UNIT-IV - Chapter-1

Estimation

Introduction:- In the sampling theory, we are primarly concerned with two types of problems (i) Estimation (ii) Testing of hypothesis. In estimation some characterstic (or) feature of the population in which we are interested may be completely unknown to us and we may like to make a guess about this characterstics entirely on the basis of a random sample drawn from the population. This type of problem is known as the problem of "Estimation".

→ **Estimate:-** An estimate is a statement made to find an unknown population ~~pop~~ parameter.

Estimator:- The method of determining unknown population parameter is called "Estimator"

Estimation:- The process by which we draw a conclusion about some measure of a population based on a sample value.

→ First we will briefly discuss the Estimation procedures. we have two types of Estimation procedures.

(i) Point Estimation

(ii) Interval Estimation.

Point Estimation :- If an estimate of the population parameter is given by a single value, then the estimate is called a "point Estimation" of the parameter.

→ A point estimator is a statistic for estimating a population parameter θ and will be denoted by $\hat{\theta}$.

Interval Estimation :- An Estimate of a population parameter given by two magnitudes within which a parameter can lie is called interval estimate of the parameter.

Ex:- i) If the height of a student is measured as 162 Cms, then the measurement gives a point estimation.

ii) If the height is given as (163 ± 3.5) Cms, then the height lies between 159.5 Cms and 166.5 Cms and the measurement gives interval estimation.

Note:- An Estimator is not expected to estimate the population parameter without error. An estimator should be close to the true value of unknown parameter.

→ An Interval estimate of a population parameter θ is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_u$, where $\hat{\theta}_L$ and $\hat{\theta}_u$ depend on the value of the statistic $\hat{\theta}$ for a particular sample and also on the sampling distribution of $\hat{\theta}$.

Properties of Estimation:

(OR)

Criteria for the Good Estimator.

→ The criteria for the good estimator are
(i) consistency
(ii) Efficiency
(iii) Unbiased ness
(iv) Sufficiency.

Consistency :- An estimator $\hat{\theta}_n$ of a parameter θ is consistent if it converges to θ as $n \to \infty$.

Efficiency :- A statistic $\hat{\theta}_1$ is said to be more efficient unbiased estimator of the parameter θ than the statistic $\hat{\theta}_2$ if
(i) $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ.
(ii) $V(\hat{\theta}_1) < V(\hat{\theta}_2)$.

Unbiasedness :- A statistic $\hat{\theta}$ is said to be an unbiased estimate of θ if $E(\hat{\theta}) = \theta$ for all θ.

Sufficiency :- An estimator is said to be sufficient for a parameter, if it contains all the information in the sample regarding the parameter.

→ Unbiased Estimator :- A statistic (or) point estimator $\hat{\theta}$ is said to be an "Unbiased estimator" of the parameter θ if $E(\hat{\theta}) = \theta$.

Example(1):- S.T S^2 is an unbiased estimator of the parameter σ^2 ie $[E(S^2) = \sigma^2]$

proof:- Let $\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}[(x_i-\mu)-(\bar{x}-\mu)]^2$

$$= \sum_{i=1}^{n}(x_i-\mu)^2 - 2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\mu) + n(\bar{x}-\mu)^2$$

$$= \sum_{i=1}^{n}(x_i-\mu)^2 - n\cdot(\bar{x}-\mu)^2 \quad \text{——— (1)}$$

$$E(S^2) = E\left[\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{n-1}\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}E(x_i-\mu)^2 - nE(\bar{x}-\mu)^2\right] \text{ from eq (1)}$$

$$\left[\because E(x_i-\mu)^2 = \sigma^2, \quad E(\bar{x}-\mu)^2 = \frac{\sigma^2}{n}\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}\sigma^2_{x_i} - n\,\sigma^2_{\bar{x}}\right]$$

However $\sigma^2_{x_i} = \sigma^2$ for $i = 1,2,3\ldots n$

and $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$.

$$E(S^2) = \frac{1}{n-1}\left[n\cdot\sigma^2 - \cancel{n}\cdot\frac{\sigma^2}{\cancel{n}}\right]$$

$$= \frac{1}{n-1}(n\sigma^2 - \sigma^2)$$

$$= \frac{1}{\cancel{n-1}}(\sigma^2\cancel{(n-1)}) = \sigma^2$$

$$\therefore \boxed{E(S^2) = \sigma^2}$$

Confidence Interval Estimates of Parameters

→ Confidence interval has a specified confidence (or) probability of correctly estimating the true value of the population parameter.

→ Computation of confidence interval and confidence limits is based on the sampling distribution of a statistic.

(1) Confidence interval for population mean (μ) is

$$\left(\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right).$$

$z_{\alpha/2}$ for 95% → 1.96

$z_{\alpha/2}$ for 99% → 2.58

$z_{\frac{\alpha}{2}}$ for 90% → 1.64.

$z_{\frac{\alpha}{2}}$ for 99.73% → 3.

2) Confidence interval for population proportion (P)

$$\left(P - z_{\frac{\alpha}{2}}\sqrt{\frac{PQ}{n}},\ P + z_{\frac{\alpha}{2}}\sqrt{\frac{PQ}{n}}\right).$$

3) Confidence interval for the difference ($\mu_1 - \mu_2$) of two population Means μ_1 and μ_2

$$\left((\bar{x}_1 - \bar{x}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ (\bar{x}_1 - \bar{x}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

4) Confidence interval for the difference $P_1 - P_2$ of two population proportions. $\left((P_1 - P_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{P_1Q_1}{n} + \frac{P_2Q_2}{n_2}},\ (P_1 - P_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}\right.$

Note:- (1) Sample size for estimating population mean

$$\boxed{n = \left(\frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{E}\right)^2}$$

2) Maximum Error of Estimate E for large samples is $(n>30)$

$$\boxed{E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}$$

3) Sample size for Estimating population proportion

$$n = \frac{\left(Z_{\frac{\alpha}{2}}\right)^2 \cdot PQ}{E^2} \quad \text{where } Q = 1-P.$$

4) Maximum Error of Estimate E for small samples is

$$\boxed{E = t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}}$$

5) Sample size for estimating population mean (small sample) $(n<30)$.

$$\boxed{n = \left(\frac{t_{\frac{\alpha}{2}} \cdot s}{E}\right)^2}$$

Problems
1) Assu

Problems :

1) Assume that $\sigma = 20$, how large a random sample be taken to assert with probability 0.95 that the sample mean will not differ from the true mean by more than 3 points.

Sol:- Given maximum Error $E = 3$. and $\sigma = 20$

we have $z_{\frac{\alpha}{2}}$ for $95\% = 1.96$

Sample size $n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$

$$n = \left(\frac{1.96 \times 20}{3}\right)^2$$

$$n = 170.74$$

$$\therefore \boxed{n \cong 171}.$$

2) What is the maximum error one can expect to make with probability 0.90 when using the mean of a random sample size $n=64$ to estimate the mean of population with $\sigma^2 = 2.56$.

Sol:- Here $n = 64$, and $\sigma^2 = 2.56 \Rightarrow \sigma = \sqrt{2.56} = 1.6$

The probability $= 0.90$ (or) 90%

Confident limit $= 90\%$, $z_{\frac{\alpha}{2}}$ for $90\% = 1.645$

[from z-distribution Table]

$\therefore$ Maximum Error $E = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

$$= (1.645) \times \frac{1.6}{\sqrt{64}}$$

$$\boxed{E = 0.329}$$

3) If we can assert with 95% that the maximum. The is 0.05 and $p = 0.2$ find the size of the sample.

Sol:- Given $p = 0.2$, $E = 0.05$

$$q = 1 - 0.2 = 0.8$$

and $z_{\frac{\alpha}{2}}$ for 95% = 1.96.

Maximum Error for proportion p is $E = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{pq}{n}}$.

$$0.05 = 1.96 \times \sqrt{\frac{(0.2)(0.8)}{n}} \Rightarrow (0.05)^2 = (1.96)^2 \times \frac{(0.2)(0.8)}{n}.$$

$\therefore$ Sample size $n = \dfrac{(1.96)^2 (0.2)(0.8)}{(0.05)^2}$

$$\boxed{n = 246}$$

4) A random sample of size 100 has a standard deviation of 5. what can you say about the maximum error with 95% confidence.

Sol:- Given $s = 5$, $n = 100$

$z_{\frac{\alpha}{2}}$ for 95% confidence = 1.96

Maximum Error $E = z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$

$$E = (1.96)\left(\frac{5}{\sqrt{100}}\right)$$

$$\boxed{E = 0.98}$$

(5) The mean and standard deviation of a population are 11,795 and 14,054 respectively what can one assert with 95% confidence about the maximum error if $\bar{x} = 11,795$ and $n = 50$. And also construct 95% confidence interval for the true mean.

