean life of 1570 hours at $\alpha = 0.01$ level of significance.

Solution: Let us take the null hypothesis that mean life time of bulbs is not more than 1570 hours, that is

$$H_0 : \mu \leq 1570 \text{ and } H_1 : \mu > 1570 \quad \text{(Right-tailed test)}$$

Given $n = 400$, $\bar{x} = 1600$ hours, $s = 150$ hrs and $\alpha = 0.01$. Thus using the z-test statistic.

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1600 - 1570}{150/\sqrt{400}} = \frac{30}{7.5} = 4$$

Since the calculated value $z_{cal} = 4$ is more than its critical value $z_{\alpha} = \pm 2.33$, the H_0 is rejected. Hence, we conclude that the mean lifetime of bulbs produced by the company may be higher than 1570 hours.

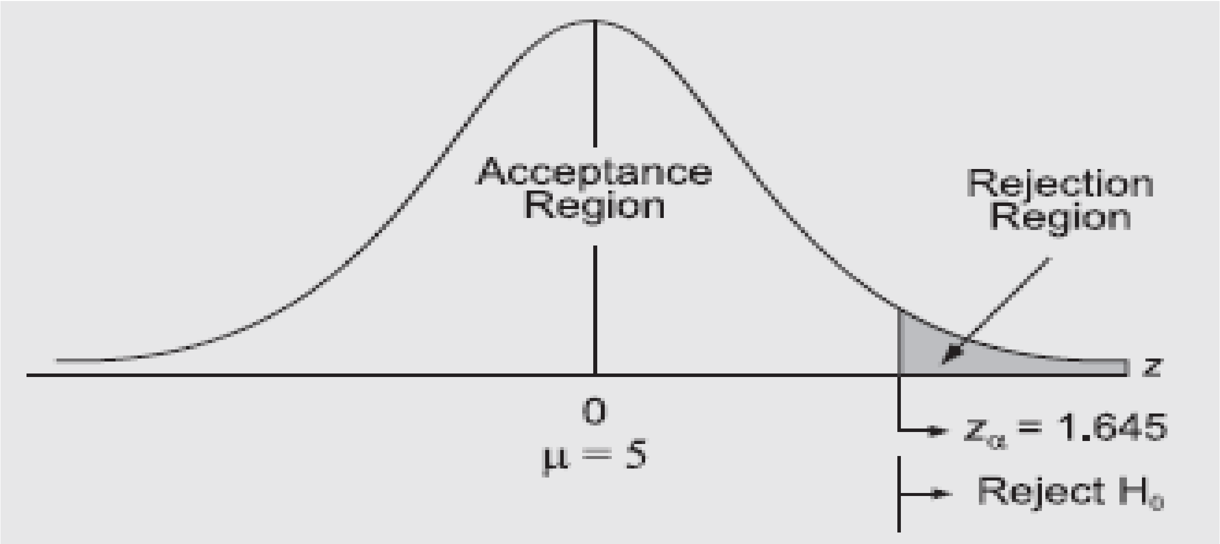
A packaging device is set to fill detergent powder packets with a mean weight of 5 kg, with a standard deviation of 0.21 kg. The weight of packets can be assumed to be normally distributed. The weight of packets is known to drift upwards over a period of time due to machine fault, which is not tolerable. A random sample of 100 packets is taken and weighed. This sample has a mean weight of 5.03 kg. Can we conclude that the mean weight produced by the machine has increased? Use a 5 percent level of significance.

Solution: Let us take the null hypothesis H_0 that mean weight has increased, that is,

$$H_0 : \mu \geq 5 \text{ and } \quad H_1 : \mu < 5$$

Given $n = 100$, $\bar{x} = 5.03$ kg, $\sigma = 0.21$ kg and $\alpha = 5$ per cent. Thus using the z-test statistic

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{5.03 - 5}{0.21\sqrt{100}} = \frac{0.03}{0.021} = 1.428$$



Since calculated value $z_{cal} = 1.428$ is less than its critical value $z_\alpha = 1.645$ at $\alpha = 0.05$, the null hypothes, H_0 is accepted as shown in Fig. 10.4. Hence we conclude that mean weight is likely to be more than 5 kg.

