UNIT - IV : Sampling Distribution

Population:

The set of well-defined objects is called population.

Sample:

A part of a population is known as a sample

Ex: We have considered population as AU, the sample is IT-B

Population size:

The no. of observations in the population is called population size and is denoted by 'N'

Sample size:

The no. of observations in the sample is called sample size and is denoted by 'n'

According to the sample size, there are 2 types of sampling

1. **Small sampling:**

   If the no. of observations in the sample is less than 30, then the sample is called small sample

2. Large sampling:

   If the no. of observations in the sample are ≥ 30, then the sample is called large sample

Notations:

$\bar{x}$ = Sample Mean

μ = Population Mean

s = Sample standard Deviation

σ = population "

σ^2 = population variance

p = sample proportion

P = population proportion

α = Level of significance

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$S = \sqrt{\frac{1}{n-1} \Sigma (x_i - \bar{x})^2} \quad (n < 30)$$

$$S = \sqrt{\frac{1}{n} \Sigma (x_i - \bar{x})^2} \quad (n \geq 30)$$

$$S^2 = \frac{1}{n-1} \Sigma (x_1 - \bar{x})^2 \quad (n < 30)$$

$$S^2 = \frac{1}{n} \Sigma (x_i - \bar{x})^2 \quad (n \geq 30)$$

- **To find the mean of the sampling distribution of the means:** let $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots \bar{x}_n$ be the means of the sample and let μ and σ^2 be the mean and variance of the population

Let $\bar{x}$ be the mean of sampling distribution of the means.

$$E(\bar{x}) = E\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right)$$

$$= \frac{1}{n}\left[E(x_1) + E(x_2) + \cdots E(x_n)\right]$$

$$= \frac{1}{n}\left[\mu + \mu + \cdots + \mu\right] \quad \left[\because E(x) = \mu\right]$$

$$= \frac{1}{\cancel{n}} \cancel{n}\mu$$

$$E(\bar{x}) = \mu$$

- To find the variance of the sampling distribution of the means.

Let σ^2 is the variance of population and s^2 be the variance of sample, then the variance of the sampling distribution of the means

$$V(\bar{x}) = V\left(\frac{x_1+x_2+\cdots+x_n}{n}\right)$$

$$= \frac{1}{n^2}\left[V(x_1)+V(x_2)+\cdots V(x_n)\right]$$

$$= \frac{1}{n^2}\left[\sigma^2+\sigma^2+\cdots+\sigma^2\right]$$

$$= \frac{1}{n^{\cancel{2}}}\cancel{n}\cdot\sigma^2$$

$$V(\bar{x}) = \frac{\sigma^2}{n}$$

$$S.D(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

Standard Error:

The S.D of the sampling distribution is known as standard Error.

$$S.E = \frac{\sigma}{\sqrt{n}}$$

Parameter:

Parameter is a statistical measure based on all the units of a population

Ex: The statistical constant of a population namely mean (μ), variance (σ^2) etc., are usually refered to as a parameter.

Statistic:

Statistic is a statistical measure based only on all the units selected in a sample

Ex: Statistical measures computed from the sample observations like sample mean ($\bar{x}$), sample

Estimate:

To find unknown population parameter a statement is made which is an Estimate (or)

A sample constant representing a population parameter is known as Estimate

Estimator:

The method / rule to determine an unknown population parameter is called Estimator

The Estimations are mainly two types:

1. point estimation
2. Interval estimation

Point estimation:

If the population parameter is being at exact point, that estimation is called point estimation

Ex: If the height of a student is measured 162 cm, then the estimate is called point estimation

Interval estimation:

If the population parameter is being an interval, it is known as Interval estimation

Ex: The height is given as (162 ± 3.5) cm, then the height is lies, then the estimate is called an Interval estimation

Maximum error of estimation for large samples:

The maximum error of estimate with $(1-\alpha)$ probability is given by:

Maximum Error $E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ (or) $Z_{\alpha/2} \frac{s}{\sqrt{n}}$

where, σ = population standard deviation

$Z_{\alpha/2}$ = Z table value at $\alpha/2$ level of significance

n = sample size

Maximum error of estimation for small sample:

The maximum error of estimate with $(1-\alpha)$ probability is given by:

Maximum error $E = t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$ ('σ' always unknown)

where,

$t_{\alpha/2}$ = t table value at $\alpha/2$ level of significance

Confidence interval for large sample:

Confidence interval for large sample with $1-\alpha$ probability is given by:

$C.I = (\bar{x} - E, \bar{x} + E)$,

where $\bar{x}$ = sample mean

$E = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ (or) $Z_{\alpha/2} \frac{s}{\sqrt{n}}$

confidence interval for small sample:

$C.I = (\bar{x} - E, \bar{x} + E)$

where $\bar{x}$ = sample mean

$E = t_{\alpha/2} \frac{s}{\sqrt{n}}$

Problem:

1) A random sample of size 81 was taken whose variance is 20.25 and mean is 32. Construct 98% confidence level and also find maximum error and CI!