Sol:- Given $\mu = 11,795$

~~Sample~~ S.D $\sigma = 14,054$

$\bar{x} = 11,795$

Sample size $n = 50$

Maximum Error $= Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

$Z_{\frac{\alpha}{2}}$ for 95% $= 1.96$

$\therefore E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = (1.96) \times \frac{14,054}{\sqrt{50}} = 3,899.$

Confidence interval $= \left(\bar{x} - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{x} + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$

$= (11,795 - 3899,\ 11,795 + 3899)$

$= (7896, 15694)$

6) A random sample of 400 items is found to have mean 82 and S.D of 18. Find the maximum error of estimation at 95% confidence interval. Find the confidence limits for the mean if $\bar{x} = 82$.

Sol:- Given S.D $\sigma = 18$, $\bar{x} = 82$

Sample size $n = 400$

$Z_{\frac{\alpha}{2}}$ for 95% Confidence $= 1.96$ (from z-distribution Table)

Maximum Error $E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

$$E = 1.96 \times \frac{18}{\sqrt{20}} = 1.764.$$

The confidence limits are $= \left(\bar{x} - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$

$$= (82 - 1.764, 82 + 1.764)$$

$$= (80.236, 83.764)$$

$\therefore$ confidence limits are 80.236 and 83.764.

7) Find 95% confidence limits for the mean of a normality distributed population from which the following samples was taken 15, 17, 10, 18, 16, 9, 7, 11, 13, 14.

Sol:- we have $\bar{x} = \frac{15+17+10+18+16+9+7+11+13+14}{10} = 13.$

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} = \frac{1}{9}\left[(15-13)^2 + (17-13)^2 + (10-13)^2 + (18-13)^2 + (16-13)^2 + (9-13)^2 + (7-13)^2 + (11-13)^2 + (13-13)^2 + (14-13)^2\right]$$

$$s^2 = \frac{40}{3} \Rightarrow S = \sqrt{\frac{40}{3}} = \frac{\sqrt{40}}{\sqrt{3}}$$

$Z_{\frac{\alpha}{2}}$ for 95% $= 1.96.$

Confidence limits are $= \left(\bar{x} - Z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}, \bar{x} + Z_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}\right)$

$$= \left(13 - 1.96 \cdot \frac{\sqrt{40}}{\sqrt{10}\cdot\sqrt{3}}, 13 + 1.96 \frac{\sqrt{40}}{\sqrt{10}\cdot\sqrt{3}}\right)$$

$$= (13 - 2.26, 13 + 2.26) = (10.74, 15.26).$$

In a study of an automobile insurance a random sample 80 body repair costs had a mean 472.36 Rs and the S.D of 62.35 Rs. If $\bar{x}$ is used as a point estimate to the true average repair costs, with what confidence we can assert that the maximum error doesn't exceed 10 Rs.

Given $n = 80$, $\bar{x} = 472.36$

$$\sigma = 62.35$$

$$E = 10$$

we have $\boxed{E = Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}$

$$10 = Z_{\frac{\alpha}{2}} \cdot \frac{62.35}{\sqrt{80}}$$

$$Z_{\frac{\alpha}{2}} = \frac{10 \times \sqrt{80}}{62.35} = 1.4345$$

$$Z_{\frac{\alpha}{2}} = 1.43$$

from normal distribution table Area when $Z = 1.43$ is 0.4236.

$$\therefore \frac{\alpha}{2} = 0.4236$$

$$\alpha = 2 \times (0.4236)$$

$$\alpha = 0.8472$$

Confidence $= (1-\alpha)100\% = 84.72\%$

*) ~~In a study of an automobile insurance a random sample of body repair costs had a mean~~

7) Among 100 fish caught in a large lake, 18 were inedible due to the pollution of the environment. With what confidence can we assert the error of this estimate is at most 0.65?

Sol: given $n = 100$

$$E = 0.065$$

P = sample proportion of inedible fish $= \frac{18}{100} = 0.18$

$q = 1 - p = 1 - 0.18 = 0.82$

Maximum error of estimate for true propotion $E = z_{\frac{\alpha}{2}} \sqrt{\frac{pq}{n}}$

$$0.065 = z_{\frac{\alpha}{2}} \sqrt{\frac{(0.18)(0.82)}{100}}$$

$$z_{\frac{\alpha}{2}} = \frac{0.065}{0.038} = 1.71$$

from normal table, when $z = 1.71$ the Area is $= 0.4564$

$$z_{\frac{\alpha}{2}} = 0.4564$$

$$\Rightarrow \alpha = 2 \times (0.4564) = 0.9128$$

$$\alpha = 91.28\% \cong 91\%$$

10) A random sample of 100 teachers in a large metropolitian area revealed a mean weekly salary 487 Rs with a S.D 48 Rs. With what degree of confidence can we assert that the average weekly salary of all teachers in the metropolitian area is between 472 to 502?

Sol: Given $\mu = 487$, $\sigma = 48$, $n = 100$,

w.k.t $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x} - 487}{\frac{48}{\sqrt{100}}}$

when $\bar{x} = 472$

$$z_1 = \frac{472 - 487}{4.8} = -3.125 < 0$$

when $\bar{x} = 502$, $z_2 = \frac{502 - 487}{4.8} = 3.125 > 0$

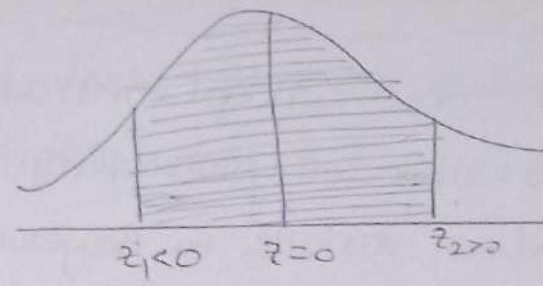
Let x be the mean salary of teacher. Then

$$P(472 < x < 502) = P(-3.125 < z < 3.125).$$

$= P(-3.125 < z < 3.125)$

$= A(-3.125) + A(3.125)$

$= A(3.125) + A(3.125)$



$= 2A(3.125)$ from Normal table

$= 2(0.4991)$ ✓

$= 0.9982$ ✓

$\therefore$ Thus we can ascertain with 99.82% confidence

11) what is ~~size~~ the size of the smallest sample required to estimate an unknown proportion to within a maximum error of 0.06 with atleast 95% confidence.

Soln Given $E = 0.06$

confidence limit = 95%

ie $(1-\alpha)100\% = 95\%$

$(1-\alpha)100 = 95$

$1-\alpha = 0.95$

$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

$z_{\frac{\alpha}{2}} = 1.96$ ✓

Here P is not given, so we take $P = \frac{1}{2}$, $Q = \frac{1}{2}$

$\therefore$ Size of the sample $n = \left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2 PQ$

$n = \left(\frac{1.96}{0.06}\right)^2 \cdot \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}\left(\frac{1.96}{0.06}\right)^2$

$n = \frac{1}{4}\left[\frac{1.96}{0.06}\right]^2 = 266.78 \simeq 267$

$\boxed{n = 267}$

Bayesian Estimation

The new concept introduced in Bayesian methods is perso (or) subjective probability. Also, parameters are considered as random variable in Bayesian method. To estimate the mean of a population, μ is treated as a random variable whose distribution is indicative of the "strong feelings".

→ Let μ_0 and σ_0 be the mean and standard deviation of such a subjective "prior distribution".

Bayesian Estimation :

combining the prior feelings about the possible values of μ with direct sample evidence the "posterior" distribution of μ in Bayesian estimation is approximated by normal distribution with

$$\mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} \quad \text{and} \quad \sigma_1 = \sqrt{\frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}}$$

when n = sample size, $\bar{x}$ = sample mean and

S = standard deviation of sample ($S = \sigma$)

Here μ_1 and σ_1 are known as the mean and standard deviation of the posterior distribution.

→ In the computation of μ_1 and σ_1, σ^2 is assumed to be known, when σ^2 is unknown, which is generally the case is replaced by sample variance s^2 provided $n \geq 30$ (large samp

→ Bayesian Interval for μ : $(1-\alpha)\,100\%$ Bayesian interval for μ is given by

$$\boxed{\left(\mu_1 - z_{\frac{\alpha}{2}}\,\sigma_1,\ \mu_1 + z_{\frac{\alpha}{2}}\cdot\sigma_1\right)}$$

examin

A professor's feelings about the mean mark in the final examination in "probability" of a large group of students is expressed subjectively by normal distribution with $\mu_0 = 67.2$ and $\sigma_0 = 1.5$.

(a) If the mean mark lies in the interval $(65.0, 75)$ determine the prior probability the professor should assign to the mean mark.

(b) Find the professor mean μ_1 and the posterior s.d σ_1 if the examinations are conducted on a random sample of 40 students yielding 74.9 and S.D 7.4. Use $S = 7.4$ as an estimate σ.

(c) Determine the posterior probability which he will thus assign to the mean mark being in the interval $(65, 70)$ using results obtained in (b).

(d) Construct a 95% Bayesian interval for μ.

