Sampling Theory.

A large collection of individuals or attributes or numerical data can be understood as a population or universe.

A finite subset of the population is called a sample. The number of individuals in a sample is called a sample size.

If the sample size : $n \geq 30$: then it is said to be large sample.
And the sample size $n < 30$ is said to be small samples.

Consider all possible samples of size 'n' which can be drawn from a given population at random. For each sample, we can compute quantities like mean, standard deviation, etc. These quantities will not be the same. Suppose we group these characteristics according to their frequencies, the frequency distributions so generated are called sampling distributions.

These can be distinguished as sampling distribution of mean, standard deviation, etc. The sampling distributions of large samples is assumed to be a normal distribution.

The standard deviation of a sampling distribution is also called the standard error. The reciprocal of the standard error is called precision.

Testing of Hypothesis.

One of the objectives of the sampling theory is to evolve appropriate rules that enable us to make decisions about populations on the basis of the information available in respect of samples. Such decisions are called statistical decisions.

For reaching statistical decisions, we start with some assumptions or guesses about the populations involved. Such assumptions / guesses, which may or may not be true, are called statistical hypotheses.

The hypothesis formulated for the purpose of its rejection under the assumption that it is true is called the Null Hypothesis denoted by H_0. Any hypothesis which is complimentary to the null hypothesis is called Alternate Hypothesis (H_1).

Procedures which enable us to decide whether to accept or reject a hypothesis or to determine whether observed samples differ significantly from expected results are called Tests of Hypothesis, Tests of significance or Rules of decision.

Type I and Type II Errors

By an error of judgement suppose we reject a hypothesis H when it should have been accepted, then we say that Type I error has occurred. On the other hand, suppose we accept a hypothesis when it should have been rejected, then we say that Type II error has occurred.

	Accepting the hypothesis	Rejecting the hypothesis
Hypotheses true	Correct decision	Wrong decision (Type-I error)
Hypotheses false	Wrong decision (Type-II error)	Correct decision

Levels of Significance.

The probability level, below which leads to the rejection of hypothesis is known as the significance level.

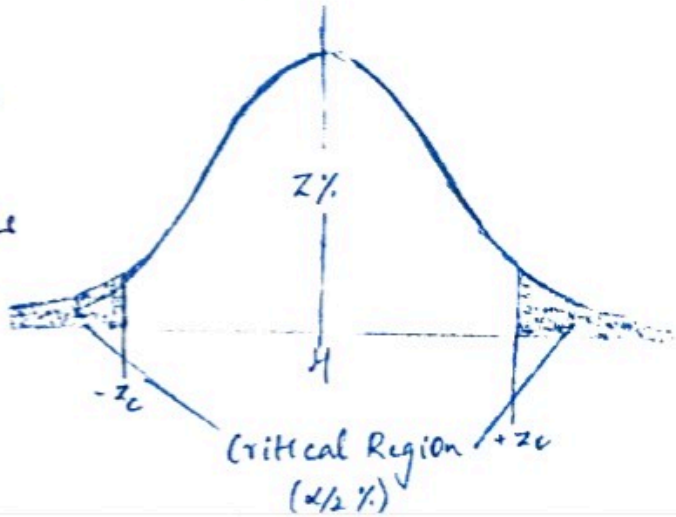
This probability is conventionally fixed at 0.05 or 0.01 ie 5% or 1%. These are called significance levels.

Suppose we have a normal population with mean μ and S.D σ. If $\bar{x}$ is the sample mean of random sample of size n, the quantity z defined by $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ is called the standard normal variate (S.N.V).

From the table of normal areas we find that 95% of the area lies between $z = -1.96$ and $z = +1.96$. In other words we can say that with 95% confidence that z lies between -1.96 and $+1.96$. 5% level of significance is denoted by $z_{0.05}$.

Similarly, we can say that with 99% confidence that z lies between -2.58 and $+2.58$. 1% level of significance is denoted by $z_{0.01}$

The constants 1.96, 2.58, etc in confidence limits are called confidence co-efficients denoted by z_c. The set of values of z that are outside the confidence interval $(-z_c, z_c)$ constitutes the region of rejection or the critical region.



One tailed and two tailed tests.

In our test of acceptance or non-acceptance of a hypothesis we concentrated on the value of z on both sides of the mean. This can be categorically stated that the focus of attention lies in the two "tails" of distribution and hence such a test is called two tailed test.

Sometimes we will be interested in the extreme values to only one side of the mean in which case the region of significance will be a region to one side of the distribution. The area of such a region will be equal to the level of significance itself. Such a test is called one tailed test.

Test	Critical values of z	
	5% level	1% level
One-tailed test	-1.645 or 1.645	-2.33 or 2.33
Two tailed test	-1.96 and 1.96	-2.58 and 2.58

Q Nos.

Statistical Inference - I

Test of significance for large samples

Working Procedure:

1. Write the null hypothesis H_0.
2. Find the calculated value of z using

$$|z| = \left|\frac{p - P}{SE(P)}\right| \quad \text{where } SE(P) = \sqrt{\frac{PQ}{n}} \quad \text{if Probability } P \text{ is known.}$$

$$|z| = \left|\frac{\bar{x} - \mu}{SE(\bar{x})}\right| \quad \text{where } SE(\bar{x}) = \sqrt{\frac{SD^2}{n}} \quad \text{if } \overset{\text{mean}}{\bar{x}} \text{ is known.}$$

3. Find the critical value using the table.
4. If Calculated value < Tabulated value, accept H_0.
   H_0 is the conclusion.
5. If Calculated value > Tabulated value, reject H_0.
   H_1 is the conclusion.

