Sampling Distribution

UNIT-II

Sampling Distribution:

Let $(x_1, x_2 \ldots x_n)$ be a r.s drawn from a popu, with dist., func, $F(x)$. The r.vs $x_1, x_2 \ldots x_n$ can be considered as indep, and identically distributed with the dist., func., $F(x)$. Let $z(x_1, x_2 \ldots x_n)$ be any fun., of these r.vs. Since $x_1, x_2 \ldots x_n$ are r.vs and $z(x_1, x_2 \ldots x_n)$ being a fun., of these r.vs is also a r.v. The dist., of $z(x_1, x_2 \ldots x_n)$ is called the sampling dist., of $z(x_1, x_2 \ldots x_n)$

A prob., dist., of all the possible means of the samples is known as sampling dist.,. If we select a no., of indep., random samples of a definite size from a given popu., and calculate some statistic (mean, S.D etc.,). From the sample we get a series of values of these statistic, These values obtained from the different samples can be put in the form of a freq., dist., . This dist., so formed of all possible values of a statistic is called the sampling dist., or probability sampling of that statistic.

Properties:

(i) AM of the sampling dist., is same as the mean of the universe (popu.) from which sample were taken.

(ie) AM of sampling dist., is equal to the popu., mean.

(ii) Sampling dist., of mean has a S.D (Std., error) equal to the popu., S.D divided by the square root of the sample size. ($\sigma/\sqrt{n}$).

Standard Error:

The S.D of the sampling dist., of a statistic is known as its std., error. SE is ~~equal~~ useful to determine within which the parameter values are expected to lie & S.E is used as an instrument in testing a given hypothesis.

Ex.,: $SE(\bar{x}) = \sigma/\sqrt{n}$

$S.E(s) = \sqrt{\sigma^2/2n}$

Unit-II continuations 3

4. TEST OF SIGNIFICANCE (Basic Concepts)

4.0 Introduction:

It is not easy to collect all the information about population and also it is not possible to study the characteristics of the entire population (finite or infinite) due to time factor, cost factor and other constraints. Thus we need sample. Sample is a finite subset of statistical individuals in a population and the number of individuals in a sample is called the sample size.

Sampling is quite often used in our day-to-day practical life. For example in a shop we assess the quality of rice, wheat or any other commodity by taking a handful of it from the bag and then to decide to purchase it or not.

4.1 Parameter and Statistic:

24 [The statistical constants of the population such as mean, (μ), variance (σ^2), correlation coefficient (ρ) and proportion (P) are called 'Parameters'.

Statistical constants computed from the samples corresponding to the parameters namely mean ($\bar{x}$), variance (S^2), sample correlation coefficient (r) and proportion (p) etc, are called statistic.] 24

Parameters are functions of the population values while statistic are functions of the sample observations. In general, population parameters are unknown and sample statistics are used as their estimates.

4.2 Sampling Distribution:

23 [The distribution of all possible values which can be assumed by some statistic measured from samples of same size 'n' randomly drawn from the same population of size N, is called as sampling distribution of the statistic] 23 (DANIEL and FERREL).

Consider a population with N values .Let us take a random sample of size n from this population, then there are

$$= \frac{\;}{K}\Sigma(t_i - t)^2$$

4

4.3 Standard Error:

25 The standard deviation of the sampling distribution of a statistic is known as its standard error. It is abbreviated as S.E. For 25 example, the standard deviation of the sampling distribution of the mean $\bar{x}$ known as the standard error of the mean,

$$\text{Where } v(\bar{x}) = v\left(\frac{x_1 + x_2 + \ldots\ldots\ldots\ldots x_n}{n}\right)$$

$$= \frac{v(x_1)}{n^2} + \frac{v(x_2)}{n^2} + \ldots\ldots + \frac{v(x_n)}{n^2}$$

$$= \frac{\sigma^2}{n^2} + \frac{\sigma^2}{n^2} + \ldots\ldots + \frac{\sigma^2}{n^2} = \frac{n\sigma^2}{n^2}$$

$\therefore$ The S.E. of the mean is $\dfrac{\sigma}{\sqrt{n}}$

5

The standard errors of the some of the well known statistic for large samples are given below, where n is the sample size, σ^2 is the population variance and P is the population proportion and $Q = 1-P$. n_1 and n_2 represent the sizes of two independent random samples respectively.