Sol:-

(a) Here $\mu_0 = 67.2$, $\sigma_0 = 1.5$ and $n = 40$

W.k.t $\quad z = \dfrac{\bar{x} - \mu}{\sigma}$

when $\bar{x} = 65$, $\quad z_1 = \dfrac{65 - 67.2}{1.5} = -1.466$

$\bar{x} = 75$, $\quad z_2 = \dfrac{75 - 67.2}{1.5} = 1.866$

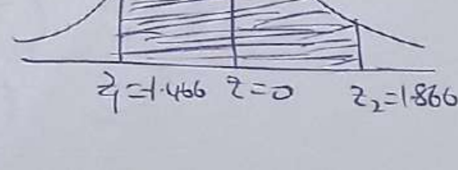
$\therefore \; P(65 < x < 70) = P(-1.466 < z < 1.866)$

$= A(-1.466) + A(1.866)$

$= A(1.466) + A(1.866)$

$= 0.4292 + 0.4693$

$= 0.8985$



b) Here $\bar{x} = 74.9 \quad \sigma = s = 7.4$

Posterior mean $\mu_1 = \dfrac{n \cdot \bar{x} \sigma_0^2 + \mu_0 \sigma^2}{n \cdot \sigma_0^2 + \sigma^2}$

$$\text{ie } \mu_1 = \frac{40(74.9)(1.5)^2 + (67.2)(7.4)^2}{40(1.5)^2 + (7.4)^2} = 71.987 \cong 72$$

posterior S.D, $\sigma_1 = \sqrt{\dfrac{\sigma^2 \cdot \sigma_0^2}{n \cdot \sigma_0^2 + \sigma^2}} = \sqrt{\dfrac{(7.4)^2 (1.5)^2}{40(1.5)^2}}$

$$\sigma_1 = 0.9225 \cong 0.923.$$

c) Here $\mu_1 = 72, \ \sigma_1 = 0.923$, w.k.t $z = \dfrac{\bar{x} - \mu}{\sigma}$

when $\bar{x} = 65$

$$z_1 = \frac{65 - 72}{0.923} = -7.5839$$

when $\bar{x} = 70$

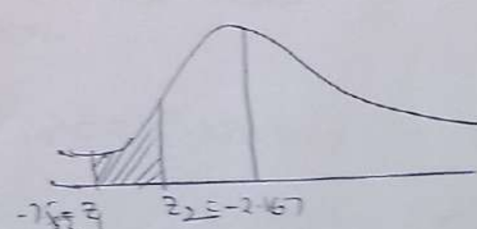
$$z_2 = \frac{70 - 72}{0.923} = -2.16684$$

posterior probability $= P(65 < x < 70)$

$= P(-7.584 < z < -2.167)$

$= \cancel{A(-7.584)} \cong 0.5 - 0.4850$

$= 0.0150$



d) 95% Bayesian Interval limits are

$= (\mu_1 - z_{\frac{\alpha}{2}} \cdot \sigma_1, \ \mu_1 + z_{\frac{\alpha}{2}} \cdot \sigma_1)$

$= (71.987 - 1.96(0.9225), \ 71.987 + 1.96(0.9225))$

$= (70.178, \ 72.909)$

UNIT - IV

Chapter-2 {Tests of Hypotheses}

Introduction:- when parametric values are known we estimate them through sample values but the problem arises when the ~~p~~ sample providers a value which is neither exactly equal to the parameter value. ~~not to fall~~.

→ In that situation one has to dovelop some procedure which is to decide whether to accept (or) Reject a value on the basis of sample values, such a procedure is known as "Testing of Hypothesis".

(OR)

→ A procedure for deciding whether to accept (or) reject a perticular hypothesis is called a "Test of Hypothesis". Also known as "Test of Significance".

Statistical Hypothesis :-

In many circumstances, to arrive at decisions about the population on the basis of sample information, we make assumptions (or guesses) about the population parameters involved. such an assumption (or statement) is called "statistical Hypothesis" which may (or) may not be true.

Testing of statistical Hypothesis

A statistical test of Hypothesis is a procedure which make one to take decision about the acceptancy (or) Rejection of the Hypothesis.

Test of Hypothesis involves the following steps.

Step (1) :- Statement (or Assumption) of Hypothesis

There are two types of Hypothesis

(i) Null Hypothesis (H_0) :- For applying the tests of Hypothesis, we first setup a Hypothesis a definite statement about the population parameter. Such a hypothesis of no- difference is called "Null Hypothesis".

(There is no Significance b/w two parameters).

It is denoted by H_0.

(ii) Alternative Hypothesis (H_1) :- Any statistical hypothesis that differs from a given "Null hypothesis" is called "Alternative Hypothesis". It is denoted by (H_1).

imp

Note :- The Alternative Hypothesis is very important to decide whether we have to use a single-tailed (right tailed (or) left tailed) (or) two tailed test.

Example

that

Example :- If we want to test the null Hypothesis that the population has a specified mean (μ_0) say ie $H_0 : \mu = \mu_0$, then the

Alternative hypothesis (i) $H_1 : \mu \neq \mu_0$

(ii) $H_1 : \mu > \mu_0$ (Right tailed)

(iii) $H_1 : \mu < \mu_0$ (left tailed).

→ The Alternative hypothesis (i) is known as a two tailed alternative hypothesis

(ii) is known as right-tailed, (iii) is known as left-tailed.

M.imp

→ Alternative hypothesis, one has to choose from the above three forms depending on the situation posed.

Step ② :- Specification of the Level of Significance

The level of significance denoted by 'α', is the confidence (%) with which we rejects (or) accepts the Null Hypothesis H_0.

→ In practice we take 5%, 2%, 1% etc level significance

Example :- 5% Level of significance in a test procedure indicates that there are about 5 cases in 100 cases that we would reject the Null Hypothesis H_0 when it is true. ie we are about 95% confident that we have made the right decision.

similarly, 1% level and 2% level of significance.

M. imp

Note:- ~~Select~~ Select the appropriate level of significance

(a). If it is not in the given problem, we choose 5% level of significance.

Step ③:- Identification of the Test statistic

There are several tests of significance, Z-Tests, t-Test, F- test etc.

→ First we have to select the right test depending on the nature of information given in the problem. Then we construct the test criterion and select the appropriate probability distibution.

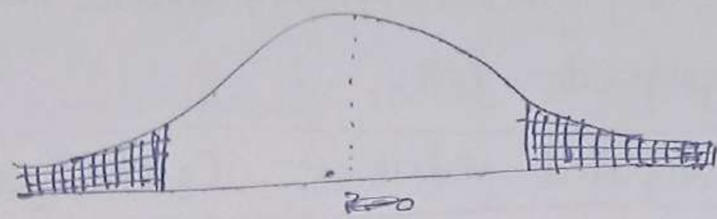
Step ④:- Critical Region.

The critical region is formed based on

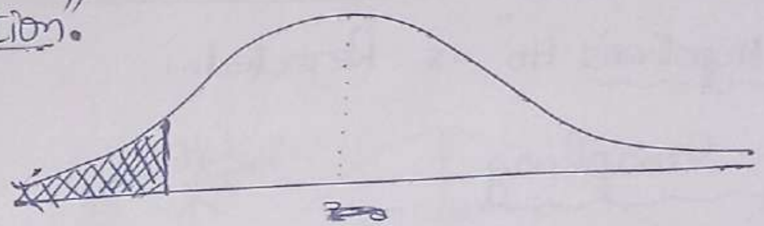
(a) Distribution of the stastistic ie whether the statistic follows the normal distribution (z), t-distribution, F-distribution.

(b) Form of Alternative hypothesis:

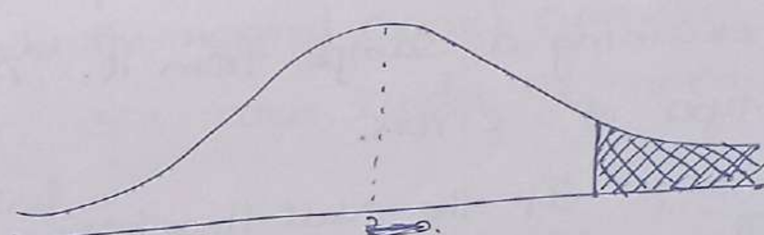
→ If the form of Alternative hypothesis has ≠ sign, the critical region is devided equally in the left and right tailes sides of the distribution.



→ If the form of Alternative hypothesis has '<' sign, the enitire critical region is taken in the "left tail of the distribution".



→ If the form of Alternative hypothesis has '>' sign, the entire critical region is taken on the 'right tail of the distribution"



Step (5) :- Making Decision (or) Conclusion.

we compute the value of appropriate test (z, t, F-test) and ascertain whether the computed value falls in acceptance (or) rejection region on the specified level of significance.

M. imp ⇒ In finding the Acceptance (or) Rejectance, we have to use Critical values given in statistical tables.

By comparing test statistic value with critical

of appropriate test.

→ If $\boxed{\text{Computed value} < \text{Critical value}}$, Then "Null Hypothesis H_0" is Accepted.

→ If $\boxed{\text{Computed value} > \text{Critical value}}$, then "Null Hypothesis H_0" is Rejected.

Errors of Sampling

The main objective in sampling theory is to find valid inferences about the population parameters on the basis of the sample results.

