Unit - VI

# TEST OF HYPOTHESIS

**Population:-** The aggregate of all the units of a the characteristic understudy is called as "population".

**Ex:-** (i) Number of universities in the state.

(ii) Number of Nationalised banks in India.

**Sample:-** A finite part of the population is called as "sample".

**Ex:-** A sample of 5 students of a class.

**Statistical Hypothesis:-** Any statement about the behaviour of the Random variables is called as "statistical Hypothesis" or "Hypothesis".

There are two types of hypothesis.

(i) **Null hypothesis:-** The null hypothesis is a hypothesis which we assumed that the difference between estimator and parameter is zero.

(or)

The hypothesis of no difference is called as "Null hypothesis". It is denoted by $H_0$.

**Ex:-** $H_0: \mu = \mu_0$

(ii) Alternative hypothesis:- Any hypothesis which is complementary to null hypothesis is called as "Alternative hypothesis".

(or)

Any hypothesis which contradicts the null hypothesis is called an "Alternative hypothesis".

It is denoted by $H_1$.

Ex: $H_1 : \mu \neq \mu_0$ (or) $H_1 : \mu > \mu_0$ (or) $H_1 : \mu < \mu_0$

Errors of sampling:- In sampling, we have two types of errors. They are

(i) Type-I error:- Reject Ho when it is true

If the null hypothesis Ho is true but it is rejected by test procedure, then the error made is called Type-I error (or) $\alpha$- error.

(ii) Type-II error:- Accept Ho, when it is wrong.

If the null hypothesis Ho is false but is accepted by test procedure, then the error made is called Type-II error (or) $\beta$-error

Scanned with CamScanner

**Level of significance:-** Probability of type-I error is called as "Level of significance" (or) Probability of rejecting $H_0$, when it is true is called as "Level of significance". It is denoted by $\alpha$.

$$\alpha = P(\text{type-I error}) = P(\text{rejecting } H_0 \text{ when it is true.})$$

**One tailed and two-tailed tests:-**

**One tailed test:-** suppose we want to test the null hypothesis $H_0: \mu = \mu_0$ against the alternative hypothesis $H_1: \mu > \mu_0$ (right-tailed) (or) $H_1: \mu < \mu_0$. (left tailed) is called as "one-tailed test."

**Two-tailed test:-** suppose we want to test the null hypothesis $H_0: \mu = \mu_0$ against the alternative hypothesis $H_1: \mu \neq \mu_0$. is called as "two tailed test".

Table: Crictical values of $Z$

| Level of significance $\alpha$ | 1% | 5% | 10% |
|---|---|---|---|
| Critical values for two-tailed test | $\lvert Z_\alpha \rvert = 2.58$ | $\lvert Z_\alpha \rvert = 1.96$ | $\lvert Z_\alpha \rvert = 1.645$ |
| Critical values for Right-tailed test. | $\lvert Z_\alpha \rvert = 2.33$ | $Z_\alpha = 1.645$ | $Z_\alpha = 1.28$ |
| Critical values for Left-tailed test | $Z_\alpha = -2.33$ | $Z_\alpha = -1.645$ | $Z_\alpha = -1.28$. |

## Procedure for Testing of Hypothesis:-

Various steps involved in testing of hypothesis are given below.

Step 1:- Null Hypothesis:- Define or set-up a Null Hypothesis $H_0$ (taking in to consideration the nature of the problem and data involved.)

Step 2:- Alternative Hypothesis:- Set up the Alternative Hypothesis $H_1$.

Step 3:- Level of significance:- Select the appropriate Level of significance ($\alpha$).

Step 4:- Test statistic:- Compute the test statistic $Z = \dfrac{t - E(t)}{S.E \text{ of } t}$ under the null hypothesis.

Here $t$ is a sample statistic and S.E is the standard error of $t$.

Step 5:- Conclusion:- we compare the computed value of the test statistic $Z$ with the critical value of the test statistic $Z$ with the Critical value of the given Level of significance. Value of $Z_\alpha$ at the given Level of Significance.

If $|Z| \leq Z_\alpha$, then we accept $H_0$ otherwise we reject it.

Scanned with CamScanner

# TEST OF SIGNIFICANCE FOR SINGLE MEAN :-

For testing the significance of single mean, we write the null hypothesis as

$H_0: \mu = \mu_0$ (or) $\bar{x} = \mu$

To test the above $H_0$, we write the Z-test statistic as

$$Z = \frac{|\bar{x} - \mu|}{\sigma/\sqrt{n}}.$$

If $Z_{cal} \leq Z_{tab}$, then we accept $H_0$ otherwise we reject it.

