Large Sample

Population or Universe: An aggregate of objects under study is called population or universe. It is thus a collection of individuals or of their attributes (qualities) or of results of operations which can be numerically specified.

Sample: Sample is a small portion of the population.

The number of individuals in a sample is called sample size.

The process of selecting a sample from a universe is called sampling.

Parameters of Statistics:- (Mean, variance etc.)

Population mean & variance are denoted by μ & σ^2, while those of the sample are given by $\bar{x}$, s^2.

Test of Significance

1. The deviation b/w the observed sample statistic & the hypothetical parameter value or
2. The deviation b/w two sample statistic is significant or might be attributed due to chance or the fluctuations of the sampling.

Before applying tests of significance, we first set up hypothesis.

Hypothesis

Null hypothesis (H_0)
which is a definite statement about the population parameter

(H_1) Alternative hypothesis
Complementary to the null hypothesis.

E.g. If we want to test the null hypothesis that the population has a specified mean μ_0, then

$$H_0 : \mu = \mu_0$$

Alternative hypothesis will be:

(i) $H_1 : \mu \neq \mu_0$ (two tailed alternative hypothesis)
(ii) $H_1 : \mu > \mu_0$ (right tailed alternative)
(iii) $H_1 : \mu < \mu_0$ (left tailed alternative).

Level of significance:-

The prob. of the value of the variate falling in the critical region is known as level of significance.

critical region → A region corresponding to a statistic available in the sample space S which amounts to rejection of the null hypothesis.

Types

① 1% $(|z|)=2.58$ → 2 tailed
$|z|=2.33$ → Right tailed.

② 5% $|z|=1.966$ → 2-tailed
$|z|=1.645$ Right tailed.

Test of Significance for Large Samples

If the sample size $n>30$, the sample is taken as large sample. [Binomial, normal distn are closely appr. to normal distn]

① Testing of significance for single proportion.
② " " " for difference of proportion.
③ " " " for single mean.
④ " " " for difference of means.
⑤ " " " " the difference of S.D.

* Here mean is zero & variance 1 for $z = \dfrac{X-E(X)}{S.E(X)}$.

Testing of Significance for Single Proportion:

This test is used to find the significant difference b/w proportion of the sample & the population.

Let X be the no. of successes in n independent trails with constant

$$E(X) = nP \; ; \; V(X) = nPQ \; ; \; Q = 1-P.$$

Let $p = \frac{X}{n}$ called the observed proportion of success.

$$E(p) = E(X/n) = \frac{1}{n}E(X) = \frac{nP}{n} = P \; ; \; E(p) = P.$$

$$V(p) = V(X/n) = \frac{1}{n^2}V(X) = \frac{nPQ}{n^2} = \frac{PQ}{n}$$

$$S.E(p) = \sqrt{\frac{PQ}{n}} \; ; \; Z = \frac{p - E(p)}{S.E(p)} = \frac{p-P}{\sqrt{PQ/n}} \sim N(0,1)$$

Ques 1) A coin was tossed 400 times & head turns up 216 times. Test the hypothesis that the coin is unbiased.

Sol^n H_0: The coin is unbiased i.e $P = 0.5$

H_1: " " " not unbiased i.e $P \neq 0.5$

Here $n = 400$; X = no. of success = 216

p = proportion of success in the sample $\frac{X}{n} = \frac{216}{400} = 0.54$

population proportion $= 0.5 = P$, $Q = 1-P = 0.5$

under H_0, test of statistic $z = \frac{p-P}{\sqrt{PQ/n}}$

$$|Z| = \frac{p-P}{\sqrt{PQ/n}} = \left| \frac{0.54-0.5}{\sqrt{\frac{0.5\times0.5}{100}}} \right| = 1.6$$

we use two tailed test

Conclusion: Since $|Z| = 1.6 < 1.96$

i.e The value 1.96 is the significant value of z at 5% level of significance.

i.e the coin is unbiased in $P = 0.5$

Ques 2) A certain cubical die was thrown 9000 times & 5 or 6 was obtained 3240 times. On the assumption of certain throwing, do the data indicate an unbiased die?