(1) A coin is tossed 1000 times and heads turned up 540 times. Decide on the hypothesis that coin is unbiased at 1% level of significance.

Soln: H_0 : Coin is unbiased.

$$P = \frac{1}{2} \quad Q = \frac{1}{2} \qquad p = \frac{540}{1000} = 0.54$$

// For coin toss

$$SE(p) = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{\frac{1}{2}\,\frac{1}{2}}{1000}} = 0.0158$$

$$|z| = \left|\frac{p - P}{SE(P)}\right| = \left|\frac{0.54 - 0.5}{0.0158}\right| = 2.5316$$

At 1% level of significance, $z_\alpha = 2.58$

$\therefore$ Calculated value < Tabulated value

$\Rightarrow$ Accept H_0.

Coin is unbiased

Q Nos.

(2) A Coin is tossed 400 times and heads turned up 216 times. Test the hypothesis that coin is unbiased at 5% level of sig.

Soln: H_0 : Coin is unbiased $P = \frac{1}{2}$ $Q = \frac{1}{2}$

$$p = \frac{216}{400} = 0.54$$

$$SE(p) = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{\frac{1}{2}\,\frac{1}{2}}{400}} = 0.025$$

$$|z| = \left|\frac{p - P}{SE(P)}\right| = \left|\frac{0.54 - 0.5}{0.025}\right| = 1.6$$

At 5% level of significance $z_\alpha = 1.96$

$\therefore$ Calculated value < Tabulated value.

$\Rightarrow$ Accept H_0. Coin is unbiased.

(3) In 324 throws of a six faced dice, an odd number turned up 181 times. Is it reasonable to think dice is unbiased at 1% level?

Soln: H_0 : Dice is unbiased

$P(\text{Odd number}) = 3/6 = 1/2$ $Q = 1/2$ $SE(P) = \sqrt{\frac{PQ}{n}} = 0.0278$

$p = 181/324 = 0.5586$

$$|z| = \left|\frac{p - P}{SE(P)}\right| = \left|\frac{0.5586 - 0.5}{0.0278}\right| = 2.1094$$

At 1% level of significance, $z_\alpha = 2.58$

$\therefore$ Calculated value < Tabulated value

$\Rightarrow$ Accept H_0. Dice is unbiased.

(4) A dice was thrown 9000 times and a throw of 5 or 6 was obtained 3240 times. On the assumption of random throwing, do the data indicate an unbiased dice at 1% level of significance?

Soln: H_0 : Dice is unbiased

$P(5 \text{ or } 6) = 2/6 = 1/3$ $Q = 2/3$ $SE(P) = \sqrt{\frac{PQ}{n}} = 0.005$

$p = 3240/9000 = 0.36$

$$|z| = \left|\frac{p - P}{SE(P)}\right| = \left|\frac{0.36 - 0.333}{0.005}\right| = 5.33$$

At 1% level of significance, $z_\alpha = 2.58$

$\therefore$ Calculated value > Tab value

$\Rightarrow$ Reject H_0.

Dice is biased.

Q Nos.

(5) A dice is thrown 960 times and it falls with 5 upwards 184 times. Is the dice biased at 5% level?

Soln: H_0 : Dice is unbiased

$P(5 \text{ upwards}) = 1/6 \quad Q = 5/6 \quad SE(P) = \sqrt{\frac{PQ}{n}} = 0.012$

$p = \frac{184}{960} = 0.1917$

$$|z| = \left|\frac{p - P}{SE(P)}\right| = \left|\frac{0.1917 - 0.1667}{0.012}\right| = 2.0861$$

At 5% level of significance, $z_\alpha = 1.96$

$\therefore$ Calculated value > Tabulated value

$\Rightarrow$ Reject H_0. $\Rightarrow$ Dice is biased.

(6) A dice was thrown 1200 times and the number '6' was observed 236 times. Can the dice be considered fair at 0.01 level of significance

Soln: H_0 : Dice is unbiased.

$P(6) = 1/6 \quad Q = 5/6 \quad SE(P) = \sqrt{\frac{PQ}{n}} = 0.0108$

$p = 236/1200 = 0.1967$

At 1% level of significance, $z_\alpha = 2.58$

$$|z| = \left|\frac{p - P}{SE(P)}\right| = \left|\frac{0.1967 - 0.1667}{0.0108}\right| = 2.7809$$

$\therefore$ Calculated value > Tabulated value

$\Rightarrow$ Reject H_0 $\Rightarrow$ Dice is biased.

(7) Results extracts revealed that in a certain school over a period of 5 years, 725 students had passed and 615 students had failed. Test the hypothesis that success and failure are in equal proportions.

