



||Jai Sri Gurudev||
S.J.C. INSTITUTE OF TECHNOLOGY, CHICKBALLAPUR
DEPARTMENT OF MATHEMATICS
MATHEMATICS-3 FOR COMPUTER SCIENCE (BCS301)
MODULE - 3
STATISTICAL INFERENCE -1

Prepared by:

Purushotham P
 Assistant Professor
 SJC Institute of Technology
 Email id: purushotham@sjcit.ac.in
 Phone: 7338005244

Introduction:

Sampling is a statistical method of obtaining representative data (observations) from a group. We have been using sampling concepts in our day to day lives knowingly or unknowingly; for instance we take a handful of rice to check the rice quality of the full lot. This is an example of random sampling from a large population.

Population (Universe):

The group of objects (individuals) under study is called population or universe. Universe may be finite or infinite.

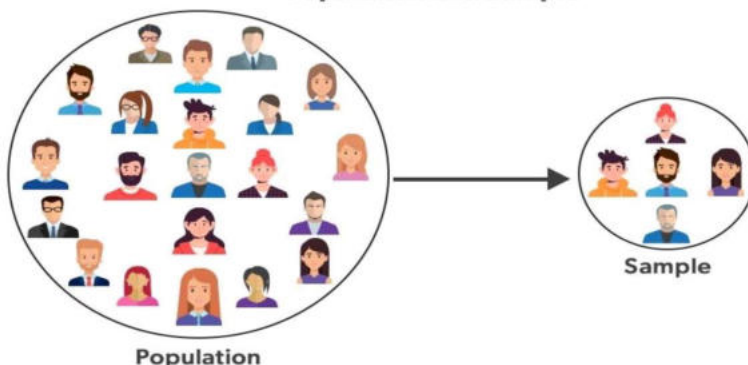
Sample:

A part containing objects(individuals), selected from the population is called a sample.

Sample size:

The number of individuals in a sample is called a sample size. If the sample size n is less than or equal to 30, then the sample is said to be small, otherwise it is called a large sample.

Population and Sample



Random Sampling:

The selection of objects (individuals) from the universe in such a way that each object (individual) of the universe has the same chance of being selected is called random sampling. Lottery system is the most common example of random sampling.

Every random sampling need not be simple. For example, if balls are drawn without replacement from a bag of balls containing different balls; the probability of success changes in every trial. Thus, the sampling though random is not simple.

Simple Sampling:

Simple sampling is a special case of random sampling in which each event has same probability of success or failure.

Hypothesis:

A hypothesis is an assumption based on insubstantial evidences that lends itself to further testing and experimentation. For example a farmer claims significant increase in crop production after using a particular fertilizer and after a season of experimenting, his hypothesis may be proved true or false. Any hypothesis may be accepted or rejected as per specific confidence levels and must be admissible to refutation.

Null Hypothesis:

The **null hypothesis** is a general statement or default position that there is no relationship between two measured phenomena or no association among groups.

Example: Given the test scores of two random samples, one of men and one of women, does one group differ from the other? A possible null hypothesis is that the mean male score is the same as the mean female score:

$$H_0: \mu_1 = \mu_2$$

where

H_0 = the null hypothesis,

μ_1 = the mean of population 1, and

μ_2 = the mean of population 2.

A stronger null hypothesis is that the two samples are drawn from the same population, such that the variances and shapes of the distributions are also equal.

Alternative Hypothesis:

It is the opposite statement of null hypothesis and denoted by $H_1: \mu_1 \neq \mu_2$

Significance levels (α):

The significance level of an event (such as a statistical test) is the probability that the event could have occurred by chance. If the level is quite low, that is, the probability of occurring by chance is quite small, we say the event is significant.

The level of significance is the measurement of the statistical significance. It defines whether the null hypothesis is assumed to be accepted or rejected. It is expected to identify if the result is statistically significant for the null hypothesis to be false or rejected.

$$\alpha = 5\% \quad \alpha = 1\% \quad \alpha = 0.27\%$$

Example: A level of significance of $p=0.05$ means that there is a 95% probability that the results found in the study are the result of a true relationship/difference between groups being compared. It also means that there is a 5% chance that the results were found by chance alone and no true relationship exists between groups.