Sol^n Here $n = 9000$

P = prob of success (i.e. getting 5 or 6 in die)

$P = 2/6 = \frac{1}{3}$, $Q = 2/3$

$$p = \frac{X}{n} = \frac{3240}{9000} = 0.36$$

H_0: is unbiased die i.e $P = 1/3$

H_1: $P \neq 1/3$ (two tailed test)

The test statistic $z = \frac{p-P}{\sqrt{PQ/n}} = \frac{0.36-0.33}{\sqrt{1/3 \times 2/3 \times \frac{1}{9000}}} = 6.09888$

$|Z| = \cancel{0.03496 < 1.96} \rightarrow \boxed{6.0988 > 1.96}$?

Conclusion: H_0 is rejected as 1.96 is the significant value of z at 5% level of significant.

i.e. die is not unbiased.

Scanned by CamScanner

Q37] A manufacturer claims that only 4% of his products supplied by him are defective. A random sample of 600 products contained 36 defectives. Test the claim of the manufacturer.

Soln $n = 600$, X = no. of defective products $= 36$

P = proportion of defectives in the population $= 0.04$

p = " " " in the sample $= \frac{36}{600} = 0.06$

H_0: claim of manufacturer i.e $P = 0.04$ is true.

i.e. claim of manufacturer is accepted.

H_1: (i) $P \neq 0.04$ (two tailed test)

(ii) If we want to reject, only if $p > 0.04$ then (right tailed).

Under H_0, $$z = \frac{p - P}{\sqrt{PQ/n}} = \frac{0.06 - 0.04}{\sqrt{\frac{0.04 \times 0.96}{600}}} = 2.5$$

Conclusion Since $|z| = 2.5 > 1.96$, we reject the hypothesis H_0 at 5% level of significance two tailed.

If H_1 is taken as $p > 0.04$ we ~~reject~~ ~~null~~ apply right tailed test

$|z| = 2.5 > 1.645$, we reject the null hypothesis

In both cases, manufacturer's claim is not acceptable.

Q4] A bag contains defective article, the exact number of which is not known. A sample of 100 from the bag gives 10 defective articles. Find the limits for the proportion of defective articles in the bag.

Soln Here p = proportion of def. article $= \frac{10}{100} = 0{\cdot}01$

$q = 0{\cdot}9$

Since the confidence limits is not given, we assume it as 95%.

$\therefore$ level of significance is 5% ie $Z_\alpha = 1{\cdot}96$

Also the proportion of population P is not given. To get the confidence limit, we use p & it is given by $p \pm Z_\alpha \sqrt{\frac{pq}{n}} = 0{\cdot}1 \pm 1{\cdot}96 \sqrt{\frac{0{\cdot}1 \times 0{\cdot}9}{100}}$

$= 0{\cdot}1 \pm 0{\cdot}0588 = 0{\cdot}1588,\ 0{\cdot}0412$

Hence 95% confidence limits for defective ~~limits~~ article in the bag are $(0{\cdot}1588, 0{\cdot}0412)$

Testing of Significance ~~test~~ for difference of Proportions

Consider two samples X_1 & X_2 of sizes n_1 & n_2 resp. taken from two different populations. To test the significance of the difference b/w the proportion p_1 & p_2. The test statistic under the null hypothesis H_0, that there is no significant difference b/w the two sample proportion, we have

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

Ques 1]: Before an increase in excise duty on tea, 800 people out of a sample of 1000 persons were found to be tea drinkers. After an increase in the excise duty, 800 persons were known to be tea drinkers in a sample of 1200 persons. Do you think that there has been a significant decrease in the consumption of tea after the increase in the excise duty.

