LARGE SAMPLE

If the size of the sample n>30, then that samplw is said to be large sample. There are four important test to test the significance of large samples.

(i). Test of significance for single mean.
(ii). Test of significance for difference of two means.
(iii). Test of significance for single proportion
(iv). Test of significance for difference of two proportions.

Note:
(i). The sampling distribution of a static is approximately normal, irrespective of whether the distribution of the population is normal or not.
(ii). The sample statistics are sufficiently close to the corresponding population parameters and hence may be used to calculate the standard errors of the sampling distribution.
(iii). Critical values for some standard LOS's (For Large Samples)

Nature of test	**1% (0.01) (99%)**	**2% (0.02) (98%)**	**5% (0.05) (95%)**	**10% (0.1) (90%)**
Two Tailed Test	$\lvert z_\alpha \rvert = 2.58$	$\lvert z_\alpha \rvert = 2.33$	$\lvert z_\alpha \rvert = 1.96$	$\lvert z_\alpha \rvert = 1.645$
One Tailed Test (Right tailed Test)	$z_\alpha = 2.33$	$z_\alpha = 2.055$	$z_\alpha = 1.645$	$z_\alpha = 1.28$
One Tailed Test (Left tailed Test)	$z_\alpha = -2.33$	$z_\alpha = -2.055$	$z_\alpha = -1.645$	$z_\alpha = -1.28$

Problem based on Test of significance for single mean:

The test statistic $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ where $\bar{x}$ =sample mean, μ=population mean, σ = standard deviation of population, n= sample size.

Note:

If standard deviation of population is not known then the static is $z = \dfrac{\bar{x} - \mu}{S / \sqrt{n}}$,

where S = standard deviation of sample.

Confident Interval:
The confident interval for μ when σ is known and sampling is done from a normal population or with a large sample is $\bar{x} \mp z_\alpha \dfrac{\sigma}{\sqrt{n}}$

$$\Rightarrow \left(\bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}} \right)$$

If s is known (σ is not known): $\bar{x} \mp z_{\alpha} \frac{s}{\sqrt{n}}$

1. A sample of 100 students is taken from a large population, the mean height in the sample is 160cm. Can it be reasonable regarded that in the population the mean height is 165cm, and s.d. is 10cm. and find confident limit. Use an level of significance at 1%

Solution:

Given n = 100, $\bar{x}$ =160cm, μ=165cm, σ =10cm

Let $H_0 : \mu = 165$

$H_1 : \mu \neq 165$ (two tailed test)

Under H_0, the test statistic is $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{160 - 165}{10 / \sqrt{100}} = -5$

$\therefore |z| = -5$

From the table, $z_{0.01}$ =2.58. Since $|z| > z_{0.01}$ $\therefore$ H_0 is rejected. hence $\mu \neq 165$.

Confident Interval:

$$\left(\bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}} \right) = \left(160 - 2.58 \frac{10}{\sqrt{100}}, 160 + 2.58 \frac{10}{\sqrt{100}} \right) = (157.42, 162.58)$$

2. The mean breaking strength of the cables supplied by a manufacture is 1800 with a S.D of 100. By a new techniques in the manufacturing process, it it claimed that the breaking strength of the cable has increased. In order to test this claim, a sample of 50 cables is tested and it is found that the mean breaking strength is 1850. Can we support the claim at 1% level of significance?

Solu:

Given n = 50, $\bar{x}$ =1850, μ=1800, σ =100

Let $H_0 : \bar{x} = \mu$

$H_1 : \bar{x} > \mu$ (one tailed test)

Under H_0, the test statistic is $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{1850 - 1800}{100 / \sqrt{50}} = 3.535$

$\therefore |z| = 3.535$

From the table, $z_{0.01}$ =2.33. Since $|z| > z_{0.01}$ $\therefore$ H_0 is rejected. hence $\bar{x} > \mu$.

3. A sample of 900 members has a mean of 3.4 cms and s.d is 2.61 cms. Is the sample from a large population of mean 3.25cm and s.d is 2.61 cms. If the population is normal and its mean is unknown find the 95% confidence limits of true mean.