**Large samples :-** If the sample size $n \geq 30$, then we consider such samples as large samples.

## Problems :

P) A sample of 64 students have a mean weight of 70 kgs. Can this be regarded as a sample from a population with mean weight 56 kgs and standard deviation 25 kgs.

Sol$^n$ :- Given

$n$ = sample size = 64

$\bar{x}$ = sample mean = 70

$\mu$ = population mean = 56.

$\sigma$ = population standard deviation = 25

Null hypothesis :- $H_0: \bar{x} = \mu$.

Alternative Hypothesis :- $H_1: \bar{x} \neq \mu$.

Level of significance :- $\alpha = 0.05$ (assume)

Then the test statistic is

$$z = \frac{|\bar{x} - \mu|}{\sigma/\sqrt{n}} = \frac{|70-56|}{\frac{25}{\sqrt{64}}} = 4.48$$

$\therefore$ ~~Zcal~~ $Z_{tab}$ Value at 5% L.O.S = 1.96

$\therefore Z_{cal} > Z_{tab}$ Value.

Hence we reject our $H_0$.

P) An oceanographer wants to check whether the depth of the ocean in a certain region is 57.4 fathoms, as had previously been recorded. What can he conclude at the 0.05 level of significance, if readings taken at 40 random locations in the given region yielded a mean of 59.1 fathoms with a standard deviation of 5.2 fathoms.

Solⁿ:- Given $n$ = Sample size = 40

$\bar{x} = 59.1$, $s = 5.2$, $\mu = 57.4$

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

$H_1: \bar{x} \neq \mu$.

$\alpha = 0.05$

$$z = \frac{|\bar{x} - \mu|}{s/\sqrt{n}} = \frac{|59.1 - 57.4|}{5.2/\sqrt{40}} = 2.067.$$

$Z_{tab}$ value at 5% L.O.S = ~~2.067~~. 1.96.

$Z_{cal} > Z_{tab}$

Hence we reject $H_0$

Hence we conclude that $\bar{x} = \mu$.

P) 10) In a random sample of 60 workers, the average time taken by them to get to work is 33.8 minutes with a standard deviation of 6.1 minutes. Can we reject the null hypothesis $\mu = 32.6$ minutes in favour of alternative null hypothesis $\mu > 32.6$ at $\alpha = 0.05$ level of significance.

Soln.:- Given,

$n = 60$, $\bar{x} = 33.8$, $s = 6.1$

$\mu = 32.6$

1. Null hypothesis :- $H_0: \mu = 32.6$

2. Alternative hypothesis :- $H_1: \mu > 32.6$ (one-tailed test)

3. Level of Significance :- $\alpha = 0.05$

4. The test statistic is,

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

$$= \frac{|33.8 - 32.6|}{6.1/\sqrt{60}} = 1.5238.$$

$Z_{tab}$ Value at 5% L.O.S = 1.645

$\therefore Z_{cal} < Z_{tab}$

Hence we accept $H_0$.

Hence we conclude that $\mu = 32.6$

P) 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.

Sol$^n$:- Given, $n = 400$, $\sigma = 10$, $\bar{x} = 40$, $\mu = 38$.

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

2. Alternative hypothesis :- $H_1 : \bar{x} \neq \mu$. (or) $\mu \neq 38$.

3. Level of Significance $\alpha = 0.05$ (5% L.O.S)

4. The test statistic, $z = \dfrac{|\bar{x} - \mu|}{\sigma/\sqrt{n}}$

$$= \frac{|40 - 38|}{10/\sqrt{400}} = 4.$$

$z_{tab}$ value at 5% L.O.S = 1.96.

$z_{cal} > z_{tab}$, Hence we reject $H_0$.

P) An ambulance service claims that it takes on the average less than 10 minutes to reach its destination in emergency calls. A sample of 36 calls has a mean of 11 minutes and the variance of 16 minutes. Test the significance at 0.05 level.

Sol$^n$:- Given $\mu = 10$ minutes, $n = 36$, $\bar{x} = 11$, $s^2 = 16$.
$s = 4$.

1. Null hypothesis :- $H_0 : \bar{x} = \mu$.

2. Alternative hypothesis :- $H_0 : \bar{x} \neq \mu$

3. Level of significance $\alpha = 0.05$

4. The test statistic is,

$$z = \frac{|\bar{x} - \mu|}{\sigma/\sqrt{n}}$$

$$= \frac{|11 - 10|}{4/\sqrt{36}} = 1.5$$

Ztab value at 0.05 L.O.S (5% L.O.S) = 1.96

Zcal < Ztab value,

Hence we accept H0

P) It is claimed that a random sample of 49 tyres has a mean life of 15200 Kms. This sample was drawn from a population whose mean is 15150 Kms and a standard deviation of 1200 Km. Test the significance at 0.05 level.