Sl.No	Statistic	Standard Error
1.	Sample mean $\bar{x}$	$\frac{\sigma}{\sqrt{n}}$
2.	Observed sample proportion p	$\sqrt{\frac{PQ}{n}}$
3.	Difference between of two samples means $(\bar{x}_1 - \bar{x}_2)$	$\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
4.	Difference of two sample proportions $p_1 - p_2$	$\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}}$

Uses of standard error

i) Standard error plays a very important role in the large sample theory and forms the basis of the testing of hypothesis.

ii) The magnitude of the S.E gives an index of the precision of the estimate of the parameter.

iii) The reciprocal of the S.E is taken as the measure of reliability or precision of the sample.

iv) S.E enables us to determine the probable limits within which the population parameter may be expected to lie.

Remark:

S.E of a statistic may be reduced by increasing the sample size but this results in corresponding increase in cost, labour and time etc.,

4.4 Null Hypothesis and Alternative Hypothesis

Hypothesis testing begins with an assumption called a Hypothesis, that we make about a population parameter. A hypothesis is a supposition made as a basis for reasoning. The conventional approach to hypothesis testing is not to construct a

6

single hypothesis about the population parameter but rather to set up two different hypothesis. So that of one hypothesis is accepted, the other is rejected and vice versa.

Null Hypothesis:

A hypothesis of no difference is called null hypothesis and is usually denoted by H_0. "Null hypothesis is the hypothesis" which is tested for possible rejection under the assumption that it is true " by Prof. R.A. Fisher. It is very useful tool in test of significance. For example: If we want to find out whether the special classes (for Hr. Sec. Students) after school hours has benefited the students or not. We shall set up a null hypothesis that "H_0: special classes after school hours has not benefited the students".

Alternative Hypothesis:

Any hypothesis, which is complementary to the null hypothesis, is called an alternative hypothesis, usually denoted by H_1. For example, if we want to test the null hypothesis that the population has a specified mean μ_0 (say).

i.e., Step 1: null hypothesis H_0: $\mu = \mu_0$

then 2. Alternative hypothesis may be

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

ii) $H_1 : \mu > \mu_0$

iii) $H_1 : \mu < \mu_0$

the alternative hypothesis in (i) is known as a two – tailed alternative and the alternative in (ii) is known as right-tailed (iii) is known as left –tailed alternative respectively. The settings of alternative hypothesis is very important since it enables us to decide whether we have to use a single – tailed (right or left) or two tailed test.

4.5 Level of significance and Critical value:

Level of significance:

In testing a given hypothesis, the maximum probability with which we would be willing to take risk is called level of significance of the test. This probability often denoted by "α" is generally specified before samples are drawn.

7 Type-I and Type-II Errors

Obviously, the first two possibilities lead to errors.

In a statistical hypothesis testing experiment, a Type I error is committed by rejecting the null hypothesis when it is true. On the other hand, a Type II error is committed by not rejecting (accepting) the null hypothesis when it is false.

If we write ,

$$\alpha = P\,(\text{Type I error}) = P\,(\text{rejecting } H_0 \mid H_0 \text{ is true})$$

$$\beta = P\,(\text{Type II error}) = P\,(\text{Not rejecting } H_0 \mid H_0 \text{ is false})$$

In practice, type I error amounts to rejecting a lot when it is good and type II error may be regarded as accepting the lot when it is bad. Thus we find ourselves in the situation which is described in the following table.

	Accept H_0	Reject H_0
H_0 is true	Correct decision	Type I Error
H_0 is false	Type II error	Correct decision

5.1 Large samples (n > 30):

8

The tests of significance used for problems of large samples are different from those used in case of small samples as the assumptions used in both cases are different. The following assumptions are made for problems dealing with large samples:

(i) Almost all the sampling distributions follow normal asymptotically.

(ii) The sample values are approximately close to the population values.

The following tests are discussed in large sample tests.

(i) Test of significance for proportion

(ii) Test of significance for difference between two proportions

(iii) Test of significance for mean
(iv) Test of significance for difference between two means.