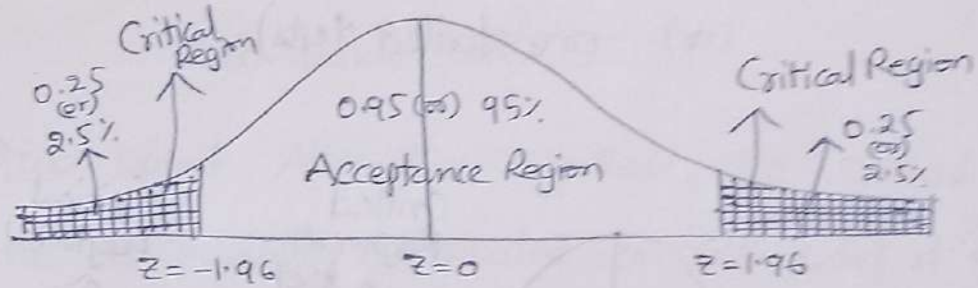
→ In practice we decide to accept (or) to reject the lo after examing a sample from it. As such we have two types of errors.

Type I Error :- If the Null Hypothesis H_0 is true, but it is Rejected by test procedure, then the Error is called Type I error.

Type II Error :- If the Null Hypothesis H_0 is False, but it is Accepted by Test procedure, Then the Error is called Type-II Error.

Critical Region :- A region corresponding to a statistic, in the sample space S which leads to the rejection of Null Hypothesis H_0 is Critical Region (or) Rejection Region. Those region which lead to the acceptance of Null Hypothesis H_0 give us a region called Acceptance Region.

Example :-



→ From the above diagram, we see that Z- lies between -1.96 to 1.96. and we are 95% (Area under the normal curve) confident that the hypothesis is true, .5% area under the normal curve, H_0 is to be rejected.

→ Z- lies inside the range -1.96 to 1.96 is called the region of Acceptance of the hypothesis.

→ Z- lies outside the range -1.96 to 1.96 is called the region of Rejectance of the hypothesis.

Similarly, we can define critical region at any other level of significance.

→ The values -1.96 and 1.96 are called critical values at 5% level of significance.

One-Tailed Test

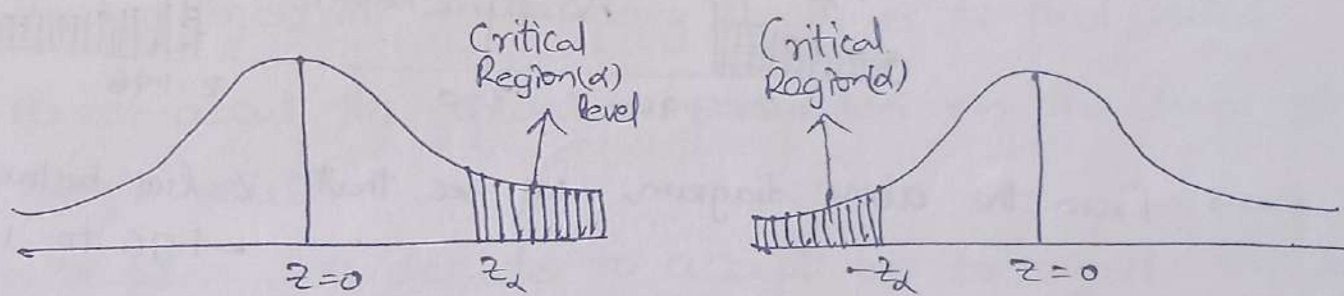
If we have to test whether the population mean μ has a specified value μ_0, then the

Null Hypothesis $H_0 : \mu = \mu_0$

Then Alternative Hypothesis $H_1 : \mu > \mu_0$ (Right-tailed)

$H_1 : \mu < \mu_0$ (Left-tailed).

(Right-tailed and Left-tailed test are called Single-tailed (or) one-tailed tests).



(Right-tailed) (Left-tailed).

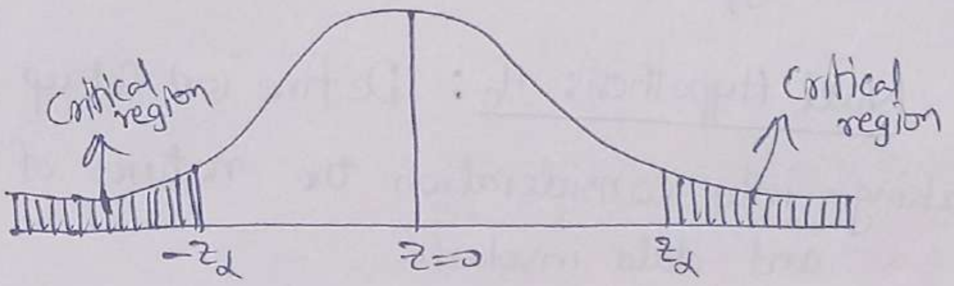
→ In the right-tail test ($H_1 : \mu > \mu_0$), the critical region (rejection region) $z > z_\alpha$ lies entirly in the right-tail of the sampling distribution of sample mean $\bar{x}$ with area equal to the level of significance (α).

→ In the left-tail test ($H_1 : \mu < \mu_0$), the critical region $z < -z_\alpha$ lies entirely in the left-tail of the sampling distribution of the sample mean $\bar{x}$ with area equal to the level of significance 'α'.

Two-tailed Test

Suppose we want to test the Null Hypothesis $H_0: \mu = \mu_0$

Alteranative Hypothesis $H_1: \mu \neq \mu_0$



(two-tailed test)

→ H_1 is two-tailed Alternative hypothesis, the critical region under the curve is equally distributed on both sides of the mean.

Thus, the critical area under the right-tail = The critical area under the left-tail

= Half of the total Area

= $\frac{1}{2}$ probability of rejection = $\frac{\alpha}{2}$

with critical statistic $z_{\alpha/2}$, where α is Level of significance.

Critical values of x for both two-tailed and one-tailed tests at 1%, 5%, 10% level of Significance

Table : Critical values of z

Level of significance (α)	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	$z_\alpha = 2.33$	$z_\alpha = 1.645$	$z_\alpha = 1.28$
Critical values for Left-tailed Test	$z_\alpha = -2.33$	$z_\alpha = -1.645$	$z_\alpha = -1.28$

Procedure for Testing of Hypothesis

Varies steps involved in testing of Hypothesis, are given below: Infact the same steps are followed for conducting all tests of Hypothesis.

Step ① :- Null Hypothesis H_0: Define (or) Set up a Null hypothesis H_0 taking into consideration the nature of the problem and data involved.

Step ② :- Set up the Alterantive Hypothesis H_1 : we ~~should~~ could ~~use~~ decide whether we should use one-tailed test (or) two tailed test.

Step ③ :- Select the appropriate Level of significance (α):

If it is not given in the problem we chose 5% level of significance.

Step ④ :- Compute the Test statistic $z = \frac{t - E(t)}{S.E \text{ of } t}$ under the null Hypothesis

Here 't' is Sample statistic, S.E is Standard Error of 't'.

Step (5) :- Conclusion: We compare the computed value of the test statistic z with the critical value z_α at given level of significance (α).

(i) If $|z| < z_\alpha$ (Caluculated value < Critical value) we conclude that it is not significant, Null hypothesis (H_0) Accepted.

(ii) If $|z| > z_\alpha$, Null Hypothesis H_0 is Rejected. at the level of significance 'α'.

Test of Hypothesis for Large Samples

→ If the sample size ($n \geq 30$), then we consider such samples as Large samples.

→ For Large samples, the sampling distribution of a statistic is approximately a Normal distribution.

→ Suppose we wish to test the hypothesis that the probability of success in such ~~trail~~ trial is p. Assuming it to be true, the mean μ and standard deviation σ of the sampling distribution of number of success are np and $\sqrt{npq}$ respectively.

If x be the observed number of success in the sample and z is the standard normal variate

$$z = \frac{x-\mu}{\sigma}.$$

Note :- Assumptions for Large Samples

The following are assumptions under which tests of hypothesis are applied:

1. The random sampling distribution of statistic has the properties of the normal curve. This may not hold in case of small samples.

2. Values (ie statistic) given by the samples are sufficiently close to the population values (ie parameters) and can be used in its place for calucalating the standard Error (S.E) of the statistic.

imp Problems :-

1). A coin was tossed 960 times and returned head 183 times. Test the hypothesis that the coin is unbiased. Use a 0.05 level of significance.

Sol:- Here $n = 960$

p = probability of getting head $= \frac{1}{2}$.

$$q = 1 - p = \frac{1}{2}$$

$$\mu = n \cdot p = \overset{480}{\cancel{960}}\left(\frac{1}{\cancel{2}}\right) = 480.$$

$$\sigma = \sqrt{npq} = \sqrt{\cancel{960} \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{240} = 15.49.$$

x = number of success $= 183$.

Step ①:- Null Hypothesis H_0: The coin is unbiased.

Step ②:- Alternative Hypothesis H_1: The coin is biased.

Step ③:- Level of significance (α): 5%

Step ④:- Test statistic $\boxed{z = \frac{x - \mu}{\sigma}} = \frac{183 - 480}{15.49} = -19.17$

$$|z| = |-19.17| = 19.17$$

$$\boxed{|z| = 19.17.}$$

Critial value at 5% level of significance $z_\alpha = 1.96$.