Sol:- A) Given,

$n = 49$, $\bar{x} = 15200$, $\mu = 15150$

$\sigma = 1200$.

1. Null hypothesis :- $H_0 : \bar{x} = \mu$ (or) $\mu = 15150$

2. Alternative hypothesis :- $H_1 : \bar{x} \neq \mu$ (or) $\mu \neq 15150$.

3. Level of significance : $\alpha = 0.05$

4. The test statistic is

$$z = \frac{|\bar{x} - \mu|}{\sigma/\sqrt{n}}$$

$$= \frac{|15200 - 15150|}{1200/\sqrt{49}} = 0.2917.$$

$\therefore$ Ztab at 5% L.O.S = 1.96.

Zcal < Ztab, Hence we accept H0

H.W

P) The mean life time of a sample of 100 light bulbs produced by a company is found to be 1560 hours with a population S.D of 90 hours. Test the hypothesis for $\alpha = 0.05$ that the mean life time of the bulbs produced by the company is 1580 hours?

Ans: $Z_{cal} = 2.22$, $Z_{tab} = 1.96$
we reject $H_0$.

Note:- The values $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$ are called 95% fiducial limits or confidence limits for the mean of the population corresponding to the given sample.

similarly, $\bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$ are called 99% confidence limits and $\bar{x} \pm 2.33 \frac{\sigma}{\sqrt{n}}$ are called 98% confidence limits.

P) The mean and standard deviation of a population (population / sample) are 11795 and 14054 respectively. If n = 50, find 95% confidence interval for the mean.

Sol$^n$:- Given, $n = 50$, $\bar{x} = 11795$ and $\sigma = 14054$.

95% Confidence interval is $\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$

i.e, $\left(11795 - \frac{(1.96)(14054)}{\sqrt{50}}, 11795 + \frac{1.96(14054)}{\sqrt{50}}\right)$

i.e, $(11795 - 3895.6, 11795 + 3895.6)$

or $(7899.4, 15690.6)$

P) A sample of 400 items is taken from a population whose standard deviation is 10. The mean of the ~~Population~~ ~~Sample~~ Sample is ~~40~~ 38. Calculate 95% confidence interval for the population.

Sol$^n$:- Given, $n = 400$, $\sigma = 10$, ~~$\bar{x} = 40$~~ $\bar{x} = 38$.

95% Confidence interval is $\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$

# TEST FOR EQUALITY OF TWO MEANS

## (TEST OF SIGNIFICANCE FOR DIFFERENCE OF MEANS OF TWO LARGE SAMPLES)

Let $\bar{x}_1$ and $\bar{x}_2$ be the sample means of two independent large random samples sizes $n_1$ and $n_2$ drawn from two populations having means $\mu_1$ and $\mu_2$ and standard deviations $\sigma_1$ and $\sigma_2$. Then we write $H_0$ as $H_0: \bar{x}_1 = \bar{x}_2$

To test whether the two population means are equal, we use the following test statistic

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

If $\sigma_1^2$ and $\sigma_2^2$ are unknown, then $\sigma_1^2$ and $\sigma_2^2$ can be replaced by sample variances $s_1^2$ and $s_2^2$.

Then

$$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

If $z_{cal} \le z_{tab}$, then we accept $H_0$ otherwise we reject it at the given L.O.S.

## Problems

1) The means of two large samples of sizes 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the same population of S.D 2.5 inches

Soln:-

Given

$n_1 = 1000$, $n_2 = 2000$

$\bar{x}_1 = 67.5$, $\bar{x}_2 = 68.0$

$\sigma = 2.5$.

Null hypothesis: $H_0 : \bar{x}_1 = \bar{x}_2$

Alternative hypothesis $H_1 : \bar{x}_1 \neq \bar{x}_2$

Level of significance: 0.05 (5%)

The test statistic $z = \dfrac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$

$$= \frac{|67.5 - 68|}{\sqrt{(2.5)^2\left(\frac{1}{1000} + \frac{1}{2000}\right)}} = 5.16$$

$z_{cal} = 5.16$

$z_{tab}$ Value at 5% L.O.S $= \cancel{5.16}\ 1.96$

$z_{cal} > \cancel{z_{tab}}$ $z_{tab}$

Hence we ~~accept~~ reject $H_0$.

