Population or Universe: An aggregate of objects under study is called population or universe.

- Universe
  - Finite Universe ← Finite no. of members
  - Infinite Universe ← infinite no. of member.
  - Existence Universe ← universe of concerte objects
  - Hypothetical Universe ← Collection of all possible ways in which a specific evnt can happen.

Sampling

Sample — Finite subset of a universe

Sample size → No. of individuals of a sample

Sampling — Process of selecting a sample from a universe.

Test of Significance

A hypothesis which is a definite statement about the population parameter called Null Hypothesis. It is denoted by H_0.

A hypothesis which is complementary to the null hypothesis is called an Alternative Hypothesis. It is denoted by H_1.

Critical Region

A region corresponding to a statistic t, in the sample space S which amounts to rejection of null hypothesis H_0 is called critical region or region of rejection.

The region of the sample space S which amounts to the acceptance H_0 is called acceptance region.

Level of Significance

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

Note The prob. α that a R.V of the statistic t belongs to the critical region is called level of significance

$P(t \in \omega | H_0) = \alpha$ ← max^m producer's risk

Error in Sampling

Type I error : when H_0 is true, we may reject it.

$P(\text{Reject } H_0 \text{ when it is true}) = P(\text{Reject } H_0 / H_0) = \alpha$

α is called the size of the type I error, also referred to as producer's risk.

Type II error : when H_0 is wrong, we may accept it.

$$P(\text{Accept } H_0 \text{ when it is wrong}) = P(\text{Accept } H_0 / H_1) = \beta$$

β is called the size of the type II error also referred to as consumer's risk.

Critical values or significant values

The values of the test statistic which seperates the critical region and acceptance region is called critical values or significance value.

Critical value depends → level of significance used
→ alternate hypothesis whether it is one tailed or two tailed.

Note : Null hypothesis

$$H_0 : \mu = \mu_0 \qquad (\text{mean} = \mu_0)$$

Alternate hypothesis will be

(i) $H_1 : \mu \neq \mu_0$ $(\mu > \mu_0$ or $\mu < \mu_0)$ ← two tailed alternate hypothesis

(ii) $H_1 : \mu > \mu_0$ ← right tailed or one tailed alt. hypo.

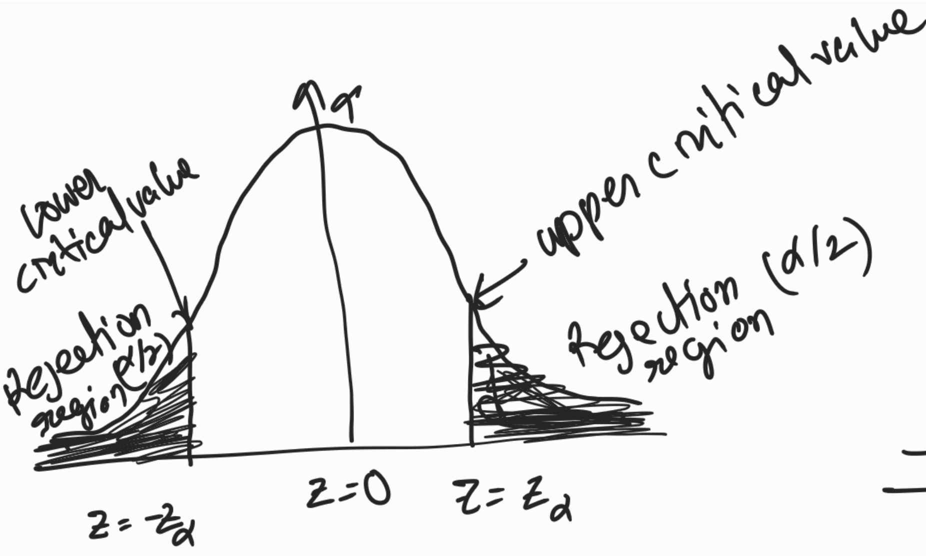
(iii) $H_1 : \mu < \mu_0$ ← left tailed or one tailed alt. hypo.

