Unit - III

Estimation and Testing of hypothesis (large Samples)

— Estimate :- A ~~sugt~~ sentences given to unknown Parameter and estimate is a statement made to find an Unknown population parameter

→ Estimator :- To estimate the value of ~~the~~ A population Parameter you can use information from the sample in the form of an estimator.

- Estimator are used in 2 different ways

① point estimator

② Intervel estimator

① point estimator :- based on sample data a single num is calculated to estimate based ~~the~~ the population parameter the rule (or) formula that discribes ~~the~~ calculation is called point estimator and resulting num is called point estimate.

② Intervel estimator :- based on sample data 2 nums are calculated to form an intervel with in which the Parameter is excepted to lie.

- The rule or formula that discribes. This calculation called intervel estimator and the resulting pair of nums is called an intervel estimate (or) Confidence intervel

— unbiased (good estimate) and biased estimate

— A Statistic is said to unbiased estimator of corresponding parameter if the mean of sampling distribution of the statistic is equal to the corresponding population parameter. otherwise the statistic is called biased estimator

- if 't' be the statistic and θ be the corresponding parameter, $E(t) = \theta$, then 't' is unbiased estimator of 'θ' otherwise 't' is biased estimator of 'θ' and the biased is $E(t) - \theta$

Problem

Q.P.T for a random sample of size n, $x_1, x_2, \dots, x_n$ taken from an infinite population $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is not unbaised estimator of the parameter σ^2 but $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is unbiased.

Sol

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

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

$$= \frac{\sum x_i^2}{n} - 2\bar{x}^2 + \bar{x}^2 = \frac{\sum x_i^2}{n} - \bar{x}^2$$

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

$$= \frac{1}{n}\left[\sum_{i=1}^{n}x_i^2 - 2\bar{x}\cdot n\sum_{i=1}^{n}\frac{x_i}{n} + \bar{x}^2\cdot n\right] \left[\because \bar{x} = \frac{\sum x_i}{n}\right]$$

$$= \frac{1}{n}\left[\sum_{i=0}^{n}x_i^2 - 2n\bar{x}\cdot x^2 + n\bar{x}^2\right]$$

$$= \frac{1}{n}\left[\sum_{i=0}^{n}x_i^2 - 2\bar{n}x^2 + \bar{n}x^2\right]$$

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

I.S 6.7

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

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

$= \frac{1}{n}\left[\sum_{i=1}^{n}(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right]$

$= \frac{1}{n}\left[n(\sigma^2+\mu^2) - \sigma^2 - n\mu^2\right]$

$= \frac{1}{n}\left[n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2\right]$

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

$var(x_i) = E(x_i^2) - E(x_i)^2$

$\sigma^2 = E(x_i)^2 - \mu^2$

$E(x_i)^2 = \sigma^2 + \mu^2$

$(var(\bar{x}) = E(\bar{x}^2) - E(\bar{x})^2$

$\frac{\sigma^2}{n} = E(\bar{x})^2 - \mu^2$

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

$E(\bar{x}^2) = \frac{\sigma^2}{n} - \mu^2$

$E(s^2) = \frac{n-1}{n}\sigma^2$ is biased estimator of σ^2

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is unbiased estimate of σ^2

i.e. $E(s^2) = \sigma^2$ //

Pg 303 - 1

Show that s^2 is an unbiased estimator of the Parameter σ^2 //

$\Rightarrow s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

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

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

$= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i - 2n\sum_{i=1}^{n}\frac{x}{n} + \bar{x}^2 \cdot n\right]$ $\left[\bar{x} = \frac{\sum x_i}{n}\right]$

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

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

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

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

$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right]$

$= \frac{1}{n-1}\left[n(\sigma^2+\mu^2) - \sigma^2 - n\mu^2\right]$

$= \frac{1}{n-1}\left[n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2\right]$

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

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

$= \sigma^2$ //

$var(x_i) = E(x_i^2) - E(x_i)^2$

$\sigma^2 = E(x_i)^2 - \mu^2$

$E(x_i)^2 = \sigma^2 + \mu^2$

$var(\bar{x}) = E(\bar{x})^2 - E(\bar{x})^2$

$\frac{\sigma^2}{n} = E(\bar{x})^2 - \mu^2$

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

$E(\bar{x}^2) = \frac{\sigma^2}{n} - \mu^2$

Confidence interval for μ, σ known

if $\bar{x}$ is the mean of a random simple of size n from the population with known variance σ^2

$(1-\alpha)100\%$.