Sol: Given $n = 81$

$s^2 = 20.25$, $s = \sqrt{20.25}$

$s = 4.5$

$\alpha = 100 - 98 = 2\%$

$\alpha = 0.02$

Maximum Error = $Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$

$Z_{\alpha/2}$ = Z - table value at $\alpha/2$ l.o.s

$\alpha = 0.02$

$\alpha/2 = 0.01$

$Z_{\alpha/2} = 0.5 - 0.01 = 0.49$

$Z_{tab} = 2.33$

$E = 2.33 \times \frac{4.5}{\sqrt{81}}$

$= \frac{2.33 \times 4.5}{9}$

$E = 1.165$

$C.I = (\bar{x} - E, \bar{x} + E)$

$= (32 - 1.165, 32 + 1.165)$

$C.I = (30.835, 33.165)$

2) If $\bar{x} = 11795$ and s.d 's' = 14054, n = 50. Calculate C.I at 5% level of significance

Sol: $\alpha = 5\% = 0.05$

Maximum Error = $Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$ $\left[Z_{\alpha/2} = 0.5 - \frac{0.05}{2} = 0.5 - 0.025 = 0.475 \right]$

$= 1.96 \times \frac{14054}{\sqrt{50}}$ $\quad Z_{tab} = 1.96$

$= 1.96 \times \frac{14054}{7.0710}$

$E = 3895.6074$

$C.I = (\bar{x} - E, \bar{x} + E)$

$= (11795 - 3895.6074, 11795 + 3895.6074)$

$C.I = (7899.393, 15690.607)$

Testing of Hypothesis:

A statistical hypothesis is a statement about the parameter one or more populations.

Testing of Hypothesis is a process for deciding whether to accept or reject the hypothesis

There are 2 types of Hypothesis:

1) Null Hypothesis
2) Alternative Hypothesis

Null Hypothesis:

Null Hypothesis is the hypothesis which is tested for possible rejection under the assumption that is true and it is denoted by H_0

Alternative Hypothesis:

Any Hypothesis which is contradict to the Null Hypothesis is called Alternative Hypothesis. and it is denoted by H_1

Ex: We want to test null hypothesis that the population has specified mean 50 (generally denoted with μ_0) then null hypothesis

$H_0 : \mu = 50$

(or)

$H_0 : \mu = \mu_0$

$\bar{x} = \mu$

$H_1 : \left.\begin{array}{l}\mu < 50 \\ \mu > 50\end{array}\right\}$ one tailed test - error α

$\mu \neq 50 \rightarrow$ two tailed test - error $\alpha/2$

Errors in Sampling:

There are 2 types of errors in Sampling. They are:

The two types of errors can be understood by the following table:

	Accept H_0	Reject H_0
H_0 is True	C.D	Type - I Error
H_0 is false	Type - II Error	C.D

Type-I Error:

The Hypothesis is true., but our test reject in this situation, we say that type I error has been made

Reject H_0 / when H_0 is true

Type-II Error:

The Hypothesis is false, but out test accept in this situation, we say that type 2 error has been made

Accept H_0 / when H_0 is false
(or)
when alternative hypothesis is true

Level of significance (α):

The probability of type-I Error is called level of significance

P(Type-I error) = α

P(Reject H_0 when H_0 is true) = α

P (Accept H_0 when H_0 is false) = β

Here α & β are the sizes of type I & type II errors.

Critical region (or) Rejection Region:

A region in the sample space 'S' which amount to rejection of

Critical value (or) Significant value:

The value of test statistic which separates the critical region and acceptance region is called critical value or significant value

The critical value depends on level of significance (α), alternative hypothesis (h_1)

One Tailed Test:

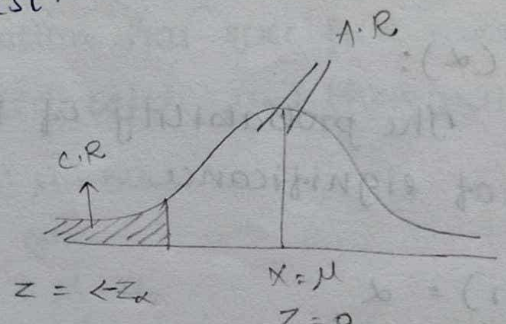
A Test of any statistical hypothesis, where alternative hypothesis $>$type or $<$type is called one Tailed Test

The one Tailed Test can be classified into 2 types. They are:

1. Left tailed test
2. Right tailed test

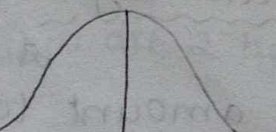
Left tailed test: ($\mu < \mu_0$)

If the critical region falls left hand side of the probability curve is known as Left tailed test.