Solution:

Given $n = 900$, $\mu = 3.25$, $\bar{x} = 3.4cm$, $\sigma = 2.61$, $s = 2.61$

Null Hypothesis H_0 : Assume that there is no significant difference between sample mean and population mean. (i.e) $\mu = 3.25$

Alternative Hypothesis H_1 : Assume that there is a significant difference between sample mean and population mean. (i.e) $\mu \neq 3.25$

Level of significance : $\alpha = 5\%$

Test Statistic :

$$z = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{3.4-3.25}{\frac{2.61}{\sqrt{900}}} = 1.724$$

Critical value: The critical value of z for two tailed test at 5% level of significance is 1.96

Conclusion:

$i.e.,\quad z = 1.724 < 1.96 \Rightarrow$ calculated value < tabulated value

Therefore We accept the null hypothesis H_0.

i.e., The sample has been drawn from the population with mean $\mu = 3.25$

To find confidence limit:

95% confidence limits are

$$\bar{x} \mp 1.96\frac{\sigma}{\sqrt{n}} = 3.4 \mp 1.96\left(\frac{2.61}{\sqrt{900}}\right) = 3.4 \mp 0.1705 = (3.57, 3.2295)$$

1. **A lathe is set to cut bars of steel into lengths of 6 centimeters. The lathe is considered to be in perfect adjustment if the average length of the bars it cuts is 6 centimeters. A sample of 121 bars is selected randomly and measured. It is determined that the average length of the bars in the sample is 6.08 centimeters with a standard deviation of 0.44 centimeters.**
   (i) Formulate the hypotheses to determine whether or not the lathe is in perfect adjustment.
   (ii) Compute the test statistic.
   (iii) What is your conclusion?

Solution:

Given $n = 121,\ \bar{x} = 6.08,\ \mu = 6,\ S = 0.44$

Null Hypothesis H_0: $\mu = 6$ i.e., Assume that the lathe is in perfect adjustment

Alternative Hypothesis H_1: $\mu \neq 6$ i.e., Assume that the lathe is not in perfect adjustment.

Level of Significance : $\alpha = 0.05$

i) Test Statistic :

$$z = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}} = \frac{6.08-6}{\frac{0.44}{\sqrt{121}}} = \frac{0.08}{0.04} = 2$$

Table value: Table value at 5% level of significance is 1.96

ii) Conclusion:

Here calculated value > tabulated value

Hence we reject H_0.

2. **The mean life time of a sample of 100 light tubes produced by a company is found to be 1580 hours with standard deviation of 90 hours. Test the hypothesis that the mean lifetime of the tubes produced by the company is 1600 hours.**

Solution:

Given $n = 100$, $\bar{x} = 1580$, $\mu = 1600$, $S = 90$

Null Hypothesis H_0: $\mu = 1600$ i.e., There is no significance difference between the sample mean and population mean

Alternative Hypothesis H1: $\mu \neq 1600$ i.e., There is a significance difference between the sample mean and population mean

Level of Significance : $\alpha = 5\% = 0.05$

Test Statistic :

$$z = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} = \frac{1580 - 1600}{\frac{90}{\sqrt{100}}} = \frac{-20}{9} = -2.22$$

$|z| = 2.22$

Table value: Table value at 5% level of significance is 1.96 (two tailed test)

Conclusion:

Here calculated value > tabulated value

Hence we reject H_0.

Hence the mean life time of the tubes produced by the company may not be 1600 hrs.

Problem based on Test of significance for difference of two means:

The test statistic $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ where σ_1, σ_2 are S.D. of populations.

Test Statistic:

i) $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ If σ is known and $\sigma_1 = \sigma_2$

ii) $Z = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$ If σ is not known and $\sigma_1 \neq \sigma_2$, S_1^2, S_2^2 are known.

Confident Interval:

The confident interval for difference between two population mean for large sample,

(1) when $\sigma(\sigma_1, \sigma_2)$ is known is $\left(\bar{x}_1 - \bar{x}_2\right) \pm z_\alpha \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

(2). when s (s_1, s_2) is known is $\left(\bar{x}_1 - \bar{x}_2\right) \pm z_\alpha \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

1. **In a random sample of size 500, the mean is found to be 20. In another independent sample of size 400, the mean is 15. Could the samples have been drawn from the same population with S.D 4?**

Solution:

Given $\bar{x}_1 = 20$, $\bar{x}_2 = 15$, $n_1 = 500$, $n_2 = 400$, $\sigma = 4$

Null hypothesis H_0 : $\mu_1 = \mu_2$ The samples have been drawn from the same population.

Alternate Hypothesis H_1 : $\mu_1 < \mu_2$ The samples could not have been drawn from same population.