Step ⑤:- Conclusion $|z| = 19.17$, $z_\alpha = 1.96$

$\therefore |z| > z_\alpha$ $\therefore$ Null Hypothesis H_0 is Rejected.

We conclude that the coin is biased.

7) A die is thrown 256 times and it turns up with an even digit 150 times. Is the die biased?

Sol:- Here $n = 256$

p = probability of getting an even digit (2 (or) 4 (or) 6) $= \frac{\cancel{3}}{\cancel{6}_2}$

$p = \frac{1}{2}$

$q = 1 - p = 1 - \frac{1}{2}$

$q = \frac{1}{2}$.

$\therefore \mu = n \cdot p = \overset{128}{\cancel{256}}\left(\frac{1}{\cancel{2}}\right) = 128$

$\boxed{\mu = 128}$

$\sigma = \sqrt{npq} = \sqrt{\overset{64}{\cancel{256}} \times \frac{1}{\cancel{4}}} = \sqrt{64} = 8$

$\boxed{\sigma = 8}$

x = number of success.

1) Null Hypothesis H_0: The die is unbiased.

2) Alternative hypothesis H_1: The die is biased.

3) Level of significance α: 5% (or) 0.05 (Assume)

4) Test statistic $z = \frac{x - \mu}{\sigma} = \frac{150 - 128}{8} = \frac{22}{8} = 2.75$

$|z| = |2.75|$

$|z| = 2.75$

5) Conclusion:

Critical value at 5% level of significance $z_\alpha = 1.96$

Here $|z| > z_\alpha$ $(2.75 > 1.96)$.

$\therefore$ H_0 is Rejected.

We conclude that die is biased.

3) In a test 2000 electrical bulbs, it was found the life of a perticular make, was normally distributed an average life of 2040 hours and S.D of 40 hrs. Estimate the number of bulbs likely to burn for more than 2140.

<u>Sol</u>:- Given $\mu = 2040$ hrs

S.D $\sigma = 40$ hrs.

$x = 2140$

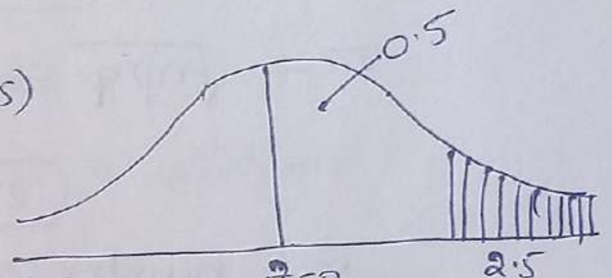
$$z = \frac{x-\mu}{\sigma} = \frac{2140 - 2040}{40} = \frac{\cancel{100}^{5}}{\cancel{40}_{2}} = 2.5 > 0$$

$$P(x > 2140) = P(z > 2.5)$$

$$= 0.5 - P(0 \le z \le 2.5)$$

$$= 0.5 - 0.4938$$

$$= 0.0062$$



Hence the number of bulbs likely to burn for more than

2140 hrs $= 0.0062 \times 2000 = 12.4$

$\cong 12$

<u>H.W</u>

(4) A die is thrown 960 times and it falls with 5 upwards 184 times. Is the die unbiased at a level of significance of 0.01?

<u>H.W</u>

(5) A coin was tossed 400 times and returned heads 216 times. Test the hypothesis that the coin is unbiased. Use a 0.05 level of significance.

Test of Hypothesis of a Single Mean - Large Samples

Let a random sample of size n $(n \geq 30)$ has the sample mean $\bar{x}$, and μ be the population mean.

Also the population mean μ has a specified value μ_0.

1) Null Hypothesis H_0: $\bar{x} = \mu$, "there is no significance difference between the sample mean and population mean", (or) the sample has been drawn from the parent population.

2) Alternative Hypothesis H_1: (i) $H_1: \bar{x} \neq \mu$ $(\mu = \mu_0)$ (Two-tailed Test)
(ii) $H_1: \bar{x} > \mu$ $(\mu > \mu_0)$ (Right-tailed Test)
(iii) $H_1: \bar{x} < \mu$ $(\mu < \mu_0)$ (Left-tailed Test).

3) Level of significance (α):

4) Test statistic: (i) when S.D σ of population is known.

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

(ii) when S.D σ of population is not known.

$$\boxed{z = \frac{\bar{x} - \mu}{s/\sqrt{n}}}$$

in this case, we take s, the standard deviation of sample.

Find the critical value Z_α of z at the level of significance α from the normal table.

5) Decision:-

(i) $|z| < z_\alpha$, we accept the Null Hypothesis H_0.

(ii) $|z| > z_\alpha$, we reject the Null Hypothesis H_0.

Note:- Confindence Interval (or) Fudicial Interval

1) 95% confidence interval for single mean

$$\boxed{\left(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)}$$

$z_{\frac{\alpha}{2}}$ for $95\% = 1.96$

$$\left(\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}\right)$$

2) 98% Confidence interval for single mean

$$\cancel{\bar{x} - 2.58} \left(\bar{x} - 2.33 \frac{\sigma}{\sqrt{n}}, \bar{x} + 2.33 \frac{\sigma}{\sqrt{n}}\right)$$

3) 99% Confidence interval for ~~sign~~ single mean

$$\left(\bar{x} - 2.58 \frac{\sigma}{\sqrt{n}}, \bar{x} + 2.58 \frac{\sigma}{\sqrt{n}}\right).$$

Problems:

1) 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. Also calucculate 95% confidence interval for the population.

Sol:- Given $n = 400$

(sample mean) $\bar{x} = 40$

(population mean) $\mu = 38$

population S.D $\sigma = 10$

1) Null Hypothesis H_0 : $\mu = 38$

2) Alternative hypothesis H_1: $\mu \neq 38$ (Two-tailed test).

3) Level of significance α: 5% (Assume)

4) Test statistic : $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

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

$|z| = |4| = 4.$

critical value at 5% level of significance $z_\alpha = 1.96$ (from Normal table)

5) Conclusion: $|z| = 4$, $z_\alpha = 1.96$

$\therefore$ $|z| > z_\alpha$. $\therefore$ Null Hypothesis H_0 Rejected.

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

$$= \left(40 - 1.96\left(\frac{10}{\sqrt{400}}\right)\right), \left(40 + 1.96\left(\frac{10}{\sqrt{400}}\right)\right)$$

$$= \left(40 - \frac{1.96}{20}\times 10, 40 + \frac{1.96}{20}\times 10\right)$$

$$= (40 - 0.98, 40 + 0.98)$$

95% confidence interval $= (39.02, 40.98)$

2). According to the norms established for a mechanical aptitude test, persons who are 18 years old have an average height of 37.2 with a standard deviation of 8.6. If 4 randomly selected persons of that age averaged 76.7, test the hypothesis $\mu = 73.2$ against the alternative hypothesis $\mu > 73.2$ at the 0.01 level of significance.

Sol:- Given $n = 4$, $\mu = 73.2$

Sample mean $\bar{x} = 76.7$

$\sigma = 8.6$

Null Hypothesis H_0: $\mu = 73.2$

Alternative hypothesis H_1: $\mu > 73.2$ (Right-tailed test)

Level of significance α: 0.01 (or) 1%

Test statistic $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{76.7 - 73.2}{\frac{8.6}{\sqrt{4}}} = \dfrac{3.5}{4.3} = 0.814$

Critical value (Tabulated value) at 1% Level $z_\alpha = 2.33$

Conclusion: $|z| = 0.814$, $z_\alpha = 2.33$

$|z| < z_\alpha$ Hence H_0 is Accepted.

ie $\bar{x}$ and μ do not differ significantly.

(10)

In 64 randomly selected hours of production, the mean and the standard deviation of the number of acceptance pieces by an automatic stamping machine are $x = 1.038$ and $\sigma = 0.146$

At the 0.05 level of significance does this enable us to reject the null hypothesis $\mu = 1.000$ against the alternative hypothesis $\mu > 1.000$?

Sol:- Given $\bar{x} = 1.038$

$\sigma = 0.146$

$\mu = 1.000$

$n = 64$. (large sample)

Null Hypothesis H_0: $\mu = 1.000$

Alternative Hypothesis H_1: $\mu > 1.000$ (Right tailed Test)

Level of significance α : 0.05

Test statistic $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$$= \frac{1.038 - 1.000}{\frac{0.146}{\sqrt{64}}} = 2.082$$

$|z| = |2.082| = 2.082.$

critical value at 0.05 level of significance $z_\alpha = 1.645$ (Right-tailed Test)

Conclusion: $|z| = 2.08$, $z_\alpha = 1.645$

$|z| > z_\alpha$

Hence H_0 is Rejected.

The mean of the population $\mu > 1.000$.

4) 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:- Given $\bar{x} = 70$

$\mu = 56$

$\sigma = 25$

$n = 64$

Null Hypothesis H_0: A sample of 64 students with mean weight 70 kgs can be regarded as a sample from a population with mean weight 56 kgs and S.D 25

Alternative Hypothesis H_1: Sample can't be regarded as one coming from the population.

Level of significance α : 0.05 (Assumption).