4) The mean yield of wheat from district A was 210 pounds with S.D 10 pounds per acre from a sample of 100 plots. In another district the mean yield was 220 pounds with S.D 12 pounds from a sample of 150 plots. Assuming that the S.D of yield in the entire state was 11 pounds, test whether there is any significant difference between the mean yield of crops in the two district

Soln:- Given

$\bar{x}_1 = 210$, Sample S.D $(s_1) =$

P): p) Samples of students were drawn from two University and from their weights in kilograms, mean and standard deviations are calculated and shown below. Make a large sample test to test the significance of the difference between the means.

| | Mean | S.D | Size of the sample |
|---|---|---|---|
| University A | 55 | 10 | 400 |
| University B | 57 | 15 | 100 |

Soln:- Given,

$\bar{x}_1 = 55, \quad \bar{x}_2 = 57, \quad s_1 = 10, \quad s_2 = 15$

$n_1 = 400, \quad n_2 = 100.$

Null hypothesis: $H_0: \bar{x}_1 = \bar{x}_2$

Alternative hypothesis: $H_1: \bar{x}_1 \neq \bar{x}_2$.

Level of significance: $\alpha = 0.05$

Test statistic $\quad z = \dfrac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

$$= \frac{|55-57|}{\sqrt{\frac{100}{400} + \frac{225}{100}}} = 1.26$$

$z_{cal} = 1.26$

$z_{tab} = 1.96$

$z_{cal} < z_{tab}$

Hence we ~~reject~~ accept $H_0$.

Scanned with CamScanner

2) The research investigator is interested in studying whether there is a significant difference in the salaries of MBA grads in two metropolitan cities. A random sample of size 100 from Mumbai yield an average income of Rs 20,150. Another random sample of 60 from chennai result in an average income of Rs 20,250. If the variances of both the population are given as $\sigma_1^2$ = Rs 40,000 and $\sigma_2^2$ = Rs 32,400 respectively.

Soln:- A) Given,

$n_1 = 100$, $\bar{x}_1 = 20,150$, and $n_2 = 60$, $\bar{x}_2 = 20,250$

and $\sigma_1^2 = 40,000$, $\sigma_2^2 = 32,400$.

Null hypothesis : $H_0 : \bar{x}_1 = \bar{x}_2$

Alternative hypothesis $H_1 : \bar{x}_1 \neq \bar{x}_2$.

Level of significance : $\alpha = 0.05$.

The test statistic $z = \dfrac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

$= \dfrac{|20150 - 20250|}{\sqrt{\dfrac{40,000}{100} + \dfrac{32400}{60}}}$

$= 3.26$.

$z_{cal} = 3.26$

$z_{tab} = 1.96$

$z_{cal} >$ ~~$z_{tab}$~~

$z_{cal} > z_{tab}$.

Hence we reject $H_0$.

F) A researcher wants to know the intelligence of students in a school. He selected two groups of students. In the first group, there are 150 students having mean IQ of 75 with a S.D of 15 in the second group there are 250 students having mean IQ of 70 with S.D of 20.

Soln:- Given

$n_1 = 150, \quad \bar{x}_1 = 75, \quad s_1 = 15$

$n_2 = 250, \quad \bar{x}_2 = 70, \quad s_2 = 20.$

Null hypothesis $H_0: \bar{x}_1 = \bar{x}_2$

Alternative hypothesis $H_1: \bar{x}_1 \neq \bar{x}_2$

Level of Significance : $\alpha = 0.05$.

Test statistic $z = \dfrac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

$= \dfrac{|75 - 70|}{\sqrt{\dfrac{(15)^2}{150} + \dfrac{(20)^2}{250}}} = 2.7116$

$z_{cal} = 2.7116$

$z_{tab} = 1.96$

$z_{cal} > z_{tab}$.

Hence we reject $H_0$.

Scanned with CamScanner

HW

2) P) Two types of new cars produced in USA are tested for petrol mileage, one sample is consisting of 42 cars averaged 15 kmpl while the other sample consisting of 80 cars averaged 11.5 kmpl with population variance as $\sigma_1^2 = 2.0$ and $\sigma_2^2 = 1.5$ respectively. Test whether there is any significant difference in the petrol consumption of these two types of cars? (use $\alpha = 0.01$)

Sol:- A) Given

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

$s_1^2 = 2.0$, $s_2^2 = 1.5$

Null hypothesis $H_0: \bar{x}_1 = \bar{x}_2$

Alternative hypothesis $H_1: \bar{x}_1 \neq \bar{x}_2$

Level of significance $\alpha = 0.01$

Test statistic $$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{|15 - 11.5|}{\sqrt{\frac{2}{42} + \frac{1.5}{80}}} = 13.587$$

$z_{cal} = 13.587$, $z_{tab}$ at 0.01 LOS = 2.58

$z_{cal} > z_{tab}$

Hence we reject $H_0$.