— confidence interval for μ is given

$$\boxed{\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}$$

*where $Z_{\frac{\alpha}{2}}$ is Z value ~~leaving a ve~~ leaving an area of $\frac{\alpha}{2}$

so, the maximum error of estimate E with $(1-\alpha)$ Probability is given by $E = Z_{\alpha/2} \cdot \left[\frac{\sigma}{\sqrt{n}}\right]$

→ saple size $n = \left(Z_{\frac{\alpha}{2}} \cdot \frac{1}{E}\right)^2 PQ$ when P and Q are known. (σ is not known).

μ & σ 90%

$(1-\alpha)100\% = 90\%$

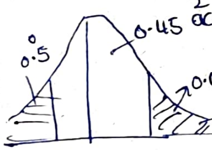
$(1-\alpha)10\cancel{0} = 9\cancel{0}$

$1-\alpha = \frac{9}{10}$

$1-\alpha = 0.9$

$\alpha = 0.1$

$\frac{\alpha}{2} = \frac{0.1}{2} = 0.05$



$\int_0^{Z_{\alpha/2}} \phi(z)\,dz = 0.45$

where $Z = 1.65$

95%

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

$(1-\alpha)100 = 95$

$1-\alpha = \frac{95}{100}$

$1-\alpha = 0.95$

$\alpha = 0.05$

$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

$\int_0^{Z_{\alpha/2}} \phi(z)\,dz = 0.495$

where $Z_{\frac{\alpha}{2}} = 1.96$

99

$(1-\alpha)100\% = 99$

$(1-\alpha)100 = 99$

$1-\alpha = \frac{99}{100}$

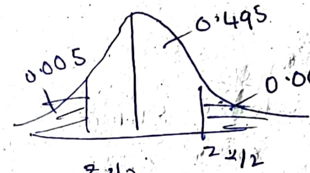
$1-\alpha = 0.99$

$\alpha = 1-0.99$

$\alpha = 0.01$

$\frac{\alpha}{2} = \frac{0.01}{2}$

$= 0.005$



$\int_0^{Z_{\alpha/2}} \phi(z)\,dz = 0.495$

where $Z_{\frac{\alpha}{2}} = 2.58$

Ex

1.	16
2	18
3	20
4	
5	
6	

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

Ex1 In a study of an automobile insurance a random sample of 80 body repair costs had a mean of RS. 472.36 and the S.D of Rs. 62.35. 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 RS 10.

Sol $n=80$

mean $\bar{x}$ = Rs 472.36

S.D, σ = Rs 62.35

E_{max} = Rs. 10.

$E_{max} = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = \frac{E_{max} \cdot \sqrt{n}}{\sigma} = \frac{10\sqrt{80}}{62.35} = \frac{89.4427}{62.35} = 1.4345$

$Z_{\alpha/2} = 1.43$

area $Z=1.43$ from table 0.4236.

$\frac{\alpha}{2} = 0.4236 \Rightarrow \alpha = 0.8472$

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

~~20) A prof~~ Bayesian Estimation:

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

$$\mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$$

$$\sigma_1 = \sqrt{\frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}}$$

$n \to$ sample size, $\bar{x}$ = sample mean, S is a S.D of sample

Bayesian interval for μ:

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

$$\mu_1 - Z_{\alpha/2} \cdot \sigma_1 < \mu < \mu_1 + Z_{\alpha/2} \cdot \sigma_1$$

Ex20 A proffesor feelings about the mean mark on the final examination in "probability" of a large group of statements is expressed subjectivly by normal distribution with $\mu_0 = 67.2$ & $\sigma_0 = 1.5$

a) If the mean mark lies in the interval (65.0, 70.0) determine the prior probability the professor should assign to the mean mark.