Soln: H_0 : Success and failure are in equal proportions

$P = 1/2 \quad Q = 1/2 \quad SE(P) = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{1/2 \times 1/2}{725+615}} = 0.0137$

$n =$ Total no. of students $= 725 + 615 = 1340$

$p = 725/1340 = 0.5410$

$$|z| = \left|\frac{p - P}{SE(P)}\right| = \left|\frac{0.5410 - 0.5}{0.0137}\right| = 2.996$$

$z = 2.996 > 2.58$, > 1.96 } Hypothesis is rejected at 1% and 5% levels of significance.

Note: If atleast / atmost $\rightarrow$ One tailed test
If not equal $\rightarrow$ Two tailed test

Q Nos.

(8) A manufacturer claimed that atleast 95% of the equipment which he supplied to a factory confirmed to specifications. An exam of 200 pieces of equipments revealed that 18 of them were faulty. Test his claim at a significance level of 1% and 5%.

Soln: H_0 : Only 95% of equipment confirmed to specifications.

$P = 95/100 \quad Q = 5/100 \quad SE(P) = \sqrt{\frac{PQ}{n}} = 0.0154$

$p = \frac{200-18}{200} = 0.91$

$|z| = \left|\frac{p-P}{SE(P)}\right| = \left|\frac{0.91-0.95}{0.0154}\right| = 2.5974$

At 1% level of significance $z_\alpha = 2.33$

5% level of significance $z_\alpha = 1.645$

$\therefore$ Calculated value of > Tabulated value.

$\therefore$ H_0 is rejected at both levels using one-tailed test.

The claim that only 95% of equipment confirmed to specifications is not true

(9) In a sample 400 parts manufactured by a factory the number of defective parts was found to be 30. The company however claimed that atmost 5% of their products is defective. Test the claim at 5% level.

Soln: H_0 : 5% of Products is defective.

$P = 5/100 \quad Q = 95/100 \quad SE(P) = \sqrt{\frac{PQ}{n}} = 0.0109$

$p = \frac{30}{400} = 0.075 \quad |z| = \left|\frac{p-P}{SE(P)}\right| = 2.2936$

$\Rightarrow$ Reject H_0

At 5% level of significance, $z_\alpha = 1.645$ $\therefore$ Calculated value > Tab value

The claim that atmost 5% of Products is defective is not true.

(10) A manufacturer claims that only 4% of his products are defective. A random sample of 600 products contained 36 defectives. Test the claim of the manufacturer at 5% level of significance.

Soln: H_0 : 4% of Products are defective.

$P = 4/100 \quad Q = 96/100 \quad SE(P) = \sqrt{\frac{PQ}{n}} = 0.008$

$p = 36/600 = 0.06 \quad |z| = \left|\frac{p-P}{SE(P)}\right| = 2.5$

Not just

At 5% level of significance, $z_\alpha = 1.96$

$\therefore$ Calculated value > Tabulated value

$\Rightarrow$ Reject H_0. The claim that only 4% of Products are defective is not true

	Mean	S.D
Sample	$\bar{x}$	s
Population	μ	σ

Q Nos.

11. A sample of 900 members is found to have a mean of 3.4 cm. Can it be reasonably regarded as a truly random sample from a large population with mean 3.25 cm and SD 1.61 cm.

Soln: H_0: Sample is taken from a large population with mean 3.25

$\mu = 3.25 \quad \sigma = 1.61 \quad \bar{x} = 3.4 \quad n = 900$

$SE(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = 0.0537$

$$|z| = \left|\frac{\bar{x} - \mu}{SE(\bar{x})}\right| = \left|\frac{3.4 - 3.25}{0.0537}\right| = 2.795$$

At 5% level of significance, $z_c = 1.96$

$\therefore$ Calculated value > Tabulated value

$\Rightarrow$ Reject H_0. Sample is not taken from large population with mean 3.25

12. The mean life time of a sample of 100 fluroscent bulbs manufactured by a company is found to be 1570 hours with SD of 120 hours. Test the hypothesis that mean life time of bulbs produced by company is 1600 hrs.

Soln: H_0: Mean life-time of bulbs is 1600 hours.

$\mu = 1600 \quad \bar{x} = 1570 \quad s = 120 \quad n = 100$

$SE(\bar{x}) = \sqrt{\frac{s^2}{n}} = 12$

$$|z| = \left|\frac{\bar{x} - \mu}{SE(\bar{x})}\right| = \left|\frac{1570 - 1600}{12}\right| = 2.5$$

$|z| = 2.5 > 1.96 \rightarrow H_0$ rejected at 5% level.

$< 2.58 \rightarrow H_0$ accepted at 1% level.

$\therefore$ Mean life-time of bulbs is 1600 hours, is accepted at 1% level.

13. The average marks in mathematics of sample of 100 students were 51 with SD of 6 marks. Could this have been a random sample from a population with average marks 50 at 5% level of significance.

Soln: H_0: Sample is drawn from Population ~~has~~ whose average is 50 marks.

$\mu = 50 \quad \bar{x} = 51 \quad s = 6 \quad n = 100 \quad SE(\bar{x}) = 0.6$

$|z| = 1.667 < 1.96$ H_0 is accepted.

$\therefore$ Sample is drawn from population whose average is 50 marks.

Q Nos.

14. The mean weight obtained from random sample of size 100 is 64 grams. The SD of weight distribution of population is 3 grm. Test the hypothesis that the mean weight of population is 67 grms at 5% level of significance.

Soln: H_0: Mean weight of population is 67 g.