P) P) A simple sample of the height of 6400 English men has a mean of 67.85 inches and a S.D of 2.56 inches while a simple sample of heights of 1600 Australians has a mean of 68.55 inches and S.D of 2.52 inches. Do the data indicate the Australians are on the average taller than the English men?

Solⁿ:-

Given

$n_1 = 6400, \quad \bar{x}_1 = 67.85, \quad s_1 = 2.56$

$n_2 = 1600, \quad \bar{x}_2 = 68.55, \quad s_2 = 2.52.$

Null hypothesis $H_0: \bar{x}_1 = \bar{x}_2$

Alternative hypothesis: $H_1: \cancel{\bar{x}_1 \neq \bar{x}_2}\ \bar{x}_1 < \bar{x}_2$

Level of significance $\alpha: 0.05$

Test statistic $z = \dfrac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

$= \dfrac{|20150 - 20250|}{\sqrt{\dfrac{40000}{100} + \dfrac{32400}{60}}} = \dfrac{|67.85 - 68.55|}{\sqrt{\dfrac{(2.56)^2}{6400} + \dfrac{(2.52)^2}{1600}}}$

$= \cancel{3.26}\ 9.9.$

$z_{tab}$ Value at 0.05 (5%) L.O.S = $\cancel{3.196}$ 1.645

$z_{cal} > z_{tab}.$

Hence we reject the $H_0$ at 5% L.O.S.

P) The mean life of a sample of 10 electric bulbs was found to be 1456 hours with S.D of 423 hours. A second sample of 17 bulbs chosen from a different batch showed the mean life of 1280 hours with S.D of 398 hours. Is there any significant difference between the means of two batches?

Solⁿ:-

We are given

$n_1 = 10, \quad \bar{x}_1 = 1456, \quad s_1 = 423.$

$n_2 = 17, \quad \bar{x}_2 = 1280, \quad s_2 = 398$

Null hypothesis $H_0: \bar{x}_1 = \bar{x}_2$

Alternative hypothesis $H_1: \cancel{\bar{x}_1 \neq \bar{x}_2}\ \bar{x}_1 < \bar{x}_2\ \ \bar{x}_1 \neq \bar{x}_2$

Level of significance $\alpha = 0.05$

Test statistic $z = \dfrac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

$$= \frac{|1456 - 1280|}{\sqrt{\dfrac{(423)^2}{10} + \dfrac{(376)^2}{17}}} = 1.067$$

$z_{cal} = 1.067$

$z_{tab}$ at $0.05$ L.O.S $= 1.96$

$z_{cal} < z_{tab}$

Hence we accept our $H_0$ at 5% L.O.S

Hence we conclude that $\bar{x}_1 = \bar{x}_2$

# TEST OF SIGNIFICANCE FOR SINGLE PROPORTION

For testing the significance of single proportion, we write the null hypothesis as

$H_0: p = P$ (or) $P = P_0$ ~ $H_1: p \neq P$ (or) $P \neq P_0$.

To test the above $H_0$, we write the Z-test statistic as

$$z = \frac{|p - P|}{\sqrt{\frac{PQ}{n}}},$$

If $Z_{cal} \leq Z_{tab}$, then we accept $H_0$ otherwise we reject it.

## Problems

P) 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% level of significance?

Soln:- Given, $n = 1000$

$p$ = sample proportion of rice eaters $= \frac{540}{1000} = 0.54$

$P$ = population proportion of rice eaters $= \frac{1}{2} = 0.5$

$Q = 1 - P = 1 - 0.5 = 0.5$

Null hypothesis : $H_0: p = P$ (or) $P = 0.5$

Alternative hypothesis : $H_1: p \neq P$ (or) $P \neq 0.5$

Level of significance : $\alpha = 0.05$

Test statistic $z = \dfrac{|p - P|}{\sqrt{\frac{PQ}{n}}}$

$$= \frac{|0.54 - 0.5|}{\sqrt{\frac{0.5 \times 0.5}{1000}}} = 2.532$$

$Z_{cal} = 2.532$

$Z_{tab}$ at 1% L.O.S = 2.58.

$Z_{cal} < Z_{tab}$

Hence we accept $H_0$.

P) 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?