Test Statistic

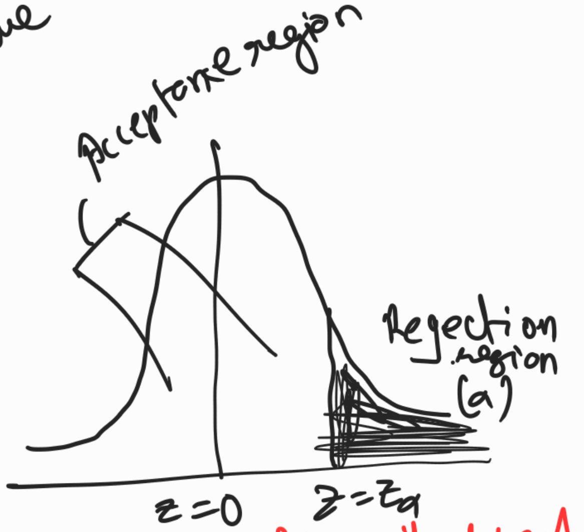
For larger samples corresponding to the statistic t, the variable $z = \dfrac{t - E(t)}{S.E(t)}$ is normally distributed with mean 0 and variance 1. The value of z given above under null hypothesis is called test statistic.

E ← error

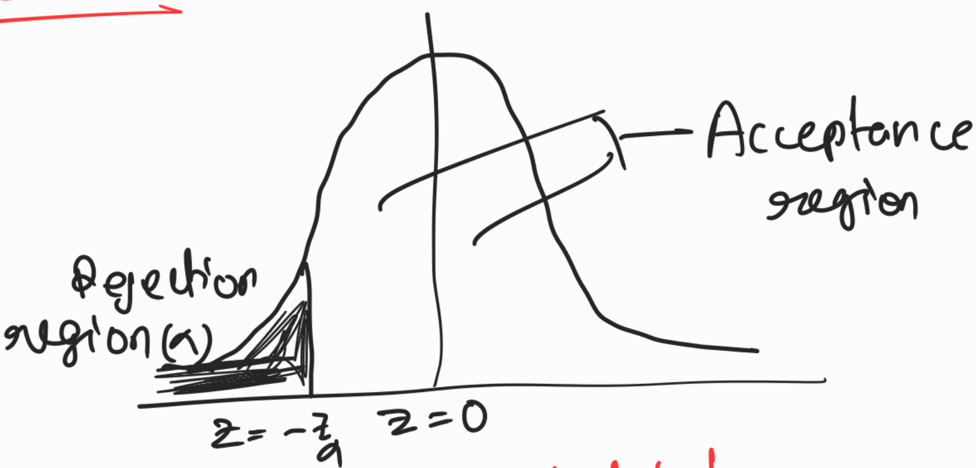
$S.E$ ← standard error.



Two tailed test



Right tailed test



Left tailed test

Level of Significance			
	1% (0.01)	5% (0.05)	10% (0.10)
Two tailed test	$\lvert z_\alpha \rvert = 2.58$	$\lvert z_\alpha \rvert = 1.966$	$\lvert z_\alpha \rvert = 0.645$
Right tailed	$z_\alpha = 2.33$	$z_\alpha = 1.645$	$z_\alpha = 1.28$
Left tailed	$z_\alpha = -2.33$	$z_\alpha = -1.645$	$z_\alpha = -1.28$

Algorithm

Step 1: **Null hypothesis**: Set up H_0 in clear terms

Step 2: **Alternate hypothesis**: Set up H_1 so that we could decide whether we should use one-tailed or two tailed test

Step 3: **level Significance**: Select the appropriate level of significance in advance depending on the reliability of the estimates.

Step 4: **Test statistic**: Compute $z = \frac{t - E(t)}{S.E(t)}$

Step 5: **Conclusion**: Compare the computed value of z with the critical value z_α at level of significance α

If $|Z| > Z_\alpha$ then reject H_0

and $|Z| < Z_\alpha$ then accept H_0.

conclude that there is significant difference

conclude that there is no significant difference.