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

$|z| = |4.48| = 4.48.$

Critical value $z_\alpha = 1.96$ at 5% level.

Conclusion :- $|z| = 4.48, \quad z_\alpha = 1.96.$

$|z| > z_\alpha$

Hence H_0 is Rejected.

H.W

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

Test For Equality of Two Means - 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 μ_1 and μ_2 and standard deviations σ_1 and σ_2.

To test whether the two population means are equal.

Let the Null Hypothesis $H_0 : \mu_1 = \mu_2$

Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$.

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

Note :- (1) If the samples have been drawn from the population with common S.D σ then $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Hence Test statistic $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}}} = \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\dfrac{1}{n_1^2} + \dfrac{1}{n_2^2}}}$.

→ If σ is not known we can use estimate of σ^2 given by $\sigma^2 = \dfrac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$.

Note :- (2) If the two samples are drawn from two populations with unknown standard deviations σ_1^2 and σ_2^2, then σ_1^2 and σ_2^2 can be replaced by sample variances S_1^2 and S_2^2 provided both the samples n_1 and n_2 are large.

In this case, the test statistic $\boxed{z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}}$ (or)

Problems :

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

Sol:- Let μ_1 and μ_2 be the means of the two populations.

Given $n_1 = 1000$, $n_2 = 2000$

$\bar{x}_1 = 67.5$, $\bar{x}_2 = 68$

population S.D $\sigma = 2.5$ inches

Null Hypothesis H_0: The samples have been drawn from the same population of S.D 2.5 inches.

ie $\mu_1 = \mu_2$, $\sigma = 2.5$

Alternative Hypothesis H_1: $\mu_1 \neq \mu_2$, Level of significance $\alpha = 0.05$ level.

Test statistic $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \dfrac{67.5 - 68}{\sqrt{(2.5)^2\left(\frac{1}{1000} + \frac{1}{2000}\right)}}$

$z = \dfrac{-0.5}{0.0968} = -5.16$

$|z| = |-5.16| = 5.16.$

Critical value at 5% level of significance $z_\alpha = 1.96$

$\therefore$ $|z| > z_\alpha$

$\therefore$ H_0 is Rejected.

Samples of students were drawn from two universities and from their weights in kilograms, mean and standard deviations are caluculated and shown below. Make a large sample test to test the significance of the difference between the means.

	Mean	Standard deviation	Size of the sample
University A	55	10	400
University B	57	15	100

Sol:- Given $n_1 = 400$, $n_2 = 100$

$\bar{x}_1 = 55$, $\bar{x}_2 = 57$

$s_1 = 10$, $S_2 = 15$

Null Hypothesis H_0: $\bar{x}_1 = \bar{x}_2$ (there is no difference)

Alternative Hypothesis H_1: $\bar{x}_1 \neq \bar{x}_2$ (Two-tailed Test)

Level of significance α: 0.05 (level)

Test statististic $$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{55 - 57}{\sqrt{\frac{(10)^2}{400} + \frac{(15)^2}{100}}}$$

$$z = \frac{-2}{\sqrt{\frac{1}{4} + \frac{9}{4}}} = -1.26$$

$$|z| = |-1.26| = 1.26.$$

Conclusion: Critical value at 0.05 level of significance in two-tailed test $z_\alpha = 1.96$.

$\therefore |z| < z_\alpha$ $\therefore$ H_0 is Accepted.

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

Sol:- Given $n_1 = 150$, $n_2 = 250$

$\bar{x}_1 = 75$, $\bar{x}_2 = 70$

$\sigma_1 = 15$, $\sigma_2 = 20$

Null Hypothesis H_0: The groups have been come from the same population

ie $\mu_1 = \mu_2$

Alternative Hypothesis H_1: $\mu_1 \neq \mu_2$

Level of significance α: 0.05 (Assume)

Test statistic $$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{75-70}{\sqrt{\frac{(15)^2}{150} + \frac{(20)^2}{250}}}$$

$$z = \frac{5}{\sqrt{\frac{9}{5} + \frac{8}{5}}} = 2.7116.$$

$|z| = |2.7116| = 2.7116.$

Conclusion: Critical value $z_\alpha = 1.96$ at 5% level

$\therefore$ $|z| = 2.7116$, $z_\alpha = 1.96$.

$\therefore |z| > z_\alpha$.

$\therefore$ H_0 is Rejected.

In a certain factory there are independent processes for manufacturing the same item. The average weight in a sample of 700 items produced from one process is found to be 250 gms with a standard deviation of 30 gms while the corrsponding sample of 300 items from the other process are 300 and 40. Is there significant difference between the mean at 1% level?

Sol:- Let the average weight in the two independent processes be μ_1 & μ_2 respectively.

Null Hypothesis $H_0 : \mu_1 = \mu_2$

Alternative Hypothesis $H_1 : \mu_1 \neq \mu_2$

Level of significance $\alpha = 1\%$.

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

given $\bar{x}_1 = 250, \; n_1 = 700, \; \sigma_1 = 30$

$\bar{x}_2 = 300, \; n_2 = 300, \; \sigma_2 = 40$

$$z = \frac{250 - 300}{\sqrt{\frac{(30)^2}{700} + \frac{(40)^2}{300}}} = \frac{-50}{\sqrt{\frac{9}{7} + \frac{16}{93}}} = -19.43$$

$|z| = |-19.43| = 19.43.$

Critical value at 1% level of two tailed test $z_\alpha = 2.5$

$\therefore$ $|z| = 19.43, \quad z_\alpha = 2.58.$

$\therefore$ $|z| > z_\alpha$

$\therefore$ H_0 is Rejected.

imp

H.W (5) The mean Height of ~~of~~ 50 male students who participated in sports is 68.2 inches with a S.D of 2.5. The mean height of 50 male students who have not participated in spoot is 67.2 inches with a S.D of 2.8. Test the hypothesis that the height of students who participated in spools is more than the students who have not participate in sports.

6). Two types new cars prduced in U.S.A. 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 variances as $\sigma_1^2 = 2.0$ and $\sigma_2^2 = 1.5$ respectively. Test whether there is any significance difference in the petrol consumption of these two types of cars (use $\alpha = 0.01$).

—— * ——

Test of Significance (Hypothesis) for Single-Proportion (Large-Samples)

Suppose a large random sample of size 'n' has a sample proportion p of members possessing a certain attribute (ie proportion of successes).

To test the hypothesis that the proportion P in the population has a specified value P_0.

Let the Null Hypothesis $H_0 : P = P_0$ (P_0 is a perticular value of P).

Alternative Hypothesis $H_1 : P \neq P_0$ (Two tailed-Test)

$H_1 : P > P_0$ (Right-tailed test)

$H_1 : P < P_0$ (Left-tailed Test)

Test statistic $z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}}$ where p is sample propotion.

is approximately Normally distributed.

<u>Note</u>:- (1) Confidence interval for proportion P for large sample at 'α' level of significance

$$\left(p - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{pq}{n}} \, , \; p + z_{\frac{\alpha}{2}} \sqrt{\frac{pq}{n}} \right)$$

(2) $z_{\frac{\alpha}{2}} = 1.96$ (for 95%),

$z_{\frac{\alpha}{2}} = 2.33$ (for 98%), $z_{\frac{\alpha}{2}} = 2.58$ (for 99%).

Problems :-

(1) 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 $n = 500$, $x = 100$

Sample proportion $p = \frac{x}{n} = \frac{100}{500} = \frac{1}{5} = 0.2$

and $P = 4\% = 0.04$

$Q = 1 - P = 1 - 0.04 = 0.96$.

Null Hypothesis $H_0 : P = 0.04$

Alternative hypothesis $H_1 : P \neq 0.04$

Level of significance $\alpha : 0.05$

Test statistic $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.2 - 0.04}{\sqrt{\frac{(0.04)(0.96)}{500}}}$

$$Z = \frac{-0.16}{\sqrt{\frac{(0.04)(0.96)}{500}}} = -18.26$$

$|Z| = |-18.26| = 18.26$.

Critical value at 5% level of significance of Two-tailed test $Z_\alpha = 1.96$.

$\therefore |Z| = 18.26$, $Z_\alpha = 1.96$

$\therefore |Z| > Z_\alpha$

$\therefore H_0$ is Rejected.

problems :-

(1) 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 $n = 500$, $x = 100$

Sample proportion $p = \frac{x}{n} = \frac{100}{500} = \frac{1}{5} = 0.2$

and $P = 4\% = 0.04$

$Q = 1 - P = 1 - 0.04 = 0.96.$

Null Hypothesis H_0: $P = 0.04$.

Alternative Hypothesis H_1: $P \neq 0.04$

Level of significance α: 0.05

Test statistic $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.2 - 0.04}{\sqrt{\frac{(0.04)(0.96)}{500}}}$

$z = \frac{-0.16}{\sqrt{\frac{(0.04)(0.96)}{500}}} = -18.26$

$|z| = |-18.26| = 18.26.$

Critical value at 5% level of significance of Two-tailest test $z_\alpha = 1.96$.