Right tailed test: ($\mu > \mu_0$)

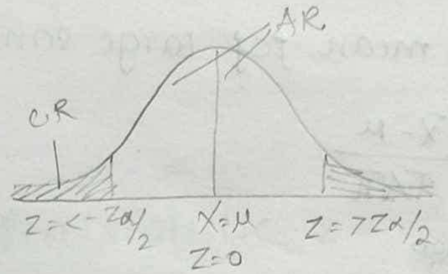
If the critical region falls right hand side of the probability curve is known as Right tailed test.



Two Tailed Test:

If the critical region falls under the probability curve is equally distributed on both sides of the probability curve

i.e,



Working rule for testing of hypothesis:

Step 1: Null Hypothesis H_0

To setup null hypothesis H_0 i.e, $\mu = \mu_0$ (or) $\bar{X} = \mu$

Step 2: Alternative Hypothesis

To setup alternative hypothesis H_1 so that we could decide whether we should use one tailed or two tailed test

i.e,

H_1: $\mu < \mu_0$ (LTT) left tailed test

$\mu > \mu_0$ (RTT) right "

$\mu \neq \mu_0$ (TTT) two "

Step 3: Level of significance (α)

Select the appropriate level of significance α depending on reliability of the estimates and permissable risk that is suitable to α is selected in advance i.e, tableted value at the given α and type of test

Step 4: Test Statistic

To compute the test statistic under the null hypothesis

Test statistic = $t - E(t)$

Here,

$t \rightarrow$ test statistic

$E(t) \rightarrow$ expected value of test statistic

$S.E(t) \rightarrow$ standard error of test statistic

Note:

Z-test for single mean of large sample

Test statistic $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

(or)

$$Z = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

where, $n \rightarrow$ sample size $(n \geq 30)$

Step 5: Conclusion

we compare calculated and tabulated values

1. If $|Z|_{cal} \leq Z_{tab}$ we accept H_0
2. If $|Z|_{cal} > Z_{tab}$ we reject H_0

Problems:

1. A sample of 400 items is taken from a population whose standard deviation is 10. The mean of the sample is 40. Test whether the sample has come from a population with mean 38. Also calculate 95% confidence interval for the population.

Sol: Given that

sample size $n = 400$

population S.D $\sigma = 10$

sample mean $\bar{x} = 40$

population mean $\mu = 38$

$\alpha = 100 - 95 = 5\%$

$\alpha = 0.05$

St 1: Null hypothesis is $H_0 : \mu = 38$ (or) $\bar{x} = \mu$

$40 = 38$

St 2: Alternative hypothesis $H_1 : \mu \neq 38$ (TTT)

The sample is not come from the population

St 3: Level of significance $\alpha = 5\% = 0.05$

$$\alpha/2 = 0.025$$

$$Z_{\alpha/2} = 0.5 - 0.025$$

$$Z_{\alpha/2} = 0.475$$

$$Z_{tab} = 1.96$$

St 4: Test statistic $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$$= \frac{40-38}{10/\sqrt{400}} = \frac{20(2)}{10} = 4$$

$$Z_{cal} = 4$$

St 5:

Conclusion:

$$Z_{cal}\,(4) > Z_{tab}\,(1.96)$$

$\therefore$ we reject H_0

$\therefore$ The sample is not come from the population

$$C.I = (\bar{x} - E, \bar{x} + E)$$

Maximum Error: $E = Z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

$$= \frac{1.96 \times 10}{\sqrt{400}} = \frac{19.6}{20} = 0.98$$

$$C.I = (\bar{x} - E, \bar{x} + E) = (40 - 0.98, 40 + 0.98)$$

$$C.I = (39.02, 40.98)$$

2. An ambulence service claims that it takes on the average less than 10 min to reach its destination in emergency call. A sample of 36 calls has a mean of 11 min and the variance of 16 min. Test the claim of 0.05 level of significance

population mean $\mu = 10$, $\alpha = 0.05$

St 1 : Null hypothesis $H_0 : \mu = 10$ min

i.e, The ambulence service to reach the destination is 10 min.

St 2 : Alternative hyp. $H_1 : \mu < 10$ (LTT)

i.e., The ambulence service to reach the destination in emergency calls less than 10 min.