Level of Significance : $\alpha = 5\% = 0.05$ (Two tailed test)

Test statistic: $z = \dfrac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\dfrac{1}{n_2}+\dfrac{1}{n_1}}} = \dfrac{20-15}{4\sqrt{\dfrac{1}{500}+\dfrac{1}{400}}} = 18.6$

Critical value: The critical value of t at 1% level of significance is 2.58
Conclusion: calculated value > table value

H_0 is rejected

The samples could not have been drawn from same population.

2. **Test significance of the difference between the means of the samples, drawn from two normal populations with the same SD using the following data:**

	Size	Mean	Standard Deviation
Sample I	**100**	**61**	**4**
Sample II	**200**	**63**	**6**

Solution:

Given $\bar{x}_1 = 60,\ \bar{x}_2 = 63, s_1 = 4,\ s_2 = 6, n_1 = 100,\ n_2 = 200$

Null hypothesis $H_0:\ \mu_1 = \mu_2$ there is no significance difference between the means of the samples.

Alternate Hypothesis $H_1:\ \mu_1 \neq \mu_2$ there is a significance difference between the means of the samples.

Level of Significance : $\alpha = 5\% = 0.05$ (two tailed test)

Test statistic: $z = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{s_1^{\,2}}{n_2}+\dfrac{s_2^{\,2}}{n_1}}} = \dfrac{61-63}{\sqrt{\dfrac{4^2}{200}+\dfrac{6^2}{100}}} = -3.02 \Rightarrow |z| = 3.02$

Critical value: The critical value of t at 5% level of significance is 1.96
Conclusion: calculated value > table value

H_0 is rejected .Therefore the two normal populations, from which the samples are drawn, may not have the same mean though they may have the same S.D.

3. **A sample of heights of 6400 Englishmen has a mean of 170cm and a S.D of 6.4cm, while a simple sample of heights of 1600 Americans has a mean of 172cm and a S.D of 6.3cm. D the data indicate that Americans are on the average, taller than Englishmen?**

Solution:

Given $\bar{x}_1 = 170,\ \bar{x}_2 = 172, s_1 = 6.4,\ s_2 = 6.3,\ n_1 = 6400,\ n_2 = 1600$

Null hypothesis $H_0:\ \mu_1 = \mu_2$ there is no significance difference between the heights of Americans and Englishmen.

Alternate Hypothesis $H_1:\ \mu_1 < \mu_2$ Americans are on the average, taller than Englishmen

Level of Significance : $\alpha = 5\% = 0.05$ (one tailed test)

Test statistic: $z = \dfrac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{s_1^{\ 2}}{n_2} + \dfrac{s_2^{\ 2}}{n_1}}} = \dfrac{170 - 172}{\sqrt{\dfrac{6.4^2}{6400} + \dfrac{6.3^2}{1600}}} = -11.32 \Rightarrow |z| = 11.32$

Critical value: The critical value of t at 5% level of significance is 1.645

Conclusion: calculated value > table value

H_0 is rejected. We conclude that the data indicate that Americans are on the average, taller than Englishmen.

4. The aveage marks scored by 32 boys is 72 with a S.D of 8, while that for 36 girls is 70 with a S.D of 6. Test at 1%level of significance whether the boys perform beter than girls.

Solution:

Given $\overline{x_1}=72,\ \overline{x_2}=70, s_1=8,\ s_2=6,\ n_1=32,\ n_2=36$

$H_0:\ \mu_1=\mu_2$ (Both perfom are equal)

$H_0:\ \mu_1>\mu_2$ (Boys are better than girls) (one tailed test)

The test statistic: $z=\dfrac{\overline{x_1}-\overline{x_2}}{\sqrt{\dfrac{s_1^{\,2}}{n_2}+\dfrac{s_2^{\,2}}{n_1}}}=\dfrac{72-70}{\sqrt{\dfrac{8^2}{32}+\dfrac{6^2}{36}}}=1.15$

Critical value: The critical value of t at 1% level of significance is 2.33

Conclusion: calculated value < table value

H_0 is accepted. Hence both are equal.