Standard Error:

The standard deviation of the sampling distribution of a statistic is Known as Standard Error (S.E.).

Precision:

Reciprocal of standard error is known as precision.

Confidence Limits:

In short, confidence limits show how accurate an estimation of the mean is or is likely to be. Confidence limits are the lowest and the highest numbers at the end of a **confidence interval**.

Confidence Interval:

A confidence interval is a range around a measurement that conveys how precise the measurement is. A confidence interval, in statistics, refers to the probability that a population parameter will fall between a set of values for a certain proportion of times. Analysts often use confidence intervals that contain either 95% or 99% of expected observations.

Critical Value:

A critical value is the value of the test statistic which defines the upper and lower bounds of a confidence interval, or which defines the threshold of statistical significance in a statistical test.

Types of test	Level of Significance		
	1%	5%	10%
Two tailed test	2.58	1.96	1.645
One tailed test	2.33	1.645	1.28

Critical Region:

A critical region, also known as the rejection region, is a set of values for the test statistic for which the null hypothesis is rejected. i.e. if the observed test statistic is in the critical region then we reject the null hypothesis and accept the alternative hypothesis.

Type I and Type II Errors:

When we test a statistic at specified confidence level, there are chances of taking wrong decisions due to small sample size or sampling fluctuations etc.

Type I error is the incorrect rejection of a true null hypothesis, i.e. we reject H_0 , when it is true, whereas Type II error is the incorrect acceptance of a false null hypothesis, i.e. we accept H_0 when it is false.

One Tailed and Two Tailed Tests:

While testing statistical significance levels; one- tailed test and a two-tailed test are used for accepting or rejecting a hypothesis. One- tailed tests are used for asymmetric distributions (reference value is unidirectional) which have a single tail; such as the chi-square distribution.

A two-tailed test is appropriate if the estimated value may lie on both sides of reference value. Two-tailed tests are only applicable when the probability curve has two tails; such as normal distribution.

Test of hypothesis:

Let x be the observed number of successes in a sample size of n and $\mu = np$ be the expected number of successes. Then the standard normal variate Z is defined as

$$Z = \frac{x - \mu}{\sigma} = \frac{x - np}{\sqrt{npq}}$$

Test of hypothesis for means:

Let μ_1, μ_2 be the means, σ_1, σ_2 be the standard deviations of two populations and $\bar{x}_1, \bar{x}_2$ are the means of the samples, then

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

If the samples are drawn from the same population, then $\sigma_1 = \sigma_2 = \sigma$ we have

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

PROBLEMS

- 1) A coin is tossed 1000 times and head turns up 540 times. Decide on the hypothesis that the coin is unbiased at 1 % level of significance.

Sol.

Let us suppose that the coin is unbiased.

and let p = the probability of getting a head in one toss = $1/2 = 0.5$

Since $p+q=1$, $q=1-p=1/2=0.5$

Expected number of heads in 1000 tosses = $np = 1000 \times 0.5 = 500$, $npq = 250$

$\therefore$ The difference is $x - \mu = 540 - 500 = 40$

$\therefore$ Consider $Z = \frac{x - \mu}{\sigma} = \frac{x - np}{\sqrt{npq}}$

$$\Rightarrow Z = \frac{40}{\sqrt{250}} = 2.53 < 2.58$$

1% level of significance = 99% confidence level.

Therefore accept the hypothesis that the coin is unbiased.

- 2) A coin is tossed 400 times and turns up head 216 times. Test the hypothesis that the coin is unbiased at 5% level of significance.

Sol.

Let us suppose that the coin is unbiased.

and let p = the probability of getting a head in one toss = $1/2 = 0.5$

Since $p+q=1$, $q=1-p=1/2=0.5$

Expected number of heads in 400 tosses = $np = 400 \times 0.5 = 200$, $npq = 100$

$\therefore$ The difference is $x - \mu = 216 - 200 = 16$

$\therefore$ Consider $Z = \frac{x - \mu}{\sigma} = \frac{x - np}{\sqrt{npq}}$

$$\Rightarrow Z = \frac{16}{\sqrt{100}} = \frac{16}{10} = 1.6 < 1.96$$

Critical value of z at $\alpha = 0.05$ is 1.96

Therefore accept the hypothesis that the coin is unbiased at the 5% level of significance.