5.2 Test of Significance for Proportion:

Test Procedure

Set up the null and alternative hypotheses

$H_0 : P = P_0$

$H_1 = P \neq P_0$ $(P > P_0$ or $P < P_0)$

Level of significance:

Let $\alpha = 0.05$ or 0.01

Calculation of statistic:

Under H_0 the test statistic is

$$Z_0 = \left| \frac{p - P}{\sqrt{\frac{PQ}{n}}} \right|$$

Expected value:

$$Z_e = \left| \frac{p - P}{\sqrt{\frac{PQ}{n}}} \right| \sim N(0, 1)$$

$$= 1.96 \quad \text{for } \alpha = 0.05 \ (1.645)$$
$$= 2.58 \quad \text{for } \alpha = 0.01 \quad (2.33)$$

Inference:

(i) If the computed value of $Z_0 \leq Z_e$ we accept the null hypothesis and conclude that the sample is drawn from the population with proportion of success P_0

(ii) If $Z_0 > Z_e$ we reject the null hypothesis and conclude that the sample has not been taken from the population whose population proportion of success is P_0.

Example 1:

In a random sample of 400 persons from a large population 120 are females.Can it be said that males and females are in the ratio 5:3 in the population? Use 1% level of significance

Solution:

We are given

$n = 400$ and

x = No. of female in the sample = 120

p = observed proportion of females in the sample = $\frac{120}{400} = 0.30$

Null hypothesis:

The males and females in the population are in the ratio 5:3

i.e., H_0: P = Proportion of females in the population = $\frac{3}{8} = 0.375$

Alternative Hypothesis: $Q = \frac{5}{8} = 0.625$

$H_1 : P \neq 0.375$ (two-tailed)

Level of significance:

$\alpha = 1\%$ or 0.01

Calculation of statistic:

Under H_0, the test statistic is

$$Z_0 = \left| \frac{p - P}{\sqrt{\frac{PQ}{n}}} \right|$$

$$= \left| \frac{0.300 - 0.375}{\sqrt{\frac{0.375 \times 0.625}{400}}} \right|$$

$$= \frac{0.075}{\sqrt{0.000586}} = \frac{0.075}{0.024} = 3.125$$

Expected value:

$$Z_e = \left| \frac{p - P}{\sqrt{\frac{PQ}{n}}} \right| \sim N(0,1) = 2.58$$

Inference :

Since the calculated $Z_0 > Z_e$ we reject our null hypothesis at 1% level of significance and conclude that the males and females in the population are not in the ratio 5:3

Example 2:

In a sample of 400 parts manufactured by a factory, the number of defective parts was found to be 30. The company, however, claimed that only 5% of their product is defective. Is the claim tenable? ratio = 1:1

Solution:

We are given

$n = 400$

x = No. of defectives in the sample = 30

p = proportion of defectives in the sample

$$= \frac{x}{n} = \frac{30}{400} = 0.075$$

Null hypothesis:

The claim of the company is tenable H_0: $P = 0.05$

Alternative Hypothesis:

$H_1 : P > 0.05$ (Right tailed Alternative)

Level of significance: 5%

Calculation of statistic:

Under H_0, the test statistic is

$$Z_0 = \left| \frac{p - P}{\sqrt{\frac{PQ}{n}}} \right|$$

$$= \left| \frac{0.075 - 0.050}{\sqrt{\frac{0.05 \times 0.95}{400}}} \right|$$

5/100 (5% defective)

assume 5% defects means 95% non-

$$= \frac{0.025}{\sqrt{0.0001187}} = 2.27$$

11A

Expected value:

$$Z_e = \left| \frac{p - P}{\sqrt{\frac{PQ}{n}}} \right| \sim N(0, 1)$$

$$= 1.645 \text{ (Single tailed)}$$

since we assumed 95% one value, so it is single tailed

Inference :

Since the calculated $Z_0 > Z_e$ we reject our null hypothesis at 5% level of significance and we conclude that the company's claim is not tenable.