$\therefore |z| = 18.26$, $z_\alpha = 1.96$

$\because |z| > z_\alpha$

$\therefore H_0$ is Rejected.

(15)

) 40 people were attacked by a disease and only 18 survived. Will you Reject the hypothesis that survival rate if attacked by this disease is 85% in favour of the hypothesis that is more at 5% level.?

Sol:- given Sample size $n = 40$

$x =$ number of survived people $= 18$

$p =$ proportion of survived people $= \frac{x}{n} = \frac{18}{40} = 0.45$

$P = 85\% = 0.85$

$Q = 1 - 0.85 = 0.15$

Null Hypothesis H_0: $P = 0.85$

Alternative Hypothesis H_1: $P > 0.85$. (Right-tailed test)

Level of significance α: 0.05 (or) 5%

Test statistic $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.45 - 0.85}{\sqrt{\frac{(0.85)(0.15)}{40}}}$

$$Z = \frac{-0.4}{\sqrt{0.00318}} = \frac{-0.4}{0.05639} = -7.0$$

$|z| = |-7.09| = 7.09$

Critical value at 5% level of Right tailed Test $Z_\alpha = 1$

$\therefore |z| > Z_\alpha$

$\therefore$ Null Hypothesis H_0 is Rejected.

(3) A manufacturer claimed that atleast 95% of the equipment which he supplied to factory conformed to specifications. An examination of sample of 200 pieces of equipment revealed that 18 were faulty. Test his ~~claim~~ claim at 5% level of significance. (given z value for 5% level = 1.645)

Sol:- Given Sample size $n = 200$

Number of pieces confirming to specification = $200 - 18 = 182$

p = proportion of piecess confirming to specifications

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

population proportion $P = 95\% = 0.95$

$$Q = 1 - 0.95 = 0.05$$

Null Hypothesis H_0: $P = 0.95$ (The proportion of pieces confirming to specifications).

Alternative hypothesis H_1: $P < 0.95$ (left-tailed Test)

Test statistic $z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0.91 - 0.95}{\sqrt{\dfrac{(0.95)(0.05)}{200}}}$

$$z = -2.59$$

$$|z| = |-2.59| = 2.59$$

Critival value at 5% level $z_\alpha = 1.645$

$\therefore$ $|z| > z_\alpha$

Null Hypothesis H_0 is Rejected.

4) 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 percentage.

Sol:- Given $n = 160$

$x = 24$

proportion $p = \frac{x}{n} = \frac{24}{160} = 0.15$

$q = 1 - 0.15 = 0.85$

Confidence interval for 99% $= \left(p - z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}},\ p + z_{\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}\right)$

$$\boxed{z_{\frac{\alpha}{2}} \text{ for } 99\% = 3.}$$

$$= \left(0.15 - 3\sqrt{\frac{(0.15)(0.85)}{160}},\ 0.15 + 3\sqrt{\frac{(0.15)(0.85)}{160}}\right)$$

$$= (0.15 - 3 \times 0.028,\ 0.15 + 3 \times 0.028)$$

$$= (0.065, 0.234)$$

Confidence interval (or) Fudicial Interval $= (0.065, 0.234)$

Note:- Confidence limit (or) Fudicial limits $= 0.065, 0.234$.

5) Experience had shown that 20% of a manufactured product is of the top quality. In one days production of 400 ~~are~~ articales only 50 are of top quality. Test the hypothesis at 0.05 level.

Sol:- Given $n = 400$, $x = 50$

$$p = \frac{x}{n} = \frac{50}{400} = \frac{1}{8} = 0.125$$

$$P = 20\% = 0.2, \quad Q = 1 - 0.2 = 0.8$$

Null Hypothesis H_0: $P = 0.2$

Alternative Hypothesis H_1: $P \neq 0.2$

Level of significance $\alpha = 0.05$

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

$$|z| = |-3.75| = 3.75$$

Criti value at 5% level of two-tailed test $z_\alpha = 1.96$

$|z| > z_\alpha$ $\therefore$ we Reject the Null Hypothesis H_0.

H.W.

(6) A random sample of 500 Apples was taken from a large consignment of 60 were found to be bad, obtain the 98% confidence limits for the percentage number of bad apples in the consignment.

H.W.

(7) In a sample of 1000 people in Karnataka 540 are rice eaters and the rest are wheat eaters. Can we assume the both rice and wheat are equally popular in this stat at 1% level of significance.

Test of Hypothesis for Equality of Two-Proportions

(OR)

Large-Samples - Test of Hypothesis for Difference of Proportion

Let p_1 and p_2 be the sample 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 samples proportions have been drawn the same population.

Null Hypothesis $H_0 : P_1 = P_2$

Alternative Hypothesis $H_1 : P_1 \neq P_2$

Test statistic $\boxed{z = \dfrac{p_1 - p_2}{\sqrt{\dfrac{P_1 Q_1}{n_1} + \dfrac{P_2 Q_2}{n_2}}}}$ (when P_1 and P_2 are known).

Note:- imp (1) when population proportions P_1 and P_2 are not known but ~~sam~~ sample proportions are known p_1 and p_2

test statistic $\boxed{z = \dfrac{p_1 - p_2}{\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}}}$

imp (2) In some cases, the estimated value for the two population proportions is obtained by pooling method the two sample proportions p_1 and p_2 into a single proportion p by the formula $p = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$, $q = 1 - p$.

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

Problems :-

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 sample size ~~of town~~ $n_1 = 1000$ (town A)

sample size $n_2 = 800$ (town B)

Number of consumers of wheat from town A $= x_1 = 400$

" " " " B $= x_2 = 400$

P_1 = Proportion of consumers of wheat in town A $= \frac{x_1}{n_1} = \frac{400}{1000} = 0.4$

P_2 = Proportion of consumers of wheat in town B $= \frac{x_2}{n_2} = \frac{400}{800} = 0.5$

we know $n_1 = 1000$, $P_1 = 0.4$ then $p = \frac{n_1 \cdot P_1 + n_2 \cdot P_2}{n_1 + n_2}$

$n_2 = 800$ $P_2 = 0.5$

$$p = \frac{(1000)(0.4) + 800(0.5)}{1000 + 800}$$

$$p = \frac{400 + 400}{1800} = \frac{\cancel{800}}{\cancel{1800}}$$

$$p = \frac{\cancel{8}4}{\cancel{18}9} = \frac{4}{9}$$

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

Null Hypothesis H_0: $P_1 = P_2$ (ie there is no difference)

Alternative Hypothesis H_1: $P_1 \neq P_2$

Level of significance α: 0.05

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

$$z = \frac{0.4 - 0.5}{\sqrt{\frac{4}{9}\times\frac{5}{9}\left(\frac{1}{1000}+\frac{1}{800}\right)}}$$

$$z = \frac{-0.1}{\sqrt{\frac{20}{81}\times\frac{1800}{(1000)(800)}}}$$

$$z = -4.242$$

$$|z| = |-4.242| = 4.242$$

critical value at 5% level $z_\alpha = 1.96$

$\therefore$ $|z| = \cancel{4.242}, 4.242, \quad z_\alpha = 1.96$

$|z| > z_\alpha$

$\therefore$ Null Hypothesis H_0 is Rejected.

2). 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.

Sol:- Given $n_1 = 1200$, $n_2 = 900$

P_1 = proportion of fair haired people in the first population

$$P_1 = \frac{30}{100} = 0.3, \quad Q_1 = 1-0.3 = 0.7$$

P_2 = Proportion of fair haired people in the second population = 25%

$$P_2 = \frac{25}{100} = 0.25, \quad Q_2 = 1-0.25 = 0.75$$

Null Hypothesis H_0: $P_1 = P_2$ (Assume that the sample proportions are equal, ie the difference of population proportions is likely to be hidden in sampling).

Alternative Hypothesis H_1: $P_1 \neq P_2$.

Level of significance α: 5% (Assume)

Test statistic
$$z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

$$z = \frac{0.3 - 0.25}{\sqrt{\frac{(0.3)(0.7)}{1200} + \frac{(0.25)(0.75)}{900}}}$$

$$z = \frac{0.05}{0.0195} = 2.56$$

$$|z| = |2.56| = 2.56$$

Critical value at 5% level $z_\alpha = 1.96$

$\therefore |z| > z_\alpha$

$\therefore$ Null Hypothesis H_0 is Rejected.

P_1 = proportion of fair haired people in the first population

$$P_1 = \frac{30}{100} = 0.3, \quad Q_1 = 1 - 0.3 = 0.7$$

P_2 = Propotion of fair haired people in the second population = 25

$$P_2 = \frac{25}{100} = 0.25, \quad Q_2 = 1 - 0.25 = 0.75$$

Null Hypothesis H_0: $P_1 = P_2$ (Assume that the sample proportions are equal. ie the difference of population proportions is likely to be hidden in sampling).

Alternative Hypothesis H_1: $P_1 \neq P_2$.

Level of significance α: 5% (Assume).