Sol$^n$:- Given $n = 600$

Number of smokers $= 325$

$p$ = Sample proportion of smokers $= \frac{325}{600} = 0.5417$

$P$ = Population proportion of smokers in the city $= \frac{1}{2} = 0.5$

$Q = 1 - P = 1 - 0.5 = 0.5$

Null hypothesis : $H_0 : p = P$ (or) $P = 0.5$

Alternative hypothesis : $H_1 : p > P$ (or) ~~$P \neq 0.5$~~ $P > 0.5$

Level of significance : $\alpha = 0.05$

Test statistic $Z = \frac{|p - P|}{\sqrt{\frac{PQ}{n}}} = \frac{|0.5417 - 0.5|}{\sqrt{\frac{0.5 \times 0.5}{600}}}$

$Z_{cal} = 2.04$.

$Z_{tab} = 1.645$ at 5% L.O.S

$Z_{cal} > Z_{tab}$

Hence we reject $H_0$

P)

P) In a random sample of 185 cool drinkers, 68 said they prefer thums up to pepsi. Test the null hypothesis $P = 0.5$ against the alternative hypothesis $P > 0.5$

Sol$^n$:- We have,

$n = 185$, $x = 68$ and $p = \frac{x}{n} = \frac{68}{125} = 0.544$

Null hypothesis : $H_0 : P = 0.5$ (ie) $p = P$

Alternative hypothesis : $H_1 : P > 0.5$

Level of significance : $\alpha = 0.05$

Test statistic : $z = \frac{|p - P|}{\sqrt{\frac{PQ}{n}}}$.

$$= \frac{0.544 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{125}}} = \frac{0.044}{0.045} = 0.9839$$

$Z_{cal} = 0.9839$

$Z_{tab} = 1.645$

$Z_{cal} < Z_{tab}$

Hence we accept $H_0$.

P) 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.

Sol$^n$:- Given $n = 400$, $x = 50$ and $p = \frac{x}{n} = \frac{50}{400} = 0.125$

Null hypothesis : $H_0 : P = 0.2$

Alternative hypothesis : $H_1 : P \neq 0.2$

Level of significance : $\alpha = 0.05$

Test statistic $z = \frac{|p - P|}{\sqrt{\frac{PQ}{n}}} = \frac{|0.125 - 0.2|}{\sqrt{\frac{(0.2)(0.8)}{400}}} = 3.75$.

$Z_{tab} = 1.96$

$Z_{cal} > Z_{tab}$

Hence we reject $H_0$

P) A manufacturer claims that only 4% of his products are defective. A random sample of 500 were taken among which 100 were defective. Test the hypothesis at 0.05 level.

Sol:- We have

$n = 500, \quad x = 100, \quad p = \frac{x}{n} = \frac{100}{500} = 0.2$

and $P = 4\% = \frac{4}{100} = 0.04, \quad Q = 1 - P = 1 - 0.04 = 0.96$

Null hypothesis : $H_0 : P = 0.04$. (or) $p = P$

Alternative hypothesis : $H_1 : P \neq 0.04$ (or) $p \neq P$.

Level of significance : $\alpha = 0.05$

Test statistic : $z = \dfrac{|p - P|}{\sqrt{\frac{PQ}{n}}}$

$= \dfrac{|0.2 - 0.04|}{\sqrt{\frac{(0.04)(0.96)}{500}}} = 18.26.$

$z_{cal} = 18.26, \quad z_{tab} = 1.96$

$z_{cal} > z_{tab}$ value.

Hence we reject $H_0$.

2) In a sample of 500 from a village in Andhra Pradesh, 280 are found to be rice eaters and the rest are wheat eaters. Can we assume that the both are equally popular?

Sol:- Given, $n = 500, \quad p = \frac{x}{n} = \frac{280}{500} = 0.56$

$P = \frac{1}{2} = 0.5$

$Q = 1 - P = 1 - 0.5 = 0.5$

Null hypothesis : $H_0 : p = P$ (or) $P = 0.5$

Alternative hypothesis : $H_1 : p \neq P$ (or) $P \neq 0.5$.

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

$$= \frac{|0.56 - 0.5|}{\sqrt{\frac{(0.5)(0.5)}{500}}} = \frac{0.06}{0.022} = 2.68$$

$\therefore Z_{cal} = 2.68$

$\therefore Z_{tab} = 2.58$

$\therefore Z_{cal} > Z_{tab}$ value

Hence we reject $H_0$

Note :- The 99% confidence interval for proportion (CI) true proportion is given by

$$\left(P - 3\sqrt{\frac{PQ}{n}}, \; P + 3\sqrt{\frac{PQ}{n}}\right)$$

P) Among 900 people in a state 90 are found to be chapati eaters. Construct 99% confidence interval for the true proportion.