$\mu = 67 \quad \bar{x} = 64 \quad \sigma = 3 \quad n = 100$

$$SE(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = \frac{3}{10} = 0.3$$

$$|z| = \left|\frac{\bar{x} - \mu}{SE(\bar{x})}\right| = \left|\frac{64-67}{0.3}\right| = 10.$$

At 5% level of significance $z_c = 1.96$

$\therefore$ Calculated value > Tabulated value

$\Rightarrow$ Reject H_0. Mean weight of population is not 67 g.

15. It is claimed that a random sample of 49 tyres has a mean life of 15,200 kms. Is the sample drawn from a population whose mean of 15,150 kms and whose SD is 1200 kms? Test the hypothesis at 0.05 significance level.

Soln: H_0: Sample is drawn from a population whose mean is 15,150.

$\mu = 15{,}150 \quad \sigma = 1200 \quad \bar{x} = 15{,}200 \quad n = 49$

$$SE(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = 171.4286$$

$$|z| = \left|\frac{\bar{x} - \mu}{SE(\bar{x})}\right| = \left|\frac{15200 - 15150}{171.4286}\right| = 0.2917$$

At 5% level of significance, $z_\alpha = 1.96$

$\therefore$ Calculated value < Tabulated value

$\Rightarrow$ Accept H_0. Sample is drawn from population of mean 15,150.

Q Nos.

16. A stenographer claims that she can type at the rate of 120 words per minute. Can we reject her claim on the basis of 100 trials in which she demonstrates a mean of 116 words with SD of 15 words? Check at 5% level.

Soln: H_0: She can type at the rate of 120 words per minute.

$\mu = 120 \quad \bar{x} = 116 \quad S = 15 \quad n = 100$

$$SE(\bar{x}) = \sqrt{\frac{s^2}{n}} = \frac{15}{10} = 1.5$$

$$|z| = \left|\frac{\bar{x} - \mu}{SE(\bar{x})}\right| = \left|\frac{116 - 120}{1.5}\right| = 2.6667$$

At 5% level, $z_\alpha = 1.96$

Calculated value > Tabulated value

$\Rightarrow$ Reject H_0.

She cannot type at the rate of 120 words per minute.

17. If a mean breaking strength of copper wire is 575 lbs with SD 8.3 lbs. How large a sample must be used in order that there be one chance in 100 that mean breaking strength of sample is less than 572 lbs ($z_\alpha = 2.33$)

One-tailed

Soln: H_0: Mean breaking strength is less than 575 lbs.

$\mu = 575 \quad \bar{x} = 572 \quad \sigma = 8.3 \quad n = ?$

$$SE(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = \frac{8.3}{\sqrt{n}}$$

$$|z| = \left|\frac{\bar{x} - \mu}{SE(\bar{x})}\right| = \left|\frac{572 - 575}{8.3/\sqrt{n}}\right|$$

To find n, such that $\mu < 572$

To find n, such that H_0 is rejected.

ie $|z| > 2.33$

$$\left|\frac{3}{8.3/\sqrt{n}}\right| > 2.33$$

$n > 41.55$

$n > 42$

Q Nos.

18. In an exam given to students at a large number of different schools the mean grade was 74.5 and SD 8. At one particular school where 200 students took exam the mean grade was 75.9. Discuss the significance of this result at both 5% and 1%.

Soln: H_0 : Sample is drawn from population whose mean = 74.5

$n = 200 \quad \mu = 74.5 \quad \sigma = 8 \quad \bar{x} = 75.9$

$SE(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = 0.5657 \qquad |z| = \left|\frac{\bar{x}-\mu}{SE(\bar{x})}\right| = 2.4748$

At 5% level of significance, $Z_\alpha = 1.96$.

Calculated value > Tabulated value. $\Rightarrow$ Reject H_0.

At 1% level of significance, $Z_\alpha = 2.58$

Calculated value < Tabulated value $\Rightarrow$ Accept H_0.

19. A Principal at a school claims that the students in his school are above average intelligence. A random sample of 30 students IQ scores have a mean score 112.5. Is there sufficient evidence to support the principal's claim. The population mean IQ is 100 with SD 15. (5% los)

Soln: H_0 : Sample is drawn from population whose mean is 100.

$n = 30 \quad \bar{x} = 112.5 \quad \mu = 100 \quad \sigma = 15$

$SE(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = 2.7386 \qquad |z| = \left|\frac{\bar{x}-\mu}{SE(\bar{x})}\right| = 4.5644$

At 5% level of significance, $Z_\alpha = 1.96$

Calculated value > Tabulated value $\Rightarrow$ Reject H_0.

Sample is not drawn from population.

20. A sample of 400 students is found to have mean height of 160 cms. Can it be reasonably regarded as a sample from large population with mean height 162.5 cm and SD 4.5 cm (1% los)

Soln: H_0 : Sample is taken from large population whose mean is 162.5 cm.

$n = 400 \quad \bar{x} = 160 \quad \mu = 162.5 \quad \sigma = 4.5$

$SE(\bar{x}) = 0.225 \quad |z| = 11.1 > 2.58$. Reject H_0.

* The mean of certain normal population is equal to the SE of the mean of samples of 100 from that distribution. Find the probability that the mean of sample of 25 from distribution will be negative.