St 3 : l.o.s 'α' = 0.05

$$Z_\alpha = 0.5 - 0.05 = 0.45$$

$$Z_{tab} = 1.645$$

St 4 : T.S $Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{11 - 10}{4/\sqrt{36}} = \frac{6}{4} = 1.5$

$$Z_{cal} = 1.5$$

St 5 : Conclusion

$$Z_{cal}(1.5) < Z_{tab}(1.645)$$

$\therefore$ we accept H_0

$\therefore$ The ambulence service claim is false.

3. A sample of 900 members has a mean of 3.4 cm and S.D 2.61 cm. Is this sample has been taken from a large population of mean 3.25 cm and S.D 2.61 cm. If the population is normal and its mean is unknown, find the 95% fiducial limits of the true mean.

Sol: Given that $n = 900$, $\bar{x} = 3.4$, $s = 2.61$, $\mu = 3.25$, $\sigma = 2.61$, $\alpha = 5\%$.

4. The mean life time of a sample of 100 light tubes produced by a company is found to be 1560 hrs with a population S.D of 90 hrs. Test the hypothesis for $\alpha = 0.05$ that the mean Lifetime of the tubes produced by the company is 1580 hrs.

Sol: Given that, $n = 100$, $\bar{x} = 1560$, $\sigma = 90$, $\mu = 1580$, $\alpha = 0.05$

St 1: Null hypothesis $H_0 : \mu = 1580$ hrs

i.e, The average length of the tubelights is equal to 1580 hrs.

St 2: Alternative hypothesis $H_1 : \mu \neq 1580$ hrs (TTT)

i.e, The average length of the tubelights is not equal to 1580 hrs.

St 3: L.O.S $\alpha = 0.05$

$\alpha/2 = 0.025$

$Z_{\alpha/2} = 0.5 - 0.025 = 0.475$

$Z_{tab} = 1.96$

St 4: T.S $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{1560 - 1580}{90/\sqrt{100}} = \dfrac{\cancel{10} \times (-20)}{\cancel{90}}$

$Z_{cal} = -2.2$

$|Z|_{cal} = 2.2$

St 5: Conclusion

$|Z|_{cal} (2.2) > Z_{tab} (1.96)$

$\therefore$ We accept H_0

(claim is false)

$\therefore$ The mean lifetime of the tubes is not equal to 1580 hrs.

5. A sample of 64 students has a mean weight of 70 kg. Can this be regardes as a sample from a population with mean weight 56 kg and S.D 25 kg.

Test for equality of 2 means of large samples or Test for significance for difference of means of two large samples:

Procedure:

Let $\bar{x}_1$ & $\bar{x}_2$ be the sample means of two independent large random samples of size n_1 & n_2 drawn from 2 populations having means μ_1 & μ_2 and s.d's σ_1 & σ_2. To test whether the two population means are equal.

St 1: Null Hypothesis $H_0 : \mu_1 = \mu_2$

i.e, population means are equal

St 2: Alternative hypothesis $H_1 : \mu_1 < \mu_2$ (LTT)

$\mu_1 > \mu_2$ (RTT)

$\mu_1 \neq \mu_2$ (TTT)

St 3: Level of significance:

To select appropriate l.o.s (α), calculate tabulated value in advance

Test statistics:

St 4: To compute test statistics under the null hypothesis:

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

Here, $\bar{x}_1$ = first sample mean

$\bar{x}_2$ = second " "

σ_1 = first population s.d

σ_2 = second " "

n_1 = first sample size

n_2 = second " "

St 5: Conclusion:

we compare calculated and tabulated values

1. If $|Z|_{cal} \leq Z_{tab}$

we accept H_0

2. If $|Z|_{cal} > Z_{tab}$

we reject H_0

Problems:

1. The means of 2 large samples of 1000 and 2000 members are 67.5 inches and 68 inches respectively. Can this sample be regarded has drawn from the same population of s.d 2.5 inches.

Sol: Given that,

$n_1 = 1000$, $n_2 = 2000$, $\sigma_1 = 2.5$ inches $= \sigma_2$

$\bar{x}_1 = 67.5$, $\bar{x}_2 = 68$

St 1: Null hypothesis $H_0 : \mu_1 = \mu_2$

i.e., The samples belong to the same population

St 2: Alternative hypothesis $H_1 : \mu_1 \neq \mu_2$ (TTT)

i.e., The samples doesn't belong to the same population

St 3: L.O.S $\alpha = 0.03$

$\therefore |Z|_{tab} = 1.96$

St 4: T.S $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{67.5 - 68}{2.5\sqrt{\dfrac{1}{1000} + \dfrac{1}{2000}}} = -5.165$

$\therefore |Z|_{cal} = 5.165$

St 5: Conclusion:

$|Z|_{cal} (5.165) > |Z|_{tab} (1.96)$

$\therefore$ We reject H_0 (or

We accept H_1

Therefore, the samples do not belong to the same population

2. Two types of new cars produced in USA are tested for petrol mileage. One sample is consisting of 42 cars averaged 15 Km/Lit while the other sample consisting of 80 cars averaged 11.5 Kmpl with population variances as $\sigma_1^2 = 2$, $\sigma_2^2 = 1.5$. Test whether there is any significant difference in the petrol consumption of these 2 types of cars. Use 1% level of significance

Sol: Given that,

$n_1 = 42$, $n_2 = 80$, $\bar{x}_1 = 15$, $\bar{x}_2 = 11.5$, $\sigma_1^2 = 2$,

$\sigma_2^2 = 1.5$, $\alpha = 0.01$

St 1: Null Hypothesis $H_0 : \mu_1 = \mu_2$

i.e., There is no significant difference b/w petrol consumption of these 2 types of cars.

St 2: Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$ (TTT)

i.e, There is significant difference between petrol consumption of these 2 types of cars

St 3: l.o.s $\alpha = 0.01$

$$\alpha/2 = \frac{0.01}{2} = 0.005$$

$$Z_{\alpha/2} = 0.495$$

$$Z_{tab} = 2.575 \text{ or } 2.58$$

St 4: TS: $Z_{cal} = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} = \dfrac{15 - 11.5}{\sqrt{\dfrac{2}{42} + \dfrac{1.5}{80}}}$

$$|Z|_{cal} = 13.586$$

Conclusion:

St 5: $|Z|_{cal}(13.586) > |Z|_{tab}(2.58)$

$\therefore$ We reject H_0 (or)

We accept H_1

i.e., There is a significant difference between

3. A simple sample of the height of 6400 englich men has a mean of 67.85 inches and a s.d of 2.56 inches while a simple sample of height of 1600 australians has a mean of 68.55 inches and s.d of 2.52 inches Do the data indicates the australians are on the average taller than the english men. Use 1% level

Sol: Given that,

$n_1 = 6400,\ n_2 = 1600,\ \bar{x}_1 = 67.85,\ \bar{x}_2 = 68.55,$

$s_1 = 2.56,\ s_2 = 2.52,\ \alpha = 0.01$

St 1: Null hypothesis $H_0 : \mu_1 = \mu_2$

i.e., Both averages of english men and australian men of the avg's are equal

St 2: Alternative hypothesis $H_1 : \mu_1 < \mu_2$ (LTT)

i.e., (English men) australian men are taller than the english men

St 3: l.o.s $\alpha = 0.01$

$Z_\alpha = 0.5 - 0.01$

$Z_\alpha = 0.49$

$Z_{tab} = 2.33$

St 4: T.S : $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \dfrac{67.85 - 68.55}{\sqrt{\dfrac{(2.56)^2}{6400} + \dfrac{(2.52)^2}{1600}}}$

$Z = -9.91$

$|Z|_{cal} = 9.91$

St 5: conclusion:

$|Z|_{cal}\ (9.9) > Z_{tab}\ (2.33)$

$\therefore$ We accept H_1

(or)

We reject H_0

$\therefore$ Australian men are taller than the english men.

4. A company claims that its bulbs are superior to those of its main competetor. If a study showed that, a sample of 40 of its bulbs have a mean life time of 647 hours of continuous use with a s.d of 27 hours. while a sample of 40 bulbs made by its main competetor had a mean life time of 638 hours of continuous use with a s.d of 31 hours. Test the significance b/w the difference of 2 means at 5% level.

Sol: Given:

$n_1 = 40$, $n_2 = 40$, $\bar{x}_1 = 647$, $\bar{x}_2 = 638$,

$s_1 = 27$, $s_2 = 31$, $\alpha = 0.05$

Step1: Null hypothesis $H_0 : \mu_1 = \mu_2$

i.e, There is no significant difference between 2 types of bulbs

Step2: Alternative hypothesis: $H_1 : \mu_1 > \mu_2$ (RTT)

i.e., Type 1 is superior to Type 2 regarding lifetime

Step3: l.o.s $\alpha = 0.05$

$Z_\alpha = 0.5 - 0.05 = 0.45$

$Z_{tab} = 1.65$

Step4: T.S: $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \dfrac{647 - 638}{\sqrt{\dfrac{729}{40} + \dfrac{961}{40}}} = \dfrac{9}{6.5}$

$\therefore |Z|_{cal} = 1.38$

Step 5: Conclusion:

$|Z|_{cal} (1.38) < |Z|_{tab} (1.65)$

$\therefore$ We reject H_1 (or) accept H_0

Test of Significance for single proportion of large sample:

If 'x' is the number that an event occure among 'n' trials, the proportion of the time that the event is x/n, it is the sample proportion. The sample proportion is denoted by p, since $P+Q=1$, $Q=1-P$