3) A coin was tossed 1600 times and the tailed turned up 864 times. Test the hypothesis that the Coin is unbiased at 1% level of significance.

Sol.

Let us suppose that the coin is unbiased.

and let p = the probability of getting a tail in one toss = $1/2 = 0.5$

Since $p + q = 1$, $q = 1 - p = 1/2 = 0.5$

Expected number of tailed in 1600 tosses = $np = 1600 \times 0.5 = 800$, $npq = 400$

$\therefore$ The difference is $x - \mu = 864 - 800 = 64$

$\therefore$ Consider $Z = \frac{x - \mu}{\sigma} = \frac{x - np}{\sqrt{npq}}$

$$\Rightarrow Z = \frac{64}{\sqrt{400}} = 3.2 > 2.58$$

1% level of significance = 99% confidence level.

Therefore accept the hypothesis that the coin is biased.

4) In 324 throws of a six faced 'die', an odd number turned up 181 times. Is it possible to think that the 'die' is an unbiased one?

Sol.

Let us suppose that the die is unbiased.

and let p = the probability of the turn up of an odd number is $= 3/6 = 1/2 = 0.5$

Since $p + q = 1$, $q = 1 - p = 1/2 = 0.5$

Expected number of successes = $np = 324 \times 0.5 = 162$, $npq = 81$

$\therefore$ The difference is $x - \mu = 181 - 162 = 19$

$\therefore$ Consider $Z = \frac{x - \mu}{\sigma} = \frac{x - np}{\sqrt{npq}}$

$$\Rightarrow Z = \frac{19}{\sqrt{81}} = \frac{19}{9} = 2.11 < 2.58$$

Thus we can that the die is unbiased

5) A die is thrown 9000 times and a throw of 3 or 4 was observed 3240 times. Show that the die Can not be regarded as an unbiased one.

Sol.

The probability of getting 3 or 4 in a single through is $p = \frac{2}{6} = \frac{1}{3}$

$$\text{And } q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$$

$$\therefore \text{Expected number of success} = \frac{1}{3} \times 9000 = 3000$$

$$\therefore \text{The difference} = 3240 - 3000 = 240$$

$$Z = \frac{x - np}{\sqrt{npq}}$$

$$\text{Consider } \Rightarrow Z = \frac{(3240) - (9000 \times \frac{1}{3})}{\sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}}}$$

$$\Rightarrow Z = \frac{240}{\sqrt{2000}}$$

$$\Rightarrow Z = 5.37$$

Since $Z = 5.37 > 2.58$,

We conclude that the die is biased.

Test of significance for proportion:

Test of significance of single proportion:

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

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}, \text{ where } P+Q=1 \Rightarrow Q=1-P, n=\text{Sample size}$$

The formulated Null and Alternative hypothesis is, $H_0: P = a$ specified value, $H_1: P \neq a$ specified value

Test of significance of Difference between two sample proportions:

To test the significance of the difference between the samples proportions, the test statistic under the null hypothesis H_0 that there is no significance difference between the two sample proportions,

$$\text{We have } Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \text{ Where } P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \text{ or } P = \frac{x_1 + x_2}{n_1 + n_2} \text{ and } P+Q=1,$$

Here p_1 and p_2 are the sample proportions in respect of an attribute corresponding to two large samples of size n_1 and n_2 drawn from the two populations.

PROBLEMS

1. A coin is tossed 400 times and it turns up head 216 times. Discuss whether the coin may be regarded as unbiased one. Sol.

Set the null hypothesis $H_0: P = \frac{1}{2}$

Set the Alternative hypothesis $H_1: P \neq \frac{1}{2}$

The level of significance $\alpha = 0.05$ (5%)

$\therefore$ The test statistic $Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$, where $P+Q=1 \Rightarrow Q=1-P$

Given, the coin is tossed and it turns up in the equal proportion

$$P = \frac{1}{2} \Rightarrow Q = 1 - P$$

$$\Rightarrow Q = 1 - \frac{1}{2} = \frac{1}{2}$$

And the coin turns up head 216 times when it tossed $n = 400$ times

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

$$\therefore Z = \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{400}}}$$

$$\Rightarrow Z = \frac{0.04}{\sqrt{0.000625}} = 1.6$$

At 5% level, the tabulated value of Z_α is 1.96

Since $|Z| = 1.6 < 1.96$

Hence, the null hypothesis is accepted at 5% level of significance and the coin may be regarded as unbiased.