Soln:- Given, $x = 90$, $n = 900$

$\therefore P = \dfrac{x}{n} = \dfrac{90}{900} = \dfrac{1}{10} = 0.1$ and

$Q = 1 - P = 1 - 0.1 = 0.9$

Now $\sqrt{\dfrac{PQ}{n}} = \sqrt{\dfrac{(0.1)(0.9)}{900}} = 0.01$

$\therefore$ Confidence interval is

$$\left(P - 3\sqrt{\frac{PQ}{n}}, \; P + 3\sqrt{\frac{PQ}{n}}\right)$$

$(0.1 - 0.03, \; 0.1 + 0.03)$

$(0.07, \; 0.13)$

P)

1) In a random sample of 160 workers exposed to a certain amount of radiation, 24 experienced some ill effects. Construct a 99% confidence interval for the corresponding true proportion.

Sol^n:- Given,

$n = 160,\ x = 24,\ n = 160$ and $P = \frac{x}{n} = \frac{24}{160} = 0.15$

$Q = 1 - P = 1 - 0.15 = 0.85$

$\therefore$ The 99% confidence interval is given by.

$\left(p - 3\sqrt{\frac{PQ}{n}},\ p + 3\sqrt{\frac{PQ}{n}}\right)$

$(0.15 - 3\times0.028,\ 0.15 + 3\times0.028)$

$(0.065,\ 0.234)$

P) In a random sample of 400 industrial accidents, it was found that 231 were due at least partially to unsafe working conditions. Construct a 99% confidence interval for the corresponding true proportion.

Sol^n:- we have $n = 400,\ x = 231,\ P = \frac{x}{n} = \frac{231}{400} = 0.5775$

$Q = 1 - P = 1 - 0.577 = 0.4225$

Now $\sqrt{\frac{PQ}{n}} = \sqrt{\frac{0.5775\times0.4225}{400}} = 0.0247.$

$\therefore$ The 99% Confidence interval is

$\left(p - 3\sqrt{\frac{PQ}{n}},\ p + 3\sqrt{\frac{PQ}{n}}\right)$

$(0.5775 - 3\times0.0247,\ 0.5775 + 3\times0.0247)$

$(0.5034,\ 0.6516).$

P) Of 80 patients are treated with an antibiotic 59 get cured. Find a 99% confidence limits to the true population of cure.

Soln: Given,

$n = 80,\ x = 59$ and $p = \frac{x}{n} = \frac{59}{80} = 0.7375$

$Q = 1 - P = 1 - 0.7375 = 0.2625$

Now $\sqrt{\frac{PQ}{n}} = \sqrt{\frac{0.7375 \times 0.2625}{80}} = 0.049.$

$\therefore$ The 99% confidence limits are given by

$\left(p - 3\sqrt{\frac{PQ}{n}},\ p + 3\sqrt{\frac{PQ}{n}}\right)$

$(0.7375 - 3 \times 0.049,\ 0.7375 + 3 \times 0.049)$

$(0.59,\ 0.88)$

## TEST FOR EQUALITY OF TWO PROPORTIONS (OR TEST FOR SIGNIFICANT DIFFERENCE BETWEEN TWO PROPORTIONS)

Let $p_1$ and $p_2$ be the proportions in two 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 population proportions $P_1$ and $P_2$ are equal

The Null hypothesis $H_0: P_1 = P_2$ (or) $p_1 = p_2$

The Alternative hypothesis $H_1: P_1 \neq P_2$. (or) $p_1 \neq p_2$

To test the above $H_0$, we write the Z-test statistic as

$$z = \frac{|p_1 - p_2|}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ and $q = 1 - p$

(or) $p = \frac{x_1 + x_2}{n_1 + n_2}$

If $z_{cal} \le z_{tab}$, then we accept ~~on~~ $H_0$ otherwise we reject it.

## Problems.

P) 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 proposal are same, at 5% level.

Sol:- Given sample sizes $n_1 = 400$, $n_2 = 600$

Proportion of men, $p_1 = \frac{200}{400} = \frac{1}{2} = 0.5$

Proportion of women, $p_2 = \frac{325}{600} = 0.541$.

$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{200+325}{400+600} = 0.525$, $q = 1 - P = 1 - 0.525 = 0.475$

(i) Null hypothesis $H_0 : P_1 = P_2$

(ii) Alternative hypothesis $H_1 : P_1 \neq P_2$

(iii) Level of significance : $\alpha = 0.05$

(iv) Test statistic $z = \dfrac{|P_1 - P_2|}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$= \dfrac{|0.5 - 0.541|}{\sqrt{(0.525 \times 0.475)\left(\frac{1}{400} + \frac{1}{600}\right)}} = 1.28$

~~Sence~~ ~~zcal~~ $z_{tab}$ at 5% L.O.S = 1.96

(v) Conclusion :- $z_{cal} < z_{tab}$

Hence we accept $H_0$

Hence we conclude that $P_1 = P_2$

P) On the basis of their total scores, 200 candidates of a Civil Service examination are divided in to two groups, the upper 30% and the remaining 70%. Consider the first question of the examination. Among the first group, 40 had the correct answer, where as among the second group, 80 had the correct answer. On the basis of these results, Can one conclude that the first question is not good at discriminating ability of the type being examined here?