Step 1: Null Hypothesis H_0 : $P = P_0$ (or) $p = P$

i.e, sample proportion is equal to population proportion

Step 2: Alternative Hypothesis H_1 : $P < P_0$ (LTT)

$P > P_0$ (RTT)

$P \neq P_0$ (TTT)

Step 3: level of significane:

To select appropriate level of significance 'α', and find out tabulated value in advance

Step 4: Test statistic:

To compute test statistic under the null hypothesis H_0:

$$\therefore \text{Test statistic } Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$$

Here, $p \rightarrow$ sample proportion

$P \rightarrow$ population "

$Q \rightarrow 1-P$

$n \rightarrow$ sample size ($n \geq 30$)

Step 5: Conclusion:

we compare calculated and tabulated values.

i) if $|Z|_{cal} \leq Z_{tab}$, we accept H_0

ii) if $|Z|_{cal} > Z_{tab}$, we reject H_0

(or)

We accept H_1

- Maximum Error : $E = Z_{\alpha/2}\sqrt{\frac{PQ}{n}}$

Problems:

1. Experience had shown that 20% of a manufactured product is of the top quality. In one day's production of 400 articles only 50 are of top quality. Test the hypothesis at 0.05 level. Also find C.I

Sol: Given that population proportion $P = 0.2$

$Q = 1 - P = 1 - 0.2 = 0.8$

$\therefore Q = 0.8$

Top quality of articles $x = 50$

$n = 400$

Sample proportion of top quality of articles $p = \frac{x}{n}$

$p = \frac{50}{400} = 0.125$

$\alpha = 0.05$

St 1: Null hypothesis $H_0 : P = 0.2$

i.e, 20% of the articles are top quality

St 2: Alternative hypothesis $H_1 : P \neq 0.2$ (TTT)

i.e, 20% of the articles are not top quality

St 3: l.o.s $\alpha = 0.05$

$\alpha/2 = 0.025$

$Z_{\alpha/2} = 0.5 - 0.025$

$Z_{\alpha/2} = 0.475$

$Z_{tab} = 1.96$

St 4: T.S $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.125 - 0.2}{\sqrt{\frac{0.2 \times 0.8}{400}}} = -3.75$

$|Z|_{cal} = 3.75$

St 5: Conclusion:

$|Z|_{cal}\ (3.75) > Z_{tab}\ (1.96)$

$\therefore$ We reject H_0 (or)

We accept H_1

$$C.I = \left(p - Z_{\alpha/2}\sqrt{\frac{PQ}{n}},\ p + Z_{\alpha/2}\sqrt{\frac{PQ}{n}}\right)$$

$$= \left(0.125 - 1.96\sqrt{\frac{0.2\times0.8}{400}},\ 0.125 + 1.96\sqrt{\frac{0.2\times0.8}{400}}\right)$$

$$= (0.0858, 0.1642)$$

2. A manufacturer claims that atleast 95% of the equipment which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at 5% level of significance

Sol: Given that population proportion $P = 0.95$

$Q = 0.05$

$n = 200$

$\alpha = 0.05$

conforming pieces $x = 200 - 18 = 182$

Sample proportion of conforming pieces $p = \frac{x}{n}$

$= \frac{182}{200} = 0.91$

St 1: Null hypothesis $H_0 : P = 0.95$

i.e, 95% of the equipment are conforming to specifications.

St 2: Alternative hypothesis $H_1 : P < 0.95$ (LTT)

i.e, Atleast 95% of the equipment is not conforming to specifications.

St 3: l.o.s $\alpha = 0.05$

$Z_\alpha = 0.5 - 0.05$

$Z_\alpha = 0.45$

$Z_{tab} = 1.645$

St 4: T.S $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91 - 0.95}{\sqrt{\frac{0.95\times0.05}{200}}}$ $\therefore Z_{cal} = -2.59$

$\therefore |Z|_{cal} = 2.59$

St 5: Conclusion:

$|Z|_{cal}\ (2.59) > Z_{tab}\ (1.645)$

3. 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 the city are smokers.

Sol: Given that, $n = 600$, $P = 0.5$, $Q = 0.5$

p = Sample proportion of smokers $= \frac{325}{600}$ $\left(\frac{x}{n}\right)$

$= 0.541$

St 1: Null hypothesis $H_0 : P = 0.5$

i.e, Both smokers and non-smokers are equal in the city

St 2: Alternative hypothesis $H_1 : P > 0.5$ (RTT)

i.e, The majority of men in the city are smokers

St 3: L.o.s 'α' $= 0.05$

$Z_\alpha = 0.5 - 0.05 = 0.45$

$Z_{tab} = 1.645$

St 4: T.S : $Z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.541 - 0.5}{\sqrt{\dfrac{0.5 \times 0.5}{600}}} = 2.008$

$\therefore |Z|_{cal} = 2.008$

St 5: Conclusion:

$|Z|_{cal}\ (2.008) > Z_{tab}\ (1.645)$

$\therefore$ We accept H_1

Hence majority of men in the city are smokers.