Soln Here $n_1 = 1000$, $n_2 = 1200$

$$p_1 = \frac{X_1}{n_1} = \frac{800}{1000} = \frac{4}{5} \; ; \; p_2 = \frac{X_2}{n_2} = \frac{800}{1200} = \frac{2}{3}$$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1+n_2} = \frac{X_1+X_2}{n_1+n_2} = \frac{800+800}{1000+1200} = \frac{8}{11}$$

$Q = \frac{3}{11}$

Null Hypothesis $H_0 : \boxed{p_1 = p_2}$ ie. there is no significant difference in the consumption of tea before & after increase of excise duty.

$H_1 : p_1 > p_2$ (right tailed test).

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.8 - 0.666}{\sqrt{\frac{8}{11}\times\frac{3}{11}\left(\frac{1}{1000}+\frac{1}{1200}\right)}} = 6.842$$

Conclusion: Since the calculated value of $|Z| > 1.645$ & also $|Z| > 2.33$, both the significant value of z at 5% & 1% level of significance. Hence Ho is rejected. ie. there is a ~~no~~ significant decrease in the consumption of tea due to increase in excise duty.

Ques 2 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?

Solⁿ: Here $n_1 = 500$, $n_2 = 100$

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

Null Hypothesis H_0: $p_1 = p_2$ i.e. The machine has not improved due to overhauling

H_1 : $P_1 > P_2$ (right tailed)

$$P = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = \frac{16+3}{500+100} = \frac{19}{600} = 0.032 \text{ (app.)}$$

The test statistic $$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.032 - 0.03}{\sqrt{(0.032)(0.968)\left(\frac{1}{500}+\frac{1}{100}\right)}}$$

$$\Rightarrow |Z| = \cancel{6.66 \text{ (app.)}}\ 0.117 \text{ (app.)}$$

<u>Conclusion</u>:- The calculated value of $|Z| < 1.645$, the significant value of Z at 5% level of significance. H_0 is ~~rejected~~ accepted, i.e. the machine has ~~not~~ not improved due to overhauling.

<u>Ques 3</u>] In two large populations there are 30% & 25% resp. of fair haired people. Is this difference likely to be hidden in samples of 1200 & 900 resp. from the two populations.

<u>Sol^n</u> p_1 = proportion of fair haired people in the first population = 30% = 0.3

$p_2 = 25\% = 0.25$

$H_0: p_1 = p_2$ i.e. the difference in population proportions is likely to be hidden in sampling

$H_1: p_1 \neq p_2$; $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{0.3 \times 1200 + 0.25 \times 900}{1200 + 900}$

$= \frac{360 + 225}{2100} = 0.2786$ (app.)

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

$$= \frac{0.3 - 0.25}{\sqrt{(0.2786)(0.7214)\left(\frac{1}{1200} + \frac{1}{900}\right)}} = \frac{0.05}{\sqrt{0.000391}} = \frac{0.05}{0.01977} = 2.529$$

<u>Conclusion</u>: since the calculated value of $|Z| > 1.96$ the significant value of Z at 5% level of significance. Ho is ~~accepted~~ rejected.

However $|Z| < 2.58$, the significant value of Z at 1% ~~of~~ level of significance. Ho is accepted.

$\therefore$ At 5% level of significance, these samples will reveal the difference in the population proportions.

Scanned by CamScanner

Test of Significance for Single Mean

This test is used to test whether the ~~level~~ difference b/w sample mean & population mean is significant or not.

$\bar{x}$ & σ are mean & S.D. of population whereas μ & s are mean & S.D of sample.

n is sample size.

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

If σ is not known, we use the test statistic

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

Ques 1] A normal population has a mean of 6.8 & S.D. of 1.5. A sample of 400 members gave a mean 6.75. Is this difference significant?