2. In a city of sample of 500 people, 280 are tea drinkers and the rest are coffee drinkers. Can we assume that both coffee and tea are equally popular in this city at 5% Los. Sol.

Set the null hypothesis $H_0: P = \frac{1}{2}$ (Both coffee and tea drinkers are equally popular)

Set the Alternative hypothesis $H_1: P \neq \frac{1}{2}$

The level of significance $\alpha = 0.05$ (5%)

$\therefore$ The test statistic $Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$, where $P+Q=1 \Rightarrow Q=1-P$

$$P = \frac{1}{2} \Rightarrow Q = 1 - P$$

$$\Rightarrow Q = 1 - \frac{1}{2} = \frac{1}{2}$$

$$\therefore p = \frac{280}{500} = 0.56, \text{ where } n = 500$$

$$\therefore Z = \frac{0.56 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{500}}}$$

$$\Rightarrow Z = \frac{0.06}{\sqrt{0.0005}} = 2.68$$

At 5% level, the tabulated value of Z_α is 1.96

Since $|Z| = 2.68 > 1.96$

Hence, the null hypothesis is rejected at 5% level of significance and both the drinkers are not popular.

3. A manufacturing company claims that at least 95% of its products supplied confirm to the specifications out of a sample of 200 products, 18 are defective. Test the claim at 5% Los.

Sol.

Set the null hypothesis $H_0: P = 95\% = 0.95$

Set the Alternative hypothesis $H_1: P \neq 0.95$

The level of significance $\alpha = 0.05$ (5%)

$\therefore$ The test statistic $Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$, where $P+Q=1 \Rightarrow Q=1-P$

Given,

$$P = 95\% = 0.95 \Rightarrow Q = 1 - P$$

$$\Rightarrow Q = 1 - 0.95 = 0.05$$

Found 18 products are defective out of 200 sample products

$\therefore$ The total defective less products (Non defective) = $200 - 18 = 182$

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

$$\therefore Z = \frac{0.91 - 0.95}{\sqrt{\frac{0.95 \times 0.05}{200}}}$$

$$\Rightarrow Z = -\frac{0.04}{\sqrt{0.0002375}} = -2.5955$$

At 5% level, the tabulated value of Z_α is 1.96

Since $|Z| = 2.5955 > 1.96$

Hence, the null hypothesis is rejected at 5% level of significance

4. If a sample of 300 units of a manufactured product 65 units were found to be defective and in another sample of 200 units, there were 35 defectives. Is there significant difference in the proportion of defectives in the samples at 5% Los.

Sol.

Set the null hypothesis $H_0: P_1 = P_2$

Set the Alternative hypothesis $H_1: P_1 \neq P_2$

The level of significance $\alpha = 0.05$ (5%)

Given

$$n_1 = 300, n_2 = 200$$

The sample of 300 units of a manufactured product 65 units were found to be defective

$$\therefore p_1 = \frac{65}{300} = 0.2166 = 0.22$$

The sample of 200 units of a manufactured product 35 units were found to be defective

$$\therefore p_2 = \frac{35}{200} = 0.1750$$

We know that $P = \frac{x_1 + x_2}{n_1 + n_2}$

$$\Rightarrow P = \frac{65 + 35}{300 + 200}$$

$$\Rightarrow P = \frac{100}{500}$$

$$\Rightarrow P = 0.2$$

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

$$\begin{aligned}
 \therefore Z &= \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \\
 \Rightarrow Z &= \frac{0.22 - 0.1750}{\sqrt{(0.2 \times 0.8)\left(\frac{1}{300} + \frac{1}{200}\right)}} \\
 \Rightarrow Z &= \frac{0.045}{\sqrt{(0.16)(0.00833)}} \\
 \Rightarrow Z &= \frac{0.045}{\sqrt{0.001328}} \\
 \Rightarrow Z &= \frac{0.045}{0.03644} \\
 \Rightarrow Z &= 1.233
 \end{aligned}$$

At 5% level, the tabulated value of Z_α is 1.96

Since $|Z| = 1.233 < 1.96$

Hence, the null hypothesis is accepted at 5% level of significance

5. In a large city A, 20% of a random sample of 900 school boys had a slight physical defect. In another large city B, 18.5% of a random sample of 1600 school boys had the same defect. Is the difference between the proportions significant?