5.3 Test of significance for difference between two proportion:

Test Procedure

Set up the null and alternative hypotheses:

$H_0 : P_1 = P_2 = P$ (say)

$H_1 : P_1 \neq P_2$ ($P_1 > P_2$ or $P_1 < P_2$)

Level of significance:

Let $\alpha = 0.05$ or 0.01

Calculation of statistic:

Under H_0, the test statistic is

$$Z_0 = \left| \frac{p_1 - p_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}} \right| \quad (P_1 \text{ and } P_2 \text{ are known})$$

$$= \left| \frac{p_1 - p_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| \quad (P_1 \text{ and } P_2 \text{ are not known})$$

where $\hat{P} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$

$\hat{Q} = 1 - \hat{P}$

Expected value:

$$Z_e = \left|\frac{p_1 - p_2}{S.E(p_1 - p_2)}\right| \sim N(0,1)$$

Inference:

(i) If $Z_0 \leq Z_e$ we accept the null hypothesis and conclude that the difference between proportions are due to sampling fluctuations.

(ii) If $Z_0 > Z_e$ we reject the null hypothesis and conclude that the difference between proportions cannot be due to sampling fluctuations

Example 3:

In a referendum submitted to the 'student body' at a university, 850 men and 550 women voted. 530 of the men and 310 of the women voted 'yes'. Does this indicate a significant difference of the opinion on the matter between men and women students?

Solution:

We are given

$n_1 = 850$ $\quad n_2 = 550$ $\quad x_1 = 530$ $\quad x_2 = 310$

$p_1 = \frac{530}{850} = 0.62$ $\quad p_2 = \frac{310}{550} = 0.56$

$\hat{P} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{530+310}{1400} = 0.60$

$\hat{Q} = 0.40$ $(1-\hat{p}) = (1-0.60)$

Null hypothesis:

H_0: $P_1 = P_2$ ie the data does not indicate a significant difference of the opinion on the matter between men and women students.

Alternative Hypothesis:

$H_1 : P_1 \neq P_2$ (Two tailed Alternative)

Level of significance:

Let $\alpha = 0.05$

13

Calculation of statistic:

Under H_0, the test statistic is

$$Z_0 = \frac{p_1 - p_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{0.62 - 0.56}{\sqrt{0.6 \times 0.4\left(\frac{1}{850} + \frac{1}{550}\right)}}$$

$$= \frac{0.06}{0.027} = 2.22$$

Expected value:

$$Z_e = \frac{p_1 - p_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1) = 1.96$$

Inference :

Since $Z_0 > Z_e$ we reject our null hypothesis at 5% level of significance and say that the data indicate a significant difference of the opinion on the matter between men and women students.

Example 4:

In a certain city 125 men in a sample of 500 are found to be self employed. In another city, the number of self employed are 375 in a random sample of 1000. Does this indicate that there is a greater population of self employed in the second city than in the first?

Solution:

We are given

$n_1 = 500$ $n_2 = 1000$ $x_1 = 125$ $x_2 = 375$

$$p_1 = \frac{125}{500} = 0.25 \qquad p_2 = \frac{375}{1000} = 0.375$$

14

$$\hat{P} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{125 + 375}{500 + 1000}$$

$$= \frac{500}{1500} = \frac{1}{3}$$

$$\hat{Q} = 1 - \frac{1}{3} = \frac{2}{3}$$

Null hypothesis:

H_0: $P_1 = P_2$ There is no significant difference between the two population proportions.

Alternative Hypothesis:

H_1 : $P_1 < P_2$ (left tailed Alternative)

Level of significance: Let $\alpha = 0.05$

Calculation of statistic:

Under H_0, the test statistic is

$$Z_0 = \left| \frac{p_1 - p_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right|$$

$$= \left| \frac{0.25 - 0.375}{\sqrt{\frac{1}{3} \times \frac{2}{3}\left(\frac{1}{500} + \frac{1}{1000}\right)}} \right| = \frac{0.125}{0.026} = 4.8$$

Expected value:

$$Z_e = \left| \frac{p_1 - p_2}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| \sim N(0,1) = 1.645$$

Inference :

Since $Z_0 > Z_e$ we reject the null hypothesis at 5% level of significance and say that there is a significant difference between the two population proportions.

Example 5:

A civil service examination was given to 200 people. On the basis of their total scores, they were divided into the upper 30% and the remaining 70%. On a certain question 40 of the upper group and 80 of the lower group answered correctly. On the basis of this question, is this question likely to be useful for discriminating the ability of the type being tested?