Soln: Given $\mu = SE(\bar{x}) = \sqrt{\frac{(SD)^2}{n}} = \sqrt{\frac{\sigma^2}{100}} = \frac{\sigma}{10}$

To find $P(\bar{x} < 0)$ when $n = 25$

$n = 25 \quad SE(\bar{x}) = \sqrt{\frac{\sigma^2}{25}} = \frac{\sigma}{5}$

$$\therefore P(\bar{x} < 0) = P\left(\frac{\bar{x} - \mu}{SE(\bar{x})} < \frac{0 - \mu}{SE(\bar{x})}\right)$$

$$= P\left(Z < -\frac{\sigma/10}{\sigma/5}\right)$$

$$= P(Z < -1/2)$$

$$= P(Z > 1/2)$$

$$= 0.5 - A(0.5)$$

$$= 0.5 - 0.1915 = 0.3085$$

Q Nos.

Test of significance for difference of mean

Let $\bar{x}_1$ and $\bar{x}_2$ be two sample means of sizes n_1 and n_2 taken from two different populations with mean μ_1 and μ_2 and SD σ_1 and σ_2. To test the significance of the difference between means $\bar{x}_1$ and $\bar{x}_2$, the test statistic under the null hypothesis H_0 that there is no significant difference between two samples means is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Comparison of large samples

Working rule

1. Write the null hypothesis H_0.
2. Find the calculated value using

$$|z| = \left| \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)} \right| \quad \text{if } \bar{x}_1, \bar{x}_2 \text{ are known}$$

$$|z| = \left| \frac{p_1 - p_2}{SE(p_1 - p_2)} \right| \quad \text{if } p_1, p_2 \text{ are known}$$

3. Find the critical value using the table.
4. If calculated value < Tabulated value
   $\Rightarrow$ Accept H_0. H_0 is the conclusion.
5. If calculated value > Tabulated value
   $\Rightarrow$ Reject H_0. H_1 is the conclusion.

where $SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ if s_1, s_2 are known

$= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ if σ_1, σ_2 are known

$= \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ if σ is known

$SE(p_1 - p_2) = \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$ if P_1, P_2 are known

$= \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ if p_1, p_2 are known

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

Q Nos.

1. A sample of 100 electric bulbs produced by manufacturer A showed a mean lifetime of 1190 hours and a SD of 90 hours. A sample of 75 bulbs produced by manufacturer B showed a mean lifetime of 1230 hours with SD of 120 hours. Is there a difference between the mean lifetime of two brands at significant level of 0.05.

Soln: H_0 : There is no difference between the mean lifetime of two brands

$n_1 = 100$ $\bar{x}_1 = 1190$ $s_1 = 90$ $\bar{x}_2 = 1230$ $s_2 = 120$ $n_2 = 75$

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{90^2}{100} + \frac{120^2}{75}} = 16.5227$$

$$|z| = \left|\frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}\right| = \left|\frac{1190 - 1230}{16.5227}\right| = 2.4209$$

At 5% level of significance $z_c = 1.96$

$\therefore$ Calculated z > Tabulated z

$\Rightarrow$ Reject H_0. There is difference between mean lifetime of 2 brands.

2. A sample of height of 6400 soldiers has a mean of 67.85 inches and SD of 2.56 inches while a sample of height of 1600 sailors has a mean of 68.55 inches and a SD of 2.52 inches. Does the data indicate that the sailors are on an average taller than soldiers? Use 0.05 level of significance.

Soln: H_0 : There is no significant difference between height of sailors & soldiers. ~~Sailors are not taller than soldiers~~

$n_1 = 6400$ $\bar{x}_1 = 67.85$ $s_1 = 2.56$ $n_2 = 1600$ $\bar{x}_2 = 68.55$ $s_2 = 2.52$

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 0.0707$$

$$|z| = \left|\frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}\right| = 9.9064$$

At 0.05 level of significance $z_c = 1.96$

$\therefore$ Calculated value > Tabulated value

$\Rightarrow$ Reject H_0. ~~Sailors are taller than soldiers.~~

There is a difference between heights of sailors & soldiers.

Q Nos.

3. The average income of people was ₹210 with SD of ₹10 in a sample of 100 people in a city. For another sample of 150 people, the average income was ₹220 with SD of ₹12. The SD of incomes of the people of city was ₹11. Test whether there is any significant difference between the average incomes of the localities. (Use 5% level)

Soln: H_0 : There is no significant difference between the average income of localities.

$n_1 = 100 \quad \bar{x}_1 = 210 \quad s_1 = 10 \qquad n_2 = 150 \quad \bar{x}_2 = 220 \quad s_2 = 12$

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 1.4$$

$$|z| = \left|\frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}\right| = 7.1429$$

At 5% level of significance, $Z_c = 1.96$

$\therefore$ Calculated value > Tabulated value.

$\Rightarrow$ Reject H_0. There is difference between average income of localities

4. The means of simple samples of size 1000 and 2000 are 67.5 cm and 68 cm respectively. Can the samples be regarded as drawn from the same population of SD 2.5 cm? (Use 5% level)

Soln: H_0 : Samples are regarded as drawn from same population.