Sol.

Set the null hypothesis $H_0: P_1 = P_2$

Set the Alternative hypothesis $H_1: P_1 \neq P_2$

The level of significance $\alpha = 0.05$ (5%)

Given

$$n_1 = 900, n_2 = 1600$$

$$x_1 = 20\% \text{ of random sample of } 900 = 0.2 \times 900 = 180$$

$$x_2 = 18.5\% \text{ of random sample of } 1600 = 0.185 \times 1600 = 296$$

$$\therefore p_1 = 20\% = \frac{20}{100} = 0.2, p_2 = 18.5\% = \frac{18.5}{100} = 0.185$$

$$\text{We know that } P = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\Rightarrow P = \frac{180 + 296}{900 + 1600}$$

$$\Rightarrow P = \frac{476}{2500}$$

$$\Rightarrow P = 0.1904 \Rightarrow Q = 1 - P = 1 - 0.1904 = 0.8096$$

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

$$\Rightarrow Z = \frac{0.2 - 0.185}{\sqrt{(0.1904 \times 0.8096)\left(\frac{1}{900} + \frac{1}{1600}\right)}}$$

$$\Rightarrow Z = \frac{0.015}{\sqrt{(0.1541)(0.00173)}}$$

$$\Rightarrow Z = \frac{0.015}{\sqrt{0.00026}}$$

$$\Rightarrow Z = \frac{0.015}{0.01612}$$

$$\Rightarrow Z = 0.9305$$

At 5% level, the tabulated value of Z_α is 1.96

Since $|Z| = 0.9305 < 1.96$

Hence, the null hypothesis H_0 is accepted at 5% level of significance and hence there is no significant difference.

6. Before an increase in excise duty on tea, 800 persons out of a sample of 1000 persons were found to be tea drinkers. After an increase in excise duty, 800 people were tea drinkers in a sample of 1200 people. Test whether there is a significant decrease in the consumption of tea after the increase in excise duty at 5% Los.

Sol.

Set the null hypothesis $H_0; P_1 = P_2$

Set the Alternative hypothesis $H_1; P_1 \neq P_2$

The level of significance $\alpha = 0.05$ (5%)

Given

$$n_1 = 1000, n_2 = 1200 \text{ \& } x_1 = 800, x_2 = 800$$

$$\therefore p_1 = \frac{800}{1000} = 0.8, p_2 = \frac{800}{1200} = 0.6670$$

$$\text{We know that } P = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\Rightarrow P = \frac{800 + 800}{1000 + 1200}$$

$$\Rightarrow P = \frac{1600}{2200}$$

$$\Rightarrow P = 0.7272 \Rightarrow Q = 1 - P = 1 - 0.7272 = 0.2728$$

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

$$\Rightarrow Z = \frac{0.8 - 0.6670}{\sqrt{(0.7272 \times 0.2728) \left(\frac{1}{1000} + \frac{1}{1200} \right)}}$$

$$\Rightarrow Z = \frac{0.133}{\sqrt{(0.1983)(0.00183)}}$$

$$\Rightarrow Z = \frac{0.133}{\sqrt{0.00036}}$$

$$\Rightarrow Z = \frac{0.133}{0.0189}$$

$$\Rightarrow Z = 7.037$$

At 5% level, the tabulated value of Z_α is 1.645.

Since $|Z| = 7.037 > 1.645$

Hence Null Hypothesis H_0 is rejected at 5% level of significance.

There is a significance decrease in the consumption of tea due to increase in excise duty.

7. In a sample of 600 men from a certain city, 450 are found smokers. In another sample of 900 men from another city, 450 are smokers. Do the indicate that the cities are significantly different with respect to the habit of smoking among men. Test at 5% significance level.

(Warning: Smoking is injurious to health, causes cancer, Tabaco causes painful death)

Sol.