Problem based on Test of significance for single proportion:

To test the significant difference between the sample proportion p and the population proportion P, then we use the test statistic

$z=\dfrac{p-P}{\sqrt{\dfrac{PQ}{n}}}$, where Q = 1 – P

Confident Interval:

The confident interval for population proportion for large sample is $p \mp z_{\alpha}\sqrt{\dfrac{PQ}{n}}$

1. In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers?

Solution:

Given n=600 , Number of smokers=325

p = sample proportion of smokers $\Rightarrow p = 325/600 = 0.5417$

P= Population proportion of smokers in the city $= 1/2 = 0.5 \Rightarrow Q = 0.5$

Null Hypothesis H_0: The number of smokers and non-smokers are equal in the city.

Alternative Hypothesis H_1: P > 0.5 (Right Tailed)

Test Statistic:

$$z=\dfrac{p-P}{\sqrt{\dfrac{PQ}{n}}}=\dfrac{0.5417-0.5}{\sqrt{\dfrac{0.5*0.5}{600}}}=2.04$$

Critical value:
Tabulated value of z at 5% level of significance for right tail test is 1.645.

Conclusion:

Since Calculated value of z > tabulated value of z.

We reject the null hypothesis. The majority of men in the city are smokers.

2. **40 people were attacked by a disease and only 36 survived. Will you reject the hypothesis that the survival rate, if attacked by this disease, is 85% at 5% level of significance?**
Solution:
Given
The Sample proportion, $p = \frac{36}{40} = 0.90$
Population proportion $P = 0.85 \Rightarrow Q = 1 - P = 1 - 0.85 = 0.15$
Null Hypothesis H_0: $P = 0.85$ i.e., There is no significance difference in survival rate
Alternative Hypothesis H_1: $P \neq 0.85$
i.e., There is a significance difference in survival rate.
Level of Significance : $\alpha = 0.05$
Test Statistic :

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.90 - 0.85}{\sqrt{\frac{0.85 \times 0.15}{40}}} = 0.886$$

Table value: Tabulated value of z at 5% level of significance is 1.96

Conclusion : The table value >calculated value
Hence we accept the null hypothesis
Conclude that the survival rate may be taken as 85%.

3. **A Manufacturer of light bulbs claims that an average 2% of the bulbs manufactured by his firm are defective. A random sample of 400 bulbs contained 13 defective bulbs. On the basis of this sample, can you support the manufacturer's claim at 5% level of significance?**
Solution:
Given $n = 400$
$p =$ Sample proportion of defectives = $\frac{X}{n} = \frac{13}{400} = 0.0325$
Null Hypothesis H_0: $P = 2\% = 0.02$ i.e., Assume that 2% bulbs are defective.
Alternative Hypothesis H_1: $P \neq 2\% \neq 0.02$ i.e., Assume that 2% bulbs are non-defective.
Level of significance: $\alpha = 5\% = 0.05$
Test Statistic : $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$

$$z = \frac{0.0325 - 0.02}{\sqrt{\frac{0.02 \times 0.98}{400}}} = \frac{0.0125}{0.0007} = 1.7857$$

Critical value : The critical value of tat 5% level of significance is 1.645 (one tailed test)

Conclusion:
Here calculated value > table value.
So we accept H_0. Hence the manufacturers claim cannot be supported.

4. **A salesman in a departmental store claims that at most 60 percent of the shoppers entering the store leave without making a purchase. A random sample of 50 shoppers should that 35 out of them left without making a purchase. Are these sample reults consistent with the claim of the salesman? Use an LOS of 0.05.**

Solution:

Let p = Sample proportion of shoppers not making a purchase = $\frac{35}{50} = 0.7$

P = Population proportion of shoppers not making a purchase = $60\% = \frac{60}{100} = 0.6$,

and Q = 1 – P = 0.4

H_0: $P = 0.6$ i.e., The claim is accepted

H_1: $P \neq 0.6$ (two tailed test)

The test Statistic is $z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.7-0.6}{\sqrt{\frac{0.6\times0.4}{50}}} = 1.445$

From the table, $z_{0.05} = 1.96$. Since $|z| < z_{0.05}$ $\therefore H_0$ is accepted

Conclusion:
The sample reults are consistent with the claim of the salesman.