Sol^n^ $H_0: \bar{x} = \mu$

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

Given $\mu = 6.8$, $\sigma = 1.5$, $\bar{x} = 6.75$ & $n = 400$

$$|Z| = \left|\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right| = \left|\frac{6.75 - 6.8}{1.5/\sqrt{400}}\right| = |-0.67| = 0.67$$

Conclusion:- As the calculated value of $|Z| < 1.96$, at 5% level of significance, H_0 is accepted, i.e. ~~the sample is drawn from the population~~ there is no significant difference between $\bar{x}$ & μ.

Ques27:- A random sample of 900 members has a mean 3.4 cms. Can it be reasonably regarded as a sample from a large population of mean 3.2 cms & S.D 2.3 cms?

Soln: Here $n = 900$, $\bar{x} = 3.4$, $\mu = 3.2$, $\sigma = 2.3$

H_0: Assume that the sample is drawn from a large population with mean 3.2 & S.D = 2.3

H_1: $\mu \neq 3.25$ (Apply two tailed test)

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{3.4 - 3.2}{2.3 / \sqrt{900}} = 2.61$$

Conclusion:- As the calculated value of $|Z| > 1.96$ the significant value of Z at 5% level of significance. H_0 is ~~accepted~~ rejected. i.e., the sample is not drawn from the population with mean 3.2 & S.D = 2.3

Ques 37. The mean weight obtained from a random sample of size 100 is 64 gms. The S.D of the weight distribution of the population is 3 gms. Test the statement that the mean weight of the population is 67 gms at 5% level of significance. Also set up 99% confidence limits of the mean weight of the population.

Sol^n $n = 100$, $\mu = 67$, $\bar{x} = 64$, $\sigma = 3$.

H_0: $\mu = \bar{x}$ i.e no significance diff. b/w sample & population mean i.e $\mu = 67$

H_1: $\mu \neq 67$ [i.e. Two tailed test].

$$|Z| = \left|\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right| = \left|\frac{64 - 67}{3/\sqrt{100}}\right| = |-10| = 10$$

Conclusion: Since the calculated value of $|Z| > 1.96$, the significant value of z at 5% level of significance. Ho is rejected i.e. the sample is not drawn from the population with mean 67 & S.D = 3

To find 99% confidence limit :- It is given by

$$\bar{x} \pm 2.58\ \sigma/\sqrt{n} = 64 \pm 2.58 \times 3/\sqrt{100} =$$
$$= 64.774,\ 63.226$$

Test of Significance for difference of Means

Let $\bar{x}_1$ & $\bar{x}_2$ are the means of a sample size n_1 & n_2 resp. from two different populations with mean μ_1 & μ_2 and variance σ_1^2 & σ_2^2 resp.

The test statistic; $$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

* If σ_1 & σ_2 are not known & $\sigma_1 \neq \sigma_2$, the test statistic in this case, $$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Ques 17 Intelligent tests were given to two groups of boys & girls.

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

Examine if the difference b/w mean score is significant.

Sol^n Here $n_1 = 60$, $n_2 = 100$, $\bar{x}_1 = 75$, $\bar{x}_2 = 73$, $s_1 = 8$, $s_2 = 10$.

$H_0: \bar{x}_1 = \bar{x}_2$ ie no significant difference b/w mean scores.

$H_1: \bar{x}_1 \neq \bar{x}_2$. [two tailed test]

$$|Z| = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{75 - 73}{\sqrt{\frac{8^2}{60} + \frac{10^2}{100}}} = 1.3912$$

Conclusion: As the calculated value of $|Z| < 1.96$, the significant value of Z at 5% level of significance H_0 is accepted i.e. there is no significant difference b/w mean scores.

Q2: The average income of a persons was ₹210 with S.D of ₹10 in sample of 100 people of a city. For another sample of 150 persons, the average incomes was ₹220 with S.D of ₹12. The S.D of incomes of the people of the city was ₹11. Test whether there is any significant difference b/w the average income of the localities.