Sol:- we have, $n_1 = 200 \times \frac{30}{100} = 60$, $n_2 = \cancel{200-60} = 200-60 = 140.$ (70%)

$x_1 = 40$, $x_2 = 80.$

$p_1 = \frac{x_1}{n_1} = \frac{40}{60} = \frac{20}{30} = 0.667$, $p_2 = \frac{x_2}{n_2} = \frac{80}{140} = 0.571.$

$$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{40+80}{60+140} = \frac{120}{200} = 0.6.$$

$$q = 1 - P = 1 - 0.6 = 0.4.$$

(i) Null hypothesis: $H_0: p_1 = p_2$

(ii) Alternative hypothesis: $H_1: p_1 \neq p_2$

(iii) Level of significance: $\alpha = 0.05$ (assumed)

(iv) the test statistic $z = \frac{|p_1 - p_2|}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$= \frac{|0.667 - 0.571|}{\sqrt{(0.6)(0.4)\left(\frac{1}{60} + \frac{1}{40}\right)}}$$

$$= \frac{0.096}{0.0756} = 1.27$$

$Z_{cal} = 1.27$

$Z_{tab}$ at 5% $Z_{0.05} = 1.96$

(v) Conclusion:- $Z_{cal} < Z_{tab}$

Hence we accept $H_0$

Hence we conclude that $p_1 = p_2$.

1) In a random sample of 1000 persons from town A, 400 are found to be consumers of wheat. In a sample of 800 from town B, 400 are found to be consumers of wheat. Do these data reveal a significant difference between town A and town B, so far as the proportion of wheat consumers is concerned?

Sol:- Given,

$n_1 = 100, \quad n_2 = 800$

$x_1 = 400, \quad x_2 = 400$

$p_1 = \frac{x_1}{n_1} = \frac{400}{1000} = 0.4$

$p_2 = \frac{x_2}{n_2} = \frac{400}{800} = 0.5$

$\therefore p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{400+400}{1000+800} = \frac{800}{1600} = \frac{8}{16} = \frac{4}{9}$

$q = 1 - p = 1 - \frac{4}{9} = \frac{5}{9}$

(i) Null hypothesis : $H_0 : P_1 = P_2$.

(ii) Alternative hypothesis : $H_1 : P_1 \neq P_2$.

(iii) Level of Significance : $\alpha = 0.05$

(iv) The test statistic :

$$z = \frac{|p_1 - p_2|}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{|0.4 - 0.5|}{\sqrt{\frac{4}{9} \times \frac{5}{9}\left(\frac{1}{1000} + \frac{1}{800}\right)}}$$

$$= 4.242.$$

$\therefore z_{tab}$ at 5% L.O.S = 1.96.

$z_{cal} > z_{tab}$

Hence we reject $H_0$

Hence we conclude that $P_1 \neq P_2$

Scanned with CamScanner

P) In two large populations, there are 30% and 25% respectively of fair haired people. Is this difference likely to be hidden in samples of 1200 and 900 respectively from the two populations.

Solⁿ:- Given

$n_1 = 1200$, $n_2 = 900$

$p_1$ = Proportion of fair haired people in the first population $= \frac{30}{100} = 0.3$.

$p_2$ = Proportion of fair haired people in the second population $= \frac{25}{100} = 0.25$

(i) Null hypothesis : $H_0 : p_1 = p_2$

(ii) Alternative hypothesis : $H_1 : p_1 \neq p_2$.

(iii) Test statistic,

$$z = \frac{|p_1 - p_2|}{\sqrt{\frac{p_1 Q_1}{n_1} + \frac{p_2 Q_2}{n_2}}}$$

where $Q_1 = 1 - p_1 = 1 - 0.3 = 0.7$

$Q_2 = 1 - P_2 = 1 - 0.25 = 0.75$.

$$\therefore Z = \frac{|0.3 - 0.25|}{\sqrt{\frac{0.3 \times 0.7}{1200} + \frac{0.25 \times 0.75}{900}}} = 2.55$$

$Z_{tab}$ value at 5% L.O.S = ~~2.5~~ 1.96

$\therefore Z_{cal} > Z_{tab}$ Value

Hence we reject $H_0$.

Hence we conclude that $p_1 \neq p_2$.