Problem based on Test of significance for Two proportion:

To test the significant difference between the sample proportion p_1 and p_2 and the population proportion P, then we use the test statistic

$$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$, where Q = 1 – P

If P is not known, then $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

Confident Interval:

The confident interval for difference between two population proportion for large sample is $(p_1 - p_2) \mp z_\alpha \sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

1. **Before an increase in excise duty on tea, 800 people out of a sample of 1000 were consumers of tea. After the increase in duty, 800 people were consumers of tea in a sample of 1200 persons. Find whether there is significant decrease in the consumption of tea after the increase in duty. Also find confident limit.**

Solution:

Given $n_1 = 1000$, $n_2 = 1200$

p_1 = proportion of tea drinkers before increase inexcise duty= $\frac{800}{1000} = 0.8$

p_2 = proportion of tea drinkers before increase inexcise duty= $\frac{800}{1200} = 0.6667$

Null hypothesis: $H_0 : P_1 = P_2$ there is no significance difference in the consumption of tea before after increase in excise duty

Alternate hypothesis: $H_1 : P_1 \neq P_2$ there is a significance difference in the consumption of tea before after increase in excise duty

Level of significance: $\alpha = 5\% = 0.05$

Test Statistic: $z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

Where

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{(0.8)(1000) + (0.67)(1200)}{1000 + 1200} = 0.7273 \quad \Rightarrow Q = 1 - P = 1 - 0.7273 = 0.2727$$

$$z = \frac{0.8 - 0.6667}{\sqrt{(0.7273)(0.2727)\left(\frac{1}{1000} + \frac{1}{1200}\right)}} = \frac{0.1333}{0.01907} = 6.99$$

Critical value: the critical value of z at 5% level of significance is 1.645

Conclusion:

Here calculated value > table value

$\therefore$ We reject H_0

Hence there is no significance difference in the consumption of tea before after increase in excise duty.

Confident Interval:

The confident interval for difference between two population proportion for large sample is

$$\left(p_1 - p_2\right) \mp z_\alpha \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \left[\left(0.8 - 0.667\right) \mp 1.645\sqrt{0.7273 \times 0.2727\left(\frac{1}{1000} + \frac{1}{1200}\right)}\right]$$
$$= (0.1016, 0.1644)$$

2. **Random samples of 400 men and 600 women asked whether they would like to have a flyover near their residence.200 men and 325 women were in favor of the proposal. Test the hypothesis that proportions of men and women in favor of the proposal are same against that they are not, at 5% level.**

Solution:

Given $n_1 = 400,\ n_2 = 600$

p_1 = proportion of men $= \frac{200}{400} = 0.5$

p_2 = proportion of women $= \frac{325}{600} = 0.541$

Null hypothesis: $H_0 : P_1 = P_2$ Assume that there is no significant difference between the option of men and women as far as proposal of flyover is concerned.

Alternate hypothesis: $H_1 : P_1 \neq P_2$ Assume that there is significant difference between the option of men and women as far as proposal of flyover is concerned

Level of significance: α =5% =0.05 (two tailed)

Test Statistic: $z = \dfrac{p_1 - p_2}{\sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

Where $P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \dfrac{(400)(0.5) + (600)(0.541)}{400 + 600} = 0.525 \quad \Rightarrow Q = 1 - P = 1 - 0.525 = 0.475$

$$z = \frac{0.5 - 0.541}{\sqrt{(0.525)(0.475)\left(\frac{1}{400} + \frac{1}{600}\right)}} = \frac{-0.041}{0.032} = -1.34 \quad \Rightarrow |z| = 1.34$$

Critical value: the critical value of z at 5% level of significance is 1.96

Conclusion:

Here calculated value < table value

$\therefore$ We accept H_0 at 5% level of significance.

Hence There is no difference between the option of men and women as far as proposal of flyover are concerned.

3. A machine puts out 16 imperfect articles in a sample of 500. After the machine is overhauled, it puts out 3 imperfect articles in a batch of 100. Has the machine improved?

Solution:

Hypothesis:

$H_0: P_1 = P_2$

$H_1: P_1 > P_2$

Level of Significance : $\alpha = 0.05$

Test Statistic : $Z = \dfrac{p_1 - p_2}{\sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

Analysis:

The Sample proportion,

$p_1 = \dfrac{16}{500} = 0.032, \quad p_2 = \dfrac{3}{100} = 0.03, \quad P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.032 \;\; \& \; Q = 1 - P = 0.968$

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.032 - 0.03}{\sqrt{0.032 \times 0.968\left(\frac{1}{500} + \frac{1}{100}\right)}} = 0.1037$$

Table value : $Z_\alpha = 1.645$

Conclusion:

Calculated value < table value

Hence we accept the null hypothesis and conclude that the machine has not improved after overhauling.