Test of Significance for Large Samples

If the sample size $n > 30$, the sample is taken as large sample.

① Test of significance for single proportion

This test is used to determine whether difference between proportion of sample and the population is significant or not.

Let X be the number of successes in n independent trials with constant probability P of success for each trial

$$E(x) = nP$$

$$V(x) = nPQ$$

$Q = 1 - P$ = prob. of failure

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

$$E(p) = E(X/n) = \frac{1}{n} E[x] = \frac{np}{n} = p$$

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

$$S.E.(P) = \sqrt{\frac{PQ}{n}}$$

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

Q. A coin is tossed 400 times and the head turned up 216 times. Test the hypothesis that the coin is unbiassed.

Soln

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

H_1: The coin is biased i.e. $P \neq 0.5$
(Two tailed test)

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

p = proportion of success in the sample

$$= \frac{X}{n}$$

$$= \frac{216}{400}$$

$\Rightarrow p = 0.54$

$Q = 1 - P = 0.5$

Under H_0, test statistic $z = \dfrac{p - P}{\sqrt{PQ/n}}$

$$\Rightarrow |z| = \left| \frac{0.54 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{400}}} \right| = 1.6$$

Conclusion : $\because |z| = 1.6 < 1.96$

i.e. $|z| < z_\alpha$, z_α is the significant value of z at 5% level of significance

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

Q. 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.

Q. A machine is producing bolts of which a certain fraction is defective. A random sample of 400 is taken from a large batch and is found to contain 30 defective bolt. Does this indicate that the proportion of defectives is larger than that claimed by manufacturer where the manufacturer claims that only 5% of his product are defective. Find 95% confidence limits of the proportion of defective bolts in batch.

Soln: Null Hypothesis, H_0: The manufacture claim is accepted i.e. $P = \frac{5}{100} = 0.05$

$$Q = 1 - P = 0.95$$

Alternative Hypothesis, H_1 : $p > 0.05$ (Right tailed test)

p = observed proportion of sample

$$= \frac{30}{400}$$

$$= 0.075$$

Under H_0, the test statistic

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

$$= \frac{0.075 - 0.05}{\sqrt{\frac{0.05 \times 0.95}{400}}}$$

$$= 2.2941$$

Conclusion: The tabulated value of z at 5% level of significance for right tailed test is $z_\alpha = 1.645$

$$\therefore |z| = 2.2941 > 1.645$$

H_0 is rejected at 5% level of significance i.e., the proportion of defective is larger than the manufacturer claim.

For calculating 95% confidence limits of the proportion

It is given by $p \pm z_\alpha \sqrt{PQ/n}$

Note: The confidence limits for the population p are $p \pm \sqrt{\frac{pq}{n}}$

$$= 0.05 \pm 1.96 \sqrt{\frac{0.05 \times 0.95}{400}} = 0.05 \pm 0.02135$$

$$= 0.07135 ,\quad 0.02865$$

Hence, 95% confidence limits for the proportion of defective bolts are 0.07136 and 0.02865.

Q. 325 men out of 600 men chosen from a big city were found to be smokers. Does this information support the conclusion that the majority of men in the city are smoker.

Test of significance for difference of two proportions

Consider two sample sizes X_1 and X_2 of sizes n_1 and n_2 respectively taken from two different populations. To test the significance between the sample proportions P_1 and P_2. To test the statistic use

$$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{where} \quad P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$$

$$Q = 1 - P$$

$$p_1 = \frac{x_1}{n_1}, \quad p_2 = \frac{x_2}{n_2}$$

Q. Random sample of 400 men and 600 women were asked whether they would like to have a school near their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that the proportion of men and women in favour of the proposal are same at 5% level of significance.

Soln: H_0 : The proportion of men and women are in favour of the proposal

H_1 : The proportion of men and women are not in favour of the proposal.