4. In a sample of 1000 people in Karnataka 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% L.o.s.

Sol: Given that, $n = 1000$, $P = 0.5$, $Q = 0.5$, $p = 0.54$

$\alpha = 0.01$

Test for equality of 2 proportions (or) Test of significance of difference between 2 sample proportions of large proportions:

Let p_1 and p_2 be the sample proportions in 2 large random samples of sizes n_1 and n_2 drawn from two populations having proportions P_1 and P_2. To test whether the two samples have been drawn from the same population:

Step 1: Null Hypothesis :- $H_0 : P_1 = P_2$

i.e, The two population proportions are equal

Step 2: Alternative Hypothesis H_1: $P_1 < P_2$ (LTT)
$P_1 > P_2$ (RTT)
$P_1 \neq P_2$ (TTT)

Step 3: Level of significance:
To select appropriate L.O.S 'α' and find out tabulated value in advance

Step 4: Test statistic:
To compute test statistic under the null hypothesis H_0:

$$\text{Test statistic} : Z = \frac{p_1 - p_2 - (P_1 - P_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

Here, $P_1 - P_2 = 0$ & P_1, P_2 are known

*Note: 1. If P_1 & P_2 are not known & p_1 & p_2 are known, then

$$T.S : Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Here, $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ (or) $\frac{x_1 + x_2}{n_1 + n_2}$

Step 5: Conclusion:
we compare calculated & tabulated values

1. If $|Z|_{cal} < Z_{tab}$ ⇒ We accept H_0

Maximum Error: $E = Z_{\alpha/2}\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ ($\because$ P_1, P_2 are unknown)

$P_1 P_2$ are known

$E = Z_{\alpha/2}\sqrt{\frac{P_1 Q_1}{n_1}+\frac{P_2 Q_2}{n_2}}$ ($\because$ P_1, P_2 are known)

Confidence Interval:

$$C.I = \left\{\left[(P_1-P_2) - Z_{\alpha/2}\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right], \left[(P_1-P_2) + Z_{\alpha/2}\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right]\right\}$$

Problems:

1 Random samples of 400 men and 600 women were asked whether they would like to have a flyover near their residence, 200 men and 325 women were in favour of the proposal. Test the hypothesis that proportions of men and women in favour of the proporsal are same at 5% l.o.s.

Sol: Given that, $n_1 = 400$ n_2 600

$p_1 = 0.5$ $p_2 = 0.541$

$\alpha = 0.05$

St1: Null hypothesis $H_0 : P_1 = P_2$

i.e., Men and women in favour of the proporsal are same

St2: Alternative hypothesis $H_1 : P_1 \neq P_2$ (TTT)

i.e, Men and women in favour of the proporsal are not same

St3: l.o.s 'α' $= 0.05$

$\alpha/2 = 0.05/2 = 0.025$

$Z_{\alpha/2} = 0.5 - 0.025 = 0.475$

$Z_{tab} = 1.96$

St4: TS: $Z = \frac{p_1 - p_2 \text{ (small)}}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

$$P = \frac{400 \times 0.5 + 600 \times 0.541}{1000} = 0.525$$

$Q = 0.4754$

$$Z = \frac{0.5 - 0.541}{\sqrt{(0.2493)\left(\frac{1}{400} + \frac{1}{600}\right)}} = 1.272$$

$|Z|_{cal} = 1.272$

St 5: Conclusion:

$Z_{cal}\ (1.272) < Z_{tab}\ (1.96)$

$\therefore$ we reject H_0

i.e., Men and women in favour of the proporsal are not same

2. A manufacture of electronic equipment samples of 2 completing brands of transistors to an accelerated performances test. If 45 of 180 transistors of the 1st kind and 34 of 120 transistors of the second kind fail the test what can we conclude that level of significance $\alpha = 0.05$ about the difference b/w the corresponding sample proportions.

Sol: Given that, $n_1 = 180$ $\quad n_2 = 120$

$P_1 = \frac{45}{180} = 0.25 \quad P_2 = \frac{34}{120} = 0.28$

$\alpha = 0.05$

St 1: Null hypothesis $H_0 : P_1 = P_2$

i.e, There is no significant difference between the accelerated performance of 2 kinds of transistors.

St 2: Alternative hypothesis $H_1 : P_1 \neq P_2$ (TTT)

i.e, There is a significant difference between the accelerated performance of 2 kinds of transistors.