$n_1 = 1000 \quad \bar{x}_1 = 67.5 \quad n_2 = 2000 \quad \bar{x}_2 = 68 \quad \sigma = 2.5$

$$SE(\bar{x}_1 - \bar{x}_2) = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 0.0968$$

$$|z| = \left|\frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}\right| = 5.164$$

At 5% level of significance $Z_\alpha = 1.96$

$\therefore$ Calculated value > Tabulated value.

$\Rightarrow$ Reject H_0. Samples cannot be regarded as drawn from same population

Q Nos.

5. Intelligent tests were given to two groups of boys and girls. Test whether there is any significant difference between mean scores (use 1% level)

	Mean	SD	Size
Girls	75	8	60
Boys	73	10	100

Soln: H_0 : There is no difference between mean scores.

$n_1 = 60 \quad \bar{x}_1 = 75 \quad S_1 = 8 \qquad n_2 = 100 \quad \bar{x}_2 = 73 \quad S_2 = 10$

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = 1.4376$$

$$|z| = \left| \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)} \right| = 1.3912$$

At 1% level of significance $Z_\alpha = 2.58$

$\therefore$ Calculated value < Tabulated value.

$\Rightarrow$ Accept H_0. There is no difference between mean scores.

6. In order to make a survey of buying habits, two markets A and B are at two different parts of a city. 400 women shoppers are choosen at random in market A. Their daily average expenditure on food is found to be ₹250 with SD of ₹40. The figures are ₹220 and ₹55 respectively in market B where also 400 women shoppers are choosen at random. Test at 1% level whether the average daily food expenditures of two shoppers are equal.

Soln: H_0 : The average daily food expenditures are equal.

$n_1 = 400 \quad \bar{x}_1 = 250 \quad S_1 = 40 \qquad n_2 = 400 \quad \bar{x}_2 = 220 \quad S_2 = 55$

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = 3.4004$$

$$|z| = \left| \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)} \right| = 8.8226$$

At 1% level of significance $Z_\alpha = 2.58$

$\therefore$ Calculated value > Tabulated value

$\Rightarrow$ Reject H_0. The average daily food expenditures are not equal.

Q Nos.

7. In an exit poll survey it was revealed that 600 voters in one locality and 400 voters from another locality favoured 55% and 48% respectively for a particular party to come to power. Test the hypothesis that there is a difference in the locality in respect of the opinion at 1% los.

Soln: H_0: There is no difference wrt the opinion.

$p_1 = 55\% = 0.55$ $\quad n_1 = 600$ $\quad p_2 = 48\% = 0.48$ $\quad n_2 = 400$

$P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.522$ $\quad Q = 1 - P = 0.478$

$$SE(p_1 - p_2) = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0.0322$$

$$|Z| = \left|\frac{p_1 - p_2}{SE(p_1 - p_2)}\right| = \left|\frac{0.55 - 0.48}{0.0322}\right| = 2.171$$

At 1% level of significance $Z_\alpha = 2.58$

$\therefore$ Calculated value < Tabulated value

$\Rightarrow$ Accept H_0. There is no difference wrt opinion.

8. One type of aircraft is found to develop engine trouble in 5 flights out of 100 and another type in 7 flights out of 200. Is there a significant difference in the two types of aircrafts so far as engine defects are concerned? Test at 5% los.

Soln: H_0: There is no significant difference in two types of aircrafts

$p_1 = \dfrac{5}{100} = 0.05$ $\quad n_1 = 100$ $\quad p_2 = \dfrac{7}{200} = 0.035$ $\quad n_2 = 200$

$P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.04$ $\quad Q = 1 - P = 0.96$

$$SE(p_1 - p_2) = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0.024$$

$$|Z| = \left|\frac{p_1 - p_2}{SE(p_1 - p_2)}\right| = 0.625$$

At 5% level of significance $Z_\alpha = 1.96$

Calculated value < Tabulated value

$\Rightarrow$ Accept H_0. There is no difference in two types of aircrafts.

9. A machine produced 16 defective articles in a batch of 500. After over heating it produced 3 defectives in a batch of 100. Has the machine improved? (At 5% l.o.s)

H_0: The machine has not improved.

$n_1 = 500 \quad p_1 = \frac{16}{500} \qquad n_2 = 100 \quad p_2 = \frac{3}{100}$

$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.0317 \qquad Q = 1 - P = 0.9683$

$SE(p_1 - p_2) = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0.0192 \qquad |z| = \left|\frac{p_1 - p_2}{SE(p_1 - p_2)}\right| = 0.1042$

At 5% level of significance $z_\alpha = 1.96$

Calculated value < Tabulated value.

$\Rightarrow$ Accept H_0. The machine has not improved.

~~10.~~

10. In a sample of 600 men from a certain city, 450 are found to be smokers. In another sample of 900 men from another city, 450 are smokers. Does this indicate that the cities are significantly different wrt smoking habit. Test at 5% los

Soln: H_0: There is no significant difference between cities wrt smoking habit of men.

$p_1 = \frac{450}{600} \quad n_1 = 600 \qquad p_2 = \frac{450}{900} \quad n_2 = 900$

$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.6 \qquad Q = 0.4$

$SE(p_1 - p_2) = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0.0258$

$|z| = \left|\frac{p_1 - p_2}{SE(p_1 - p_2)}\right| = 9.6825$

At 5% level of significance $z_\alpha = 1.96$

Calculated z > Tabulated z

$\therefore$ Reject H_0.