Set the null hypothesis $H_0; P_1 = P_2$

Set the Alternative hypothesis $H_1; P_1 \neq P_2$

The level of significance $\alpha = 0.05$ (5%)

Given

$$n_1 = 600, n_2 = 900 \text{ \& } x_1 = 450, x_2 = 450$$

$$\therefore p_1 = \frac{450}{600} = 0.75, p_2 = \frac{450}{900} = 0.5$$

$$\text{We know that } P = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\Rightarrow P = \frac{450 + 450}{600 + 900}$$

$$\Rightarrow P = \frac{900}{1500} = 0.6$$

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

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

$$\Rightarrow Z = \frac{0.75 - 0.5}{\sqrt{(0.6 \times 0.4) \left(\frac{1}{600} + \frac{1}{900} \right)}}$$

$$\Rightarrow Z = \frac{0.25}{\sqrt{(0.24)(0.00277)}}$$

$$\Rightarrow Z = \frac{0.25}{\sqrt{0.0006648}}$$

$$\Rightarrow Z = \frac{0.25}{0.02578}$$

$$\Rightarrow Z = 9.69$$

At 5% level, the tabulated value of Z_α is 1.645.

Since $|Z| = 9.69 > 1.645$

Hence Null Hypothesis H_0 is rejected at 5% level of significance.

8. One type of air craft is found to develop engine trouble in 5 flights out of a total of 100 and another type in 7 flights out of a total of 200 flights. Is there a significance difference in the two types of air craft's so far as engine defects are concerned? Test at 5% significance level.

Sol.

Set the null hypothesis $H_0: P_1 = P_2$

Set the Alternative hypothesis $H_1: P_1 \neq P_2$

The level of significance $\alpha = 0.05$ (5%)

Given

$$n_1 = 100, n_2 = 200 \text{ \& } x_1 = 5, x_2 = 7$$

$$\therefore p_1 = \frac{5}{100} = 0.05, p_2 = \frac{7}{200} = 0.35$$

$$\text{We know that } P = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\Rightarrow P = \frac{5+7}{100+200}$$

$$\Rightarrow P = \frac{12}{300}$$

$$\Rightarrow P = 0.04 \Rightarrow Q = 1 - P = 1 - 0.04 = 0.96$$

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

$$\Rightarrow Z = \frac{0.05 - 0.35}{\sqrt{(0.04 \times 0.96)\left(\frac{1}{100} + \frac{1}{200}\right)}}$$

$$\Rightarrow Z = -\frac{0.3}{\sqrt{(0.384)(0.015)}}$$

$$\Rightarrow Z = -\frac{0.3}{\sqrt{0.00576}}$$

$$\Rightarrow Z = -\frac{0.3}{0.07589}$$

$$\Rightarrow Z = -3.953$$

At 5% level, the tabulated value of Z_α is 1.645.

Since $|Z| = 3.953 > 1.645$

Hence Null Hypothesis H_0 is rejected at 5% level of significance.

9. A machine produced 16 defective articles in a batch of 500. After overhauling it produced 3 defectives in a batch of 100. Has the machine improved?

Sol.

Set the null hypothesis $H_0: P_1 = P_2$

Set the Alternative hypothesis $H_1: P_1 \neq P_2$

The level of significance $\alpha = 0.01$ (1%)

Given

$$n_1 = 500, n_2 = 100 \text{ \& } x_1 = 16, x_2 = 3$$

$$\therefore p_1 = \frac{16}{500} = 0.032, p_2 = \frac{3}{100} = 0.03$$

$$\text{We know that } P = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\Rightarrow P = \frac{16+3}{500+100}$$

$$\Rightarrow P = \frac{19}{600}$$

$$\Rightarrow P = 0.03166 \Rightarrow Q = 1 - P = 1 - 0.03166 = 0.96834$$

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

$$\Rightarrow Z = \frac{0.032 - 0.03}{\sqrt{(0.03166 \times 0.96834) \left(\frac{1}{500} + \frac{1}{100} \right)}}$$

$$\Rightarrow Z = \frac{0.002}{\sqrt{(0.03065)(0.012)}}$$

$$\Rightarrow Z = \frac{0.002}{\sqrt{0.0003678}}$$

$$\Rightarrow Z = \frac{0.002}{0.01917}$$

$$\Rightarrow Z = 0.1047$$

At 1% level, the tabulated value of Z_α is 1.96.