Test statistic $z = \dfrac{p_1 - p_2}{\sqrt{\dfrac{P_1 Q_1}{n_1} + \dfrac{P_2 Q_2}{n_2}}}$

$$z = \frac{0.3 - 0.25}{\sqrt{\dfrac{(0.3)(0.7)}{1200} + \dfrac{(0.25)(0.75)}{900}}}$$

$$z = \frac{0.05}{0.0195} = 2.56$$

$$|z| = |2.56| = 2.56$$

Critical value at 5% level $z_\alpha = 1.96$

$\therefore |z| > z_\alpha$

$\therefore$ Null Hypothesis H_0 is Rejected.

In an investigation on the machine performance the following results are obtained.

	Number of units inspected	No. of defectives
Machine 1.	375	17
Machine 2.	450	22

Test whether there is any significant performance of two machines at $\alpha = 0.05$.

Let P_1 and P_2 be the proportions of defective units in the population of units inspected in Machine 1 and Machine 2 respectively.

Given $n_1 = 375$, $\quad n_2 = 450$

$x_1 = 17$, $\quad x_2 = 22$

$$P_1 = \frac{17}{375} = 0.045, \quad P_2 = \frac{x_2}{n_2} = \frac{22}{450} = 0.049$$

$(P_1 > P_2)$

Null Hypothesis $H_0 : P_1 = P_2$

Alternative Hypothesis $H_1 : P_1 \neq P_2$

Level of significance $\alpha : 0.05$

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

where
$$p = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{(375)(0.045) + (450)(0.049)}{375 + 450}$$

$$p = \frac{39}{825} = 0.047, \quad q = 1 - 0.047 = 0.953$$

then $z = \dfrac{0.045 - 0.049}{\sqrt{(0.047)(0.953)\left(\frac{1}{375} + \frac{1}{450}\right)}}$

$z = -0.267$

$|z| = |-0.267| = 0.267$

Critical value at 5% level of significance at two-t test $z_\alpha = 1.96$

$|z| = 0.267, \quad z_\alpha = 1.96$

$\therefore \quad |z| < z_\alpha$

Then Null Hypothesis H_0 is Accepted.

We conclude that there is no significant difference in performance of two machines.

4). Random samples of 400 men and 600 women were asked whethere they would like to have a flyover near their residen 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 $n_1 = 400$, $n_2 = 600$, $x_1 = 200$, $x_2 = 3$

proportion of men $P_1 = \dfrac{x_1}{n_1} = \dfrac{200}{400} = 0.5$

$P_2 = \dfrac{x_2}{n_2} = \dfrac{325}{600} = 0.541$

we have $p = \dfrac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \dfrac{(400) \times (0.5) + (600)(0.541)}{400 + 600}$

$p = \dfrac{525}{1000} = 0.525, \quad q = 1 - 0.525 = 0.475$

H_0: Assume that there is no
significant difference between the option
of men and women as far as proposal of
flayover is concerned
i.e $P_1 = P_2$

hesis H_1: $P_1 \neq P_2$ (Two-tailed test).

nce α: 5%

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.5 - 0.541}{\sqrt{(0.525)(0.425)\left(\frac{1}{400} + \frac{1}{600}\right)}}$$

$$= \frac{-0.041}{0.032} = -1.28$$

$|-1.28| = 1.28$

e at 5% level $z_\alpha = 1.96$

z_α

$\therefore$ Ho is Accepted.

e population proportions P_1 and P_2 are given
. If we want to test the hypothesis that
$_1 - P_2)$ in population proportions is likely to be
les of sizes n_1 and n_2 from the two
tively, then test statistic $\boxed{z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}}$

(20)

Null Hypothesis H_0: Assume that there is no significant difference between the option of men and women as far as proposal of flayover is concerned.

ie $P_1 = P_2$

Alternative Hypothesis H_1: $P_1 \neq P_2$ (Two-tailed test).

Level of significance α: 5%

Test statistic $z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \dfrac{0.5 - 0.541}{\sqrt{(0.525)(0.425)\left(\frac{1}{400} + \frac{1}{600}\right)}}$

$$z = \frac{-0.041}{0.032} = -1.28$$

$$|z| = |-1.28| = 1.28$$

Critical value at 5% level $z_\alpha = 1.96$

$\therefore$ $|z| < z_\alpha$

$\therefore$ H_0 is Accepted.

Note:- Suppose the population proportions P_1 and P_2 are given and $P_1 \neq P_2$. If we want to test the hypothesis that the difference $(P_1 - P_2)$ in population proportions is likely to be hidden in samples of sizes n_1 and n_2 from the two populations respectively, then test statistic $\boxed{z = \dfrac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\dfrac{P_1 Q_1}{n_1} + \dfrac{P_2 Q_2}{n_2}}}}$

Null Hypothesis H_0: Assume that there is no significant difference between the option of men and women as far as proposal of flayover is concerned.

ie $P_1 = P_2$

Alternative Hypothesis H_1: $P_1 \neq P_2$ (Two-tailed test).

Level of significance α: 5%

Test statistic $$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.5 - 0.541}{\sqrt{(0.525)(0.425)\left(\frac{1}{400} + \frac{1}{600}\right)}}$$

$$z = \frac{-0.041}{0.032} = -1.28$$

$$|z| = |-1.28| = 1.28$$

Critical value at 5% level $z_\alpha = 1.96$

$\therefore$ $|z| < z_\alpha$

$\therefore$ H_0 is Accepted.

Note:- Suppose the population proportions P_1 and P_2 are given and $P_1 \neq P_2$. If we want to test the hypothesis that the difference $(P_1 - P_2)$ in population proportions is likely to be hidden in samples of sizes n_1 and n_2 from the two populations respectively, then test statistic $$\boxed{z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}}$$

→ If sample proportions are not known, then we use

$$|z| = \frac{|P_1 - P_2|}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}}$$

5) A cigarette manufacturing firm claims that its brand A line of cigarettes outsells its brand B by 8%. If it is found that 42 out of a sample of 200 smokers prefer brand A and 18 out another sample of 100 smokers prefer brand B, test whether the 8% difference is a valid claim.

Sol:- Given $n_1 = 200$, $n_2 = 100$, $x_1 = 42$, $x_2 = 18$

$$P_1 = \frac{42}{200} = 0.21, \quad P_2 = \frac{18}{100} = 0.18$$

and $P_1 - P_2 = 8\% = \frac{8}{100} = 0.08$

Null Hypothesis H_0: Assume that 8% difference in the sale of two brands of cigarettes is a valid claim ie

$$H_0: P_1 - P_2 = 0.08$$

Alternative Hypothesis $H_1: (P_1 - P_2) \neq 0.08$

Level of significance α: 5%

Test statistic $$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{(0.21)(200) + (0.18)(100)}{200 + 100}$$

$$p = 0.2, \quad q = 1 - 0.2 = 0.8.$$

(21)

Test statistic $z = \dfrac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$z = \frac{0.03 - 0.08}{\sqrt{(0.2)(0.8)\left(\frac{1}{200} + \frac{1}{100}\right)}}$$

$$z = \frac{-0.05}{0.0489} = -1.02$$

$$|z| = |-1.02| = 1.02$$

Critical value at 5% level $z_\alpha = 1.96$.

$\therefore$ $|z| = 1.02$, $z_\alpha = 1.96$

$\therefore$ Null Hypothesis H_0 is Accepted.

ie 8% difference in the sale of two brands of cigarettes is a valid claim.

H.W

6) A machine produced 20 defective articles in a batch of 400. After overhaulting it produced 10 defectives in a batch of 300. Has the machine being improved after overhauling?

7). During a country wide investigation the incidence of tuberculosis was found to be 1%. In a college of 400 students 3 reported to be affected, where as in another college of 1200 students 10 were affected. Does this indicate any significant difference?

8). A random sample of 300 shoppers at a supermarket includes 204 who regularly use cents of coupons. In another sample 500 shoppers at a supermarket includes 75 who regularly use cents of coupons, construct (98%) confidence interval for the probability that any one shopper at the supermarket, selected at random, will regularly use cents of coupons.

Sol:- Given $n_1 = 300$, $n_2 = 500$

$x_1 = 204$, $x_2 = 75$

P_1 = proportion of shoppers who use cents off coupons in the first sample $= \frac{204}{300} = 0.68$

P_2 = Proportion of shoppers who use cents off coupons in the second sample $= \frac{75}{500} = 0.15$

98% Confidence interval for two sample proportions is

$$\left(P_1 - Z_{\frac{\alpha}{2}}\sqrt{\frac{P_1 q_1}{n_1}},\ P_1 + Z_{\frac{\alpha}{2}}\sqrt{\frac{P_1 q_1}{n_2}}\right)$$

$Z_{\frac{\alpha}{2}}$ for 98% $= 2.33$.

$P_1 = 0.68$, $q_1 = 1 - P_1 = 1 - 0.68 = 0.32$

$P_2 = 0.15$, $q_2 = 1 - q_2 = 1 - 0.15 = 0.85$

$$= \left(0.68 - 2.33\left(\sqrt{\frac{0.68 \times 0.32}{300}}\right),\ 0.68 + 2.33\sqrt{\frac{0.68 \times 0.32}{300}}\right)$$

$$= (0.68 - (0.063),\ 0.68 + 0.063)$$

Confidence Interval $= (0.62, 0.74)$