Solution:

We are given

$$n_1 = \frac{30 \times 200}{100} = 60 \qquad n_2 = \frac{70 \times 200}{100} = 140$$

$$x_1 = 40 \qquad x_2 = 80$$

$$p_1 = \frac{40}{60} = \frac{2}{3} \qquad p_2 = \frac{80}{140} = \frac{4}{7}$$

$$\hat{P} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{40+80}{60+140}$$

$$= \frac{120}{200} = \frac{6}{10}$$

$$\hat{Q} = 1 - \hat{P} = 1 - \frac{1}{6} = \frac{4}{10}$$

Null hypothesis:

H_0: $P_1 = P_2$ (say) The particular question does not discriminate the abilities of two groups.

Alternative Hypothesis:

H_1 : $P_1 \neq P_2$ (two tailed Alternative)

Level of significance:

Let $\alpha = 0.05$

Calculation of statistics

Under H_0, the test statistic is

5.4 Test of significance for mean: 16

Let x_i (i = 1,2.....n) be a random sample of size n from a population with variance σ^2, then the sample mean $\bar{x}$ is given by

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots\ldots x_n)$$

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

$$V(\bar{x}) = V\left[\frac{1}{n}(x_1 + x_2 + \ldots\ldots x_n)\right]$$

17

$$= \frac{1}{n^2}[(V(x_1)+V(x_2)+\ldots..V(x_n)]$$

$$= \frac{1}{n^2}\, n\sigma^2 = \frac{\sigma^2}{n}$$

$$\therefore \text{ S.E}(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

Test Procedure:

Null and Alternative Hypotheses:

$H_0: \mu = \mu_0$.

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

Level of significance:

Let $\alpha = 0.05$ or 0.01

Calculation of statistic:

Under H_0, the test statistic is

$$Z_0 = \left|\frac{\bar{x}-E(\bar{x})}{S.E(\bar{x})}\right| = \left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right|$$

Expected value:

$$Z_e = \left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right| \sim N(0,1)$$

$= 1.96$ for $\alpha = 0.05$ (1.645)

or

$= 2.58$ for $\alpha = 0.01$ (2.33)

Inference :

If $Z_0 \leq Z_e$ we accept our null hypothesis and conclude that the sample is drawn from a population with mean $\mu = \mu_0$

If $Z_0 > Z_e$ we reject our H_0 and conclude that the sample is not drawn from a population with mean $\mu = \mu_0$

Example 6:

The mean lifetime of 100 fluorescent light bulbs produced by a company is computed to be 1570 hours with a standard deviation of 120 hours. If μ is the mean lifetime of all the bulbs produced by the company, test the hypothesis μ=1600 hours against the alternative hypothesis $\mu \neq 1600$ hours using a 5% level of significance.

Solution:

We are given

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

Null hypothesis:

H_0: $\mu = 1600$.ie There is no significant difference between the sample mean and population mean.

Alternative Hypothesis:

H_1: $\mu \neq 1600$ (two tailed Alternative)

Level of significance:

Let $\alpha = 0.05$

Calculation of statistics

Under H_0, the test statistic is

$$Z_0 = \left|\frac{\bar{x} - \mu}{s/\sqrt{n}}\right|$$

$$= \left|\frac{1570 - 1600}{\frac{120}{\sqrt{100}}}\right|$$

$$= \frac{30 \times 10}{120}$$

$$= 2.5$$

Expected value:

$$Z_0 = \left|\frac{\bar{x} - \mu}{s/\sqrt{n}}\right| \sim N(0,1)$$

$$= 1.96 \text{ for } \alpha = 0.05$$

Inference :

Since $Z_0 > Z_e$ we reject our null hypothesis at 5% level of significance and say that there is significant difference between the sample mean and the population mean.

Example 7:

A car company decided to introduce a new car whose mean petrol consumption is claimed to be lower than that of the existing car. A sample of 50 new cars were taken and tested for petrol consumption. It was found that mean petrol consumption for the 50 cars was 30 km per litre with a standard deviation of 3.5 km per litre. Test at 5% level of significance whether the company's claim that the new car petrol consumption is 28 km per litre on the average is acceptable.

Solution:

We are given $\bar{x} = 30$; $\mu = 28$; $n = 50$; $s = 3.5$

Null hypothesis:

H_0: $\mu = 28$. i.e The company's claim that the petrol consumption of new car is 28km per litre on the average is acceptable.