Hypothesis Testing for Difference between Two Population Means

Let two independent random samples of large size $n1$ and $n2$ be drawn from the first and second populations, respectively. Let the sample means so calculated be $x1$ and $x2$.

$$\text{Test statistic: } z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

where $\sigma_{\bar{x}_1 - \bar{x}_2}$ = standard error of the statistic $(\bar{x}_1 - \bar{x}_2)$

$\bar{x}_1 - \bar{x}_2$ = difference between two sample means, that is, sample statistic

$\mu_1 - \mu_2$ = difference between population means, that is, hypothesized population parameter

If $\sigma_1^2 = \sigma_2^2$, the above formula algebrically reduces to:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

The null and alternative hypothesis are stated as:

Null hypothesis : $H_0 : \mu_1 - \mu_2 = d_0$

Alternative hypothesis :

One-tailed Test	*Two-tailed Test*
$H_1 : (\mu_1 - \mu_2) > d_0$	$H_1 : (\mu_1 - \mu_2) \neq d_0$
$H_1 : (\mu_1 - \mu_2) < d_0$	

where d_0 is some specified difference that is desired to be tested. If there is no difference between μ_1 and μ_2, i.e. $\mu_1 = \mu_2$, then $d_0 = 0$.

A firm believes that the tires produced by process A on average last longer than tires produced by process B. To test this belief, random samples of tires produced by the two processes were tested and the results are:

Process	Sample Size	Average Lifetime (in km)	Standard Deviation (in km)
A	50	22,400	1000
B	50	21,800	1000

Is there evidence at a 5 per cent level of significance that the firm is correct in its belief ?

Solution: Let us take the null hypothesis that there is no significant difference in the average life of tyres produced by processes A and B, that is,

$$H_0 : \mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0 \quad \text{and} \quad H_1 : \mu_1 \neq \mu_2$$

Given, $\bar{x}_1$ =22,400 km, $\bar{x}_2$ = 21,800 km, $\sigma_1 = \sigma_2$ = 1000 km, and $n_1 = n_2$ = 50. Thus using the z-test statistic

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

$$= \frac{22,400 - 21,800}{\sqrt{\dfrac{(1000)^2}{50} + \dfrac{(1000)^2}{50}}} = \frac{600}{\sqrt{20,000 + 20,000}} = \frac{600}{200} = 3$$

Since the calculated value z_{cal} =3 is more than its critical value $z_{\alpha/2} = \pm 1.645$ at α = 0.05 level of significance, therefore H_0 is rejected. Hence we can conclude that the tyres produced by process A last longer than those produced by process B.

Q. An experiment was conducted to compare the mean time in days required to recover from a common cold for person given daily dose of 4 mg of vitamin C versus those who were not given a vitamin supplement. Suppose that 35 adults were randomly selected for each treatment category and that the mean recovery times and standard deviations for the two groups were as follows:

	Vitamin C	*No Vitamin Supplement*
Sample size	35	35
Sample mean	5.8	6.9
Sample standard deviation	1.2	2.9

Test the hypothesis that the use of vitamin C reduces the mean time required to recover from a common cold and its complications, at the level of significance = 0.05.

Solution: Let us take the null hypothesis that the use of vitamin C reduces the mean time required to recover from the common cold, that is

$$H_0 : (\mu_1 - \mu_2) \leq 0 \text{ and } H_1 : (\mu_1 - \mu_2) > 0$$

Given $n_1 = 35$, $\bar{x}_1 = 5.8$, $s_1 = 1.2$ and $n_2 = 35$, $\bar{x}_2 = 6.9$, $s_2 = 2.9$. The level of significance, $\alpha = 0.05$. Substituting these values into the formula for z-test statistic, we get

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

$$= \frac{5.8 - 6.9}{\sqrt{\frac{(1.2)^2}{35} + \frac{(2.9)^2}{35}}} = \frac{-1.1}{\sqrt{0.041 + 0.240}} = -\frac{1.1}{0.530} = -2.605$$

Using a one-tailed test with significance level $\alpha = 0.05$, the critical value is $z_\alpha = 1.645$. Since $z_{cal} < z_\alpha$ $(= 1.645)$, the null hypothesis H_0 is rejected. Hence we can conclude that the use of vitamic C does not reduce the mean time required to recover from the common cold.

t-test

When the sample size is small (i.e., less than 30), the central limit theorem does not assure us to assume that the sampling distribution of a statistic such as mean $\bar{x}$, proportion $\bar{p}$, is normal. Consequently when testing a hypothesis with small samples, we must assume that the samples come from a normally or approximately normally distributed population. Under these conditions, the sampling distribution of sample statistic such as $\bar{x}$ and $\bar{p}$ is normal but the critical values of $\bar{x}$ or $\bar{p}$ depend on whether or not the population standard deviation σ is known. When the value of the population standard deviation σ is not known, its value is estimated by computing the standard deviation of samples s and the standard error of the mean is calculated by using the formula, $\sigma_{\bar{x}} = s/\sqrt{n}$. When we do this, the resulting sampling distribution may not be normal even if sampling is done from a normally distributed population. In all such cases the sampling distribution turns out to be the *Student's t-distribution*.