There is difference between cities wrt smoking habits of men.

Q Nos.

11. In a city A, 20% of random sample of 900 school boys were 6ft tall. In another city B, 18.5% of random sample of 1600 school boys had same height. Is the difference between proportions significant?

Soln: H_0 : There is no significant difference between proportions

$n_1 = 900$ $\quad n_2 = 1600$ $\quad P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.1904$ $\quad Q = 0.8096$

$p_1 = 20\% = 0.2$ $\quad p_2 = 18.5\% = 0.185$

$SE(p_1 - p_2) = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0.0164$ $\quad |z| = \left|\frac{p_1 - p_2}{SE(p_1 - p_2)}\right| = 0.9169$

At 1% level of significance $Z_\alpha = 2.58$

At 5% level of significance $Z_\alpha = 1.96$

Calculated value < Tabulated value.

$\Rightarrow$ Accept H_0. There is no significant difference between proportions.

12. In 2 large populations there are 30% and 25% of left handed people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from 2 populations.

Soln: H_0 : There is no significant difference to be hidden.

$n_1 = 1200$ $\quad n_2 = 900$ $\quad P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.2786$

$p_1 = 30\% = 0.3$ $\quad p_2 = 25\% = 0.25$ $\quad Q = 0.7214$

$SE(p_1 - p_2) = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0.0198$ $\quad |z| = \left|\frac{p_1 - p_2}{SE(p_1 - p_2)}\right| = 2.5253$

At 5% level of significance $Z_\alpha = 1.96$

Calculated value > Tabulated value.

$\Rightarrow$ Reject H_0.

At 1% level of significance. $Z_\alpha = 2.58$

Calculated value < Tabulated value

$\Rightarrow$ Accept H_0.

There is no significant difference to be hidden.

Q Nos.

13. 500 articles from a factory are examined and found that 2% are defective. 800 similar articles from a second factory are found to have only 1.5% defective. Can it be reasonably concluded that the product of first factory is inferior to those of second?

Soln: ~~Ho : The product of first factory is not inferior to those of second factory.~~

Ho : There is no significant difference between the defectives of 2 factories.

$n_1 = 500$ $\quad n_2 = 800$ $\quad P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.0169$ $\quad Q = 0.9831$

$p_1 = 2\% = 0.02$ $\quad p_2 = 1.5\% = 0.015$

$SE(p_1 - p_2) = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0.0073$ $\quad |z| = \left|\frac{p_1 - p_2}{SE(p_1 - p_2)}\right| = 0.6804$

At 1% level of significance $Z_\alpha = 2.58$

At 5% level of significance $Z_\alpha = 1.96$

Calculated value < Tabulated value

$\Rightarrow$ Accept Ho. There is no significant difference

14. Random sample of 1000 engineering students from a city A and 800 from city B were taken. It was found that 400 students in each of the sample were from Bangalore. Does this data reveal a significant difference between the 2 cities in respect to this selection?

Soln: Ho : There is no significant difference in 2 cities in respect to selection

$n_1 = 1000$ $\quad p_1 = \frac{400}{1000}$ $\quad n_2 = 800$ $\quad p_2 = \frac{400}{800}$ $\quad P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{4}{9}$ $\quad Q = \frac{5}{9}$

$SE(p_1 - p_2) = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0.0236$ $\quad |z| = \left|\frac{p_1 - p_2}{SE(p_1 - p_2)}\right| = 4.2373$

At 1% level of significance $Z_\alpha = 2.58$

At 5% level of significance $Z_\alpha = 1.96$

Calculated value > Tabulated value

$\Rightarrow$ Reject Ho. There is significant difference in 2 cities in respect to selection.

Q. Nos.

Confidence limits

An interval which is likely to contain the parameter is called confidence interval. The limits of the confidence interval are called confidence limits.

Ex:- While we check BP, 80 and 120 are lower and upper limits. Interval 80-120 is called confidence interval.

The probability that the confidence interval contains the parameter is called confidence co-efficients. It is denoted by $1-\alpha$.

Confidence limits for mean

$$\bar{x} - Z_\alpha \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_\alpha \frac{\sigma}{\sqrt{n}}$$

(i) 95% confidence limits are : $\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}$

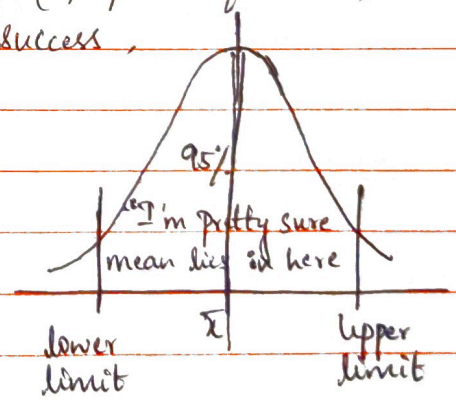
(ii) 99% confidence limits are : $\bar{x} - 2.58 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.58 \frac{\sigma}{\sqrt{n}}$

Probable limits

With p representing probability of success (Proportion of success) and $\sqrt{pq/n}$ is the standard error proportional to success, the probable limit are $p \pm Z_\alpha \sqrt{pq/n}$

(i) 95% probable limits are : $p \pm 1.96 \sqrt{\frac{pq}{n}}$

(ii) 99% probable limits are : $p \pm 2.58 \sqrt{\frac{pq}{n}}$



* A 95% confidence limit means that there is a 95% probability that the true value is within the range of the error bar. (See above)

Q. Nos.

1. The mean weight obtained from a random sample of size 100 is 64 g. The SD of weight distribution of the population is 3g. Test the statement that mean weight of population is 67g at 5% level of significance. Also set up 99% confidence limits of the mean weight of population.