Alternative Hypothesis:

H_1: $\mu < 28$ (Left tailed Alternative)

Level of significance:

Let $\alpha = 0.05$

Calculation of statistic:

Under H_0 the test statistics is

$$Z_0 = \left|\frac{\bar{x} - \mu}{s/\sqrt{n}}\right|$$

$$= \left|\frac{30 - 28}{\frac{3.5}{\sqrt{50}}}\right|$$

$$= \frac{2 \times \sqrt{50}}{3.5}$$

$$= 4.04$$

Expected value:

$$Z_e = \left|\frac{\bar{x} - \mu}{s/\sqrt{n}}\right| \sim N(0,1) \text{ at } \alpha = 0.05$$

$$= 1.645$$

Inference : 20 Evangeline.

Since the calculated $Z_0 > Z_e$ we reject the null hypothesis at 5% level of significance and conclude that the company's claim is not acceptable.

5.5 Test of significance for difference between two means:

Test procedure

Set up the null and alternative hypothesis

H_0: $\mu_1 = \mu_2$; **H_1:** $\mu_1 \neq \mu_2$ ($\mu_1 > \mu_2$ or $\mu_1 < \mu_2$)

Level of significance:

Let $\alpha\%$

Calculation of statistic:

Under H_0 the test statistic is

$$Z_0 = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right|$$

If $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (ie) If the samples have been drawn from the population with common S.D σ then under $H_0 : \mu_1 = \mu_2$

$$Z_0 = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right|$$

Expected value:

$$Z_e = \left| \frac{\bar{x}_1 - \bar{x}_2}{S.E(\bar{x}_1 - \bar{x}_2)} \right| \sim N(0,1)$$

Inference:

(i) If $Z_0 \leq Z_e$ we accept the H_0 (ii) If $Z_0 > Z_e$ we reject the H_0

Example 8:

A test of the breaking strengths of two different types of cables was conducted using samples of $n_1 = n_2 = 100$ pieces of each type of cable.

2.1

Cable I	Cable II
$\bar{x}_1 = 1925$	$\bar{x}_2 = 1905$
$\sigma_1 = 40$	$\sigma_2 = 30$

Do the data provide sufficient evidence to indicate a difference between the mean breaking strengths of the two cables? Use 0.01 level of significance.

Solution:

We are given

$\bar{x}_1 = 1925$ $\bar{x}_2 = 1905$ $\sigma_1 = 40$ $\sigma_2 = 30$

Null hypothesis

$H_0: \mu_1 = \mu_2$ .ie There is no significant difference between the breaking strengths of the two cables.

$H_1 : \mu_1 \neq \mu_2$ (Two tailed alternative)

Level of significance:

Let $\alpha = 0.01$ or 1%

Calculation of statistic:

Under H_0 the test statistic is

$$Z_0 = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right|$$

$$= \left| \frac{1925 - 1905}{\sqrt{\frac{40^2}{\cancel{190}} + \frac{30^2}{100}}} \right| = \frac{20}{5} = 4$$

100

Expected value:

$$Z_e = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right| \sim N(0,1) = 2.58$$

Inference: 22

Since $Z_0 > Z_e$, we reject the H_0 Hence the formulated null hypothesis is wrong ie there is a significant difference in the breaking strengths of two cables.

Example 9:

The means of two large samples of 1000 and 2000 items are 67.5 cms and 68.0cms respectively. Can the samples be regarded as drawn from the population with standard deviation 2.5 cms. Test at 5% level of significance.

Solution:

We are given

$n_2 = 1000$; $n_2 = 2000$ $\bar{x}_1 = 67.5$ cms ; $\bar{x}_2 = 68.0$ cms $\sigma = 2.5$ cms

Null hypothesis

H_0: $\mu_1 = \mu_2$ (i.e.,) the sample have been drawn from the same population.

Alternative Hypothesis:

H_1: $\mu_1 \neq \mu_2$ (Two tailed alternative)

Level of significance:

$\alpha = 5\%$

Calculation of statistic:

Under H_e the test statistic is

$$Z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{67.5 - 68.0}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}}$$

$$= \frac{0.5 \times 20}{2.5\sqrt{3/5}}$$

$$= 5.1$$

Expected value: 2.3.

$$Z_e = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \sim N(0,1) = 1.96$$

Inference :

Since $Z_0 > Z_e$ we reject the H_0 at 5% level of significance and conclude that the samples have not come from the same population.