Since $|Z| = 0.1047 < 1.96$

Hence Null Hypothesis H_0 is accepted at 1% level of significance

$\therefore$ the machine is not improved after overhauling.

10. A machine produced 25 defective articles in a batch of 400. After over hauling it produced 15 defectives in a batch of 200. Test at 1% level of significance whether there is a reduction of defective articles after overhauling.

Sol.

The null hypothesis $H_0; P_1 = P_2$

Set the Alternative hypothesis $H_1: P_1 \neq P_2$

The level of significance $\alpha = 0.01$ (1%)

Given

$$n_1 = 400, n_2 = 200 \text{ \& } x_1 = 25, x_2 = 15$$

$$\therefore p_1 = \frac{25}{400} = 0.0625, p_2 = \frac{15}{200} = 0.075$$

We know that $P = \frac{x_1 + x_2}{n_1 + n_2}$

$$\Rightarrow P = \frac{25 + 15}{400 + 200}$$

$$\Rightarrow P = \frac{40}{600}$$

$$\Rightarrow P = 0.0666 \Rightarrow Q = 1 - P = 1 - 0.0666 = 0.9334$$

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

$$\Rightarrow Z = \frac{0.0625 - 0.075}{\sqrt{(0.0666 \times 0.9334) \left(\frac{1}{400} + \frac{1}{200} \right)}}$$

$$\Rightarrow Z = -\frac{0.0125}{\sqrt{(0.0621)(0.0075)}}$$

$$\Rightarrow Z = -\frac{0.0125}{\sqrt{0.00046}}$$

$$\Rightarrow Z = -\frac{0.0125}{0.0214}$$

$$\Rightarrow Z = -0.5841$$

At 1% level, the tabulated value of Z_α is 1.96.

Since $|Z| = 0.5841 < 1.96$

Hence Null Hypothesis H_0 is accepted at 1% level of significance

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

Sol.

The level of significance $\alpha = 0.05$ (5%) $\Rightarrow Z_{0.05} = 1.96$

The level of significance $\alpha = 0.01$ (1%) $\Rightarrow Z_{0.01} = 1.64$

Given

$$n = 200$$

$$\sigma = 8$$

$$\mu = 74.5 \text{ and } \bar{x} = 75.9$$

We calculate Z through Test Statistic,

$$\begin{aligned}
 Z &= \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \\
 \Rightarrow Z &= \frac{75.9 - 74.5}{8 / \sqrt{200}} \\
 \Rightarrow Z &= \frac{1.4}{1.4} \\
 \Rightarrow Z &= \frac{8 / 14.1421}{1.4 \times 14.1421} \\
 \Rightarrow Z &= \frac{8}{19.799} \\
 \Rightarrow Z &= 2.4748
 \end{aligned}$$

i) Thus At 5% level, the tabulated value of Z_α is 1.645.

Since $|Z| = 2.4748 > 1.96$

Hence Null Hypothesis H_0 is rejected at 5% level of significance.

ii) Thus At 1% level, the tabulated value of Z_α is 1.645.

Since $|Z| = 2.4748 > 1.645$

Hence Null Hypothesis H_0 is rejected at 1% level of significance.

12. Intelligent tests were given to the two groups of boys and girls,

	Mean	S.D	Size
Girls	75	8	60
Boys	73	10	100

Find out if the two mean significantly differ at 5% level of significance.

Solⁿ:

Set The null hypothesis $H_0; P_1 = P_2$

Set the Alternative hypothesis $H_1; P_1 \neq P_2$

where, P1 refers the girls and P2 refers the boys

Given, the means, S.D's & sizes of both the groups of girls and boys are as follows,

$$\bar{x}_1 = 75, \bar{x}_2 = 73, \sigma_1 = 8, \sigma_2 = 10, n_1 = 60, n_2 = 100$$

$$\text{WKT, } Z = \frac{(\bar{x}_2 - \bar{x}_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(73 - 75)}{\sqrt{\frac{64}{60} + \frac{100}{100}}} = -\frac{2}{\sqrt{2.07}} = -1.3898.$$

Thus At 5% level, the tabulated value of Z_α is 1.96.

Since $|Z| = 1.3898 < 1.96$

Hence, the null hypothesis is accepted at 5% level of significance, i.e., there is no significant difference.