Soln: (i) H_0 : Mean weight of Population is 67 g.

$n = 100 \quad \bar{x} = 64 \quad \sigma = 3 \quad \mu = 67$

$SE(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = 0.3 \qquad |z| = \left|\frac{\bar{x}-\mu}{SE(\bar{x})}\right| = 10$

At 5% level of significance $Z_\alpha = 1.96$

Calculated value > Tabulated value

$\Rightarrow$ Reject H_0. Mean weight is not 67g.

(ii) 99% confidence limits

$$\bar{x} - 2.58\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}$$

$$64 - 2.58(0.3) < \mu < 64 + 2.58(0.3)$$

63.226 and 64.774 are confidence limits.

2. A coin was tossed 500 times and heads turned up 120 times. Obtain the probable limits at 95% confidence level.

Soln: $n = 500 \quad P(\text{Heads}) = \frac{120}{500} = 0.24 \quad q = 0.76$

At 95% confidence level,

Probable limits : $p \pm 1.96\sqrt{\frac{pq}{n}}$

$$= 0.24 \pm 1.96\sqrt{\frac{0.24 \times 0.76}{500}}$$

$$= 0.24 \pm 0.0374$$

$$= 0.2774 \text{ and } 0.2026$$

$$= 27.74\% \text{ and } 20.26\%$$

Probable limits of obtaining heads is 20.26% to 27.74%

Nos.

3. A sample of 900 days was taken in a coastal town and it was found that on 100 days the weather was very hot. Obtain the probable limits of the percentage of very hot weather ($Z_\alpha = 2.58$)

Soln: $n = 900$

$p =$ Probability of hot weather $= \frac{100}{900} = \frac{1}{9}$ $\quad q = \frac{8}{9}$

Probable limits : $p \pm 2.58\sqrt{\frac{pq}{n}}$

$$= \frac{1}{9} \pm 2.58\sqrt{\frac{\frac{1}{9}\times\frac{8}{9}}{900}}$$

$$= 0.1381 \text{ and } 0.0841$$

$$= \cancel{13\%}\ 13.81\% \text{ and } 8.41\%$$

Probable limits of hot weather is 8.41% to 13.81%.

4. A survey was conducted in a locality of 2000 families by selecting a sample of size 800. It was revealed that 180 families were illiterates. Find the probable limits of illiterate families in population of 2000. (At 99% confidence level)

Soln: $N = 2000$ $\quad n = 800$

population $\quad$ sample

$p =$ Probability of illiterate families $= \frac{180}{800} = 0.225$ $\quad q = 0.775$

Probable limits : $p \pm 2.58\sqrt{\frac{pq}{n}}$

$$= 0.225 \pm 2.58\sqrt{\frac{0.225\times 0.775}{800}}$$

$$= 0.225 \pm 0.038$$

$$= 0.263 \text{ and } 0.187$$

$\therefore$ Probable limits of illiterate families in population 2000 is $(0.187)(2000)$ and $(0.263)(2000)$

Thus 374 to 526 families are illiterates.

Q. Nos.

5. In a locality of 18000 families a sample of 840 families was selected at random. Of these 840 families, 206 families were found to be having monthly income of ₹ 2500 or less. It was desired to estimate how many of the 18000 families have monthly income of ₹ 2500 or less. Within what limits would you place your estimate? (At 99% confidence level)

Soln: $N = 18000 \quad n = 840$

p = proportion of families having monthly income of ₹ 2500 or less.

$= \frac{206}{840} = 0.2452$

$q = 1 - p = 0.7548$

Probable limits : $p \pm 2.58\sqrt{\frac{pq}{n}}$

$= 0.2452 \pm 2.58\sqrt{\frac{0.2452 \times 0.7548}{840}}$

$= 0.2452 \pm 0.0383$

$= 0.2835$ and 0.2069

∴ Probable limits wrt population 18000 families is given by ;
$(0.2069)(18,000)$ and $(0.2835)(18000)$

Thus 3724 to 5103 families are likely to have monthly income of ₹ 2500 or less.

6. A random sample of 500 apples was taken from a large consignment and 65 were found to be bad. Estimate the proportion of bad apples as well as standard error of the estimate. Also find the percentage of bad apples in the consignment. (99% confidence level)

Soln: $n = 500 \quad p$ = Proportion of bad apples in consignment $= \frac{65}{500} = 0.13$

$q = 0.87 \quad$ SE proportion of bad apples $= \sqrt{\frac{pq}{n}} = 0.015$

Probable limits of bad apples in consignment $= p \pm 2.58\sqrt{\frac{pq}{n}}$

$= 0.13 \pm 2.58(0.015)$

$= 0.13 \pm 0.0388$

$= 0.1688$ and 0.0912

∴ Percentage of bad apples is 9.12% to 16.88%