Sir William Gosset of Ireland in early 1900, under his pen name 'Student', developed a method for hypothesis testing popularly known as the '**t-test**'. It is said that Gosset was employed by Guinness Breway in Dublin, Ireland which did not permit him to publish his research findings under his own name, so he published his research findings in 1905 under the pen name 'Student'.

Uses of t- Distribution

There are various uses of t-distribution. A few of them are as follows:

(i) Hypothesis testing for the population mean.

(ii) Hypothesis testing for the difference between two populations means with independent samples.

(iii) Hypothesis testing for the difference between two populations means with dependent samples.

(iv) Hypothesis testing for an observed coefficient of correlation including partial and rank correlations.

(v) Hypothesis testing for an observed regression coefficient.

1.Hypothesis Testing for Single Population Mean.

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{\bar{x} - \mu}{s / \sqrt{n}}; \quad s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

2. Hypothesis Testing for Difference of Two Population Means (Independent 2 Samples)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$s^2 = \frac{\Sigma (x_1 - \bar{x}_1)^2 + \Sigma (x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

3. Hypothesis Testing for Difference of Two Population Means (Dependent Samples)

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

where

n = number of paired observations

df = $n - 1$, degrees of freedom

$\bar{d}$ = mean of the difference between paired (or related) observeations

n = number of pairs of differences

s_d = sample standard deviation of the distribution of the difference between the paired (or related observations)

$$= \sqrt{\frac{\Sigma (d - \bar{d})^2}{n-1}} = \sqrt{\frac{\Sigma d^2}{n-1} - \frac{(\Sigma d)^2}{n(n-1)}}$$

The average breaking strength of steel rods is specified to be 18.5 thousand kg. For this a sample of 14 rods was tested. The mean and standard deviation obtained were 17.85 and 1.955, respectively. Test the significance of the deviation.

Solution: Let us take the null hypothesis that there is no significant deviation in the breaking strength of the rods, that is,

$$H_0 : \mu = 18.5 \quad \text{and} \quad H_1 : \mu \neq 18.5 \quad \text{(Two-tailed test)}$$

Given, $n = 14$, $\bar{x} = 17.85$, $s = 1.955$, $df = n - 1 = 13$, and $\alpha = 0.05$ level of significance. The critical value of t at $df = 13$ and $\alpha/2 = 0.025$ is $t_{\alpha/2} = 2.16$.

Using the z-test statistic,

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{17.85 - 18.5}{\frac{1.955}{\sqrt{14}}} = -\frac{0.65}{0.522} = -1.24$$

Since t_{cal} $(= -1.24)$ value is more than its critical value $t_{\alpha/2} = -2.16$ at $\alpha/2 = 0.025$ and $df = 13$, the null hypothesis H_0 is accepted. Hence we conclude that there is no significant deviation of sample mean from the population mean.

An automobile tyre manufacturer claims that the average life of a particular grade of tyre is more than 20,000 km when used under normal conditions. A random sample of 16 tyres was tested and a mean and standard deviation of 22,000 km and 5000 km, respectively were computed. Assuming the life of the tyres in km to be approximately normally distributed, decide whether the manufacturer's claim is valid.

Solution: Let us take the null hypothesis that the manufacturer's claim is valid, that is,

$$H_0 : \mu \geq 20{,}000 \quad \text{and} \quad H_1 : \mu < 20{,}000 \text{ (Left-tailed test)}$$

Given, $n = 16$, $\bar{x} = 22{,}000$, $s = 5000$, $df = 15$ and $\alpha = 0.05$ level of significance. The critical value of t at $df = 15$ and $\alpha = 0.05$ is $t_\alpha = 1.753$. Using the z-test statistic,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{22{,}000 - 20{,}000}{5000/\sqrt{16}} = \frac{2000}{1250} = 1.60$$