Here, $n_1 = 400$ | $x_1 = 200$

$n_2 = 600$ | $x_2 = 325$

$$\therefore \quad p_1 = \frac{x_1}{n_1} = \frac{200}{400} = 0.50$$

$$p_2 = \frac{x_2}{n_2} = \frac{325}{600} = 0.541$$

Now,

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{200 + 325}{400 + 600}$$

$$= \frac{525}{1000}$$

$$= 0.52$$

$$Q = 1 - P = 1 - 0.52 = 0.48$$

$$\therefore \quad Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.50 - 0.541}{\sqrt{(0.52)(0.48)\left(\frac{1}{400}+\frac{1}{600}\right)}}$$

$$= \frac{-0.041}{0.0323}$$

$$= -1.269$$

$$\Rightarrow \quad |Z| = 1.269 < 1.96$$

$\therefore$ H_0 is accepted at 5% level of significance

Hence, proportion of men & women in favour of the proposal is same

Q. 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 population

Test of Significance for single mean

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from a large population $x_1, x_2, \ldots, x_N$ of size N with mean μ and variance σ^2.

Standard error of mean of a random sample of size n from a population with variance $\sigma^2 = \frac{\sigma}{\sqrt{n}}$

Test statistic, $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$; σ = S.D. of population

If σ is not known then use

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$; s = S.D. of sample

Q. A random sample of 200 measurements from a large population gave a mean of 50 and a S.D. of 9. Determine 95% confidence interval for the mean of popul^n

Soln:

Given,

$$n = 200$$

$$\bar{x} = 50$$

$$\sigma = 9$$

95% confidence levels are given by

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = 50 \pm 1.96 \left(\frac{9}{\sqrt{200}}\right)$$

$$= 50 \pm 1.247$$

Hence, 95% confidence limits are 48.75 and 51.20

Note

(i) At 5% level of significance, 95% confidence limits are

$$\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}$$

(ii) At 1% level of significance, 99% confidence limits are

$$\bar{x} - 2.58 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.58 \frac{\sigma}{\sqrt{n}}$$

Q. The average marks in Mathematics of a sample of 100 students was 51 with a S.D. of 6 marks Could this have a random sample from a population with average marks 50?

Soln: Here,

$$n = 100$$

$$\bar{x} = 51$$

$$s = 6 \qquad (\sigma \text{ is unknown})$$

$$\mu = 50$$

H_0: The sample size is drawn from a populn with mean 50, $\mu = 50$

H_1: $\mu \neq 50$

Now,

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{51 - 50}{6/\sqrt{100}} = 1.66$$

Conclusion: $\because |Z| = 1.66 < 1.96$

$\therefore$ H_0 is accepted.

Q. A sample of 1000 students from a university was taken and their average weight was found to be 112 pounds with S.D. of 20 pounds. Could the mean weight of students in the population be 112 pounds

Test of Significance for difference of means

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

If the samples are drawn from same population then

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

Q. The number of accidents per day were studied for 144 days in town A and for 100 days in town B and the following information were obtained

	Mean of Accidents	S.D.
Town A	4.5	1.2
Town B	5.4	1.5

Is the difference bet^n the mean of accidents of the two towns statistically significant?

Sol^n: Here,

$n_1 = 144$, $\bar{x}_1 = 4.5$, $s_1 = 1.2$

$n_2 = 100$, $\bar{x}_2 = 5.4$, $s_2 = 1.5$

H_0: There is no significant difference between the mean of accidents i.e. $\bar{x}_1 = \bar{x}_2$

H_1: There is a significant difference between the mean of accidents i.e. $\bar{x}_1 \neq \bar{x}_2$

$$Z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{4.5 - 5.4}{\sqrt{\frac{1.49}{144} + \frac{2.25}{100}}}$$

$$= \frac{-0.9}{0.192}$$

$\Rightarrow |Z_0| = 4.687 > 2.56$

(Test at 1% level of sig. can also be tested at 5%)

$\therefore$ Ho is rejected.

Hence there is a significant diff....

Q. A random sample of 200 villages from Coimbatore district gives the mean population per village at 485 with a S.D. of 50. Another random sample of the same size from the same district gives the mean population per village at 510 with a S.D. of 40. Is the difference between the mean values given by the two sample statistically significant?