St 3: LOS 'α' $= 0.05$

$\alpha/2 = 0.05/2 = 0.025$

St 4: TS : $Z = \dfrac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

$$P = \frac{n_1P_1 + n_2P_2}{n_1+n_2} \text{ (or) } \frac{x_1+x_2}{n_1+n_2} = \frac{45+34}{180+120} = \frac{79}{300} = 0.263$$

$$Q = 1 - P = 1 - 0.263 = 0.737$$

$$Z = \frac{0.25 - 0.25}{\sqrt{0.263 \times 0.737\left(\frac{1}{180}+\frac{1}{120}\right)}} = -0.647$$

$$|Z|_{cal} = 0.647$$

St 5: Conclusion:

$$|Z|_{cal}\ (0.641) < Z_{tab}\ (1.96)$$

$\therefore$ we accept H_0

i.e., There is no significant difference between the accelerated performance of 2 kinds of transistors.

*** 3. On the basis of their total scores, 200 candidates of a civil service examination are divided into 2 groups. The upper 30% and the remaining 70% consider the first question of the examination among the first group 40 had the correct answer, whereas among the second group 80 had the correct answer. On the basis of the results, can one conclude that the first question is not good at discriminating ability of the type being examined here. Also find CI.

Sol: Given that,

$n_1 = 60$ $\quad\quad$ $n_2 = 140$

$x_1 = 40$ $\quad\quad$ $x_2 = 80$

$P_1 = \dfrac{x_1}{n_1} = 0.66$ $\quad\quad$ $P_2 = \dfrac{x_2}{n_2} = 0.57$

$Q = 1 - P = 1 - 0.6 = 0.4$

$\alpha = 0.05$

St 1: Null hypothesis $H_0: P_1 = P_2$

i.e., The first question is not good in discriminating ability of the students of both groups

St 2: Alternative hypothesis $H_1: P_1 \neq P_2$

i.e., The first question is good in discriminating ability of the students of both groups.

St 3: L.O.S 'α' $= 0.05$

$\alpha/2 = 0.05/2 = 0.025$

$Z_{\alpha/2} = 0.5 - 0.025 = 0.475$

$Z_{tab} = 1.96$

St 4: T.S : $Z = \dfrac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \dfrac{0.67 - 0.57}{\sqrt{(0.6)(0.4)\left(\frac{1}{60} + \frac{1}{140}\right)}} = 1.322$

$|Z|_{cal} = 1.32$

St 5: Conclusion:

$|Z|_{cal}(1.32) < Z_{tab}(1.96)$

$\therefore$ We accept H_0

i.e, The first question is not good in discriminating ability of the students of both groups.

Confidence Interval (C.I):

$$C.I = \left((P_1 - P_2) \pm Z_{\alpha/2}\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$$

$$= \left[(0.66 - 0.57) - 1.96\sqrt{0.6 \times 0.4\left(\frac{1}{60} + \frac{1}{140}\right)},\right.$$

$$\left.(0.66 - 0.57) + 1.96\sqrt{0.6 \times 0.4\left(\frac{1}{60} + \frac{1}{140}\right)}\right]$$

Note: If sample proportion P_1 & P_2 are unknown then instead of P_1 & P_2 take $p_1 - p_2$

$$T.S = \frac{p_1 - p_2}{\sqrt{\frac{p_1 Q_1}{n_1} - \frac{p_2 Q_2}{n_2}}}$$

4. In 2 large populations there are 30% and 25% respectively of fair haired people. In this difference likely to be hidden in samples of 1200 and 900 respectively from the 2 populations.

Sol: Given that $n_1 = 1200$ $\quad n_2 = 900$

$P_1 = 0.3$ $\quad P_2 = 0.25$

$Q_1 = 0.7$ $\quad Q_2 = 0.75$

$\alpha = 0.05$

St 1: Null hypothesis $H_0 : P_1 = P_2$

i.e, The difference in population proportion is likely that the real difference will be hidden.

St 2: Alternative hypothesis $H_1 : P_1 \neq P_2$

i.e, The difference in population proportion is unlikely that the real difference will be hidden

St 3: LOS $\alpha = 0.05$

$\alpha/2 = 0.025$

$Z_{\alpha/2} = 0.5 - 0.025 = 0.475$

$Z_{tab} = 1.96$

St 4: T.S : $Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}} = \frac{0.3 - 0.25}{\sqrt{\frac{0.3 \times 0.7}{1200} + \frac{0.25 \times 0.75}{90}}}$

$|Z|_{cal} = 2.631$

St 5: Conclusion:

$Z_{cal}\ (2.631) > Z_{tab}\ (1.96)$