Soln: $n_1 = 100$, $n_2 = 150$, $\bar{x}_1 = 210$, $\bar{x}_2 = 220$,
$s_1 = 10$, $s_2 = 12$

H_0: $\bar{x}_1 = \bar{x}_2$ i.e. No difference b/w the income of the localities.

H_1: $\bar{x}_1 \neq \bar{x}_2$ [two tailed test]

$$|Z| = \left|\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right| = \left|\frac{210 - 220}{\sqrt{\frac{10^2}{100} + \frac{12^2}{150}}}\right| = |-7.1428|$$
$$= 7.1428$$

Conclusion:- As the calculated value of $|Z| > 1.96$, the significant value of Z at 5% level of significance. H_0 is rejected i.e. there is significant difference b/w the average incomes of the localities

Ques 3] Compute the standard error of the difference of the two sample means & find out if the two means significantly differ at 5% level of significance

	No. of items	Mean	S.D
Group I	50	181.5	3
Group II	75	179	3.6

Sol^n

$H_0: \bar{x}_1 = \bar{x}_2$; There is no significant diff b/w the samples.

$H_1: \bar{x}_1 \neq \bar{x}_2$. [Two tail test]

$$|Z| = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right| = \left| \frac{181.5 - 179}{\sqrt{\frac{9}{50} + \frac{(3.6)^2}{75}}} \right| = 4.2089$$

Conclusion: As the calculated value of $|Z| > 1.96$, at 5% level of significance. Ho is rejected ie there is significant difference b/w the samples.

Test of significance for the difference of standard deviations.

If s_1 & s_2 are the standard deviations of two independent samples then under H_0, the

$$|Z| = \left| \frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}} \right|$$

), σ_1 & σ_2 are S.D of population.

* If σ_1 & σ_2 are not known then

$$|Z| = \left| \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} \right|$$

Ques 1) Random samples drawn from two countries gave the following data relating to the heights of adult males.

Sol:

	Country A	Country B
Mean height	67.42	67.25
S. D	2.58	2.50
n	1000	1200

Is the diff. b/w the S.D significants?

Sol^n: H_0: $s_1 = s_2$ i.e. There is no significant difference b/w the S.D. of sample.

H_1: $s_1 \neq s_2$ (two tailed test)

$$|Z| = \left| \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} \right| = \left| \frac{2.58 - 2.50}{\sqrt{\frac{(2.58)^2}{2000} + \frac{(2.50)^2}{2400}}} \right| = 1.0387$$

since the calculated value of $|z| < 1.96$, at 5% level of significance. H_0 is accepted. i.e. there is no significance difference b/w S.D. of samples.

Ques2/ Intelligence test of two groups of boys & girls gives the following results.

	Mean	S.D	n
Girls	84	10	121
Boys	81	12	81

(a) Is the difference in mean scores significant.

(b) Is the difference b/w the S.D. significant?

Solⁿ (a) $H_0: \bar{x}_1 = \bar{x}_2$ i.e, the sample mean do not differ significantly.

$H_1: \bar{x}_1 \neq \bar{x}_2$ (two tailed)

$$|z| = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right| = \left| \frac{84 - 81}{\sqrt{\frac{10^2}{121} + \frac{(12)^2}{81}}} \right| = 0.1859$$

The calculated value of $|z| < 1.96$, we accept null hypothesis at 5% level of significance.

(b) $H_0: \sigma_1 = \sigma_2$ i.e the sample S.D do not differ significantly.

$H_1: \sigma_1 \neq \sigma_2$ (two tailed)

$$|z| = \left| \frac{s_1 - s_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}} \right| = \left| \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} \right| = \left| \frac{10 - 12}{\sqrt{\frac{100}{242} + \frac{144}{162}}} \right| =$$

$$= 1.7526$$

Scanned by CamScanner

$\therefore$ The Calculated value of $|Z| < 1.96$, we accept the null hypothesis at 5% level of significance. (11)