Sol μ_0 ~~=~~

$x_1 = 65$, $x_2 = 70$

$Z_1 = \frac{x_1 - \mu}{\sigma} = \frac{65 - 67.2}{1.5} = -1.466$

$Z_2 = \frac{x_2 - \mu}{\sigma} = \frac{70 - 67.2}{1.5} = 1.866$

$P(Z_1 \le Z \le Z_2) = P(-1.466 \le Z \le 1.866)$

$= \int_0^{Z_1 = 1.466} \phi(z)dz + \int_0^{Z_2 = 1.866} \phi(z)dz$

$= 0.4279 + 0.4686$

$= 0.8965$

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

Sol $n = 40$

$\bar{x} = 74.9$

$S.D (\sigma = s) = 7.4$

$$\mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$$

$$= \frac{40(74.9)(1.5)^2 + (67.2)(7.4)^2}{40(1.5)^2 + (7.4)^2} = 71.987$$

$$\sigma_1 = \sqrt{\frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}} = \sqrt{\frac{(1.5)^2(7.4)^2}{40(1.5)^2 + (7.4)^2}}$$

$= 0.922568$

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)

Sol $\mu_1 = 71.987$, $\sigma_1 = 0.9225682$

$$z_1 = \frac{x_1 - \mu_0}{\sigma} = \frac{65 - 71.987}{0.9225682} = -7.5134$$

$$z_2 = \frac{70 - 71.987}{0.9225682} = -2.15377$$

d) ~~construct a 95% Bayesian interval for μ.~~

Tests of hypothesis for large sample :-

A statistical hypothesis is a statement about the parameter of one or more population

- Testing is a process for deciding whether to accept or reject.
- There are 2 types of hypothesis
  1. Null hypothesis
  2. Alternative "

— **Null hypothesis (h_0)** :-

A hypothesis of No difference is called null hypothesis it is denoted by h_0.

- It is in the form of H_0 : $\mu \neq \mu_0$

— **Alternative hypothesis (h_1)** :- A hypothesis which is complementary to Null hypothesis is called an alternative hypothesis, it is denoted by h_1

* It is in the form $\mu \neq H_1$, $\mu < \mu_0$ or $\mu > \mu_0$

— **Critical region** :- A region corresponding to a statistic in the sample space "S" which leads to reject h_0 is called critical region, the region which leads to accepts h_0 is called acceptance region, the value of test statistic, which separate rejection region and acceptance region is called critical value or statistical value.

Type I error:-

reject h_0 when it is true i.e. rejecting a correct hypothesis it is denoted by α.

Type ② error:-

Accept h_0 when it is wrong

- It is denoted by β, the only way to reduce both types of errors to increase the sample size if possible.

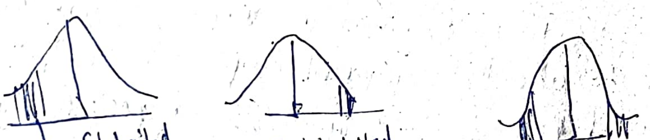
	H_0 True	H_0 False
reject	Type-I error	correct inference
accept	correct inference	Type-II error

* Typ-I error - reject H_0 when it true in 100%.
* Type-II error : accept H_0 when it false

I tail and II tail tes

for I tail test critical region is represent at only one size (left or right side).

* Alternative hypothesis for I tail is $>$ (or) $<$ tail



left tailed Right-tailed Two tailed

- working rule for testing of hypothesis

step① NULL hypothesis

define null hypothesis h_0 in clear terms.

step② Alternative hypothesis

define h_1 so, that one can decide we should use one tail or 2 tailed test

step③ level of significant:-

level of significantès i.e sutabli α is selected in advance, generally it is 1%, 2%, 3%, 5%, 10%.

step④ - Write test statistic which is formula to find calculated value.

step5 :- Conclusion

compare the calculated value with table value at given level of significant.

→ if $|Z_{cal}| \leq |Z_{tab}|$ accept H_0

→ **large sample test** :- (Z test)

type ① Ztest for single mean (mean and S.D. with 1 sample)

type ② Ztest for difference of means (mean and S.D with 2 sample)

③ Ztest for single proportion (No mean, No S.D. with one sample)

④ Z test for different of proportion (No mean, No S.D. with 2 sample)

Step 1 null hypothesis . H_0 $\mu = \mu_0$

Step 2 Alternative hypothesis $\mu \neq \mu_0$

Step 3 level of significance. (α = 1% or 2% or 5% or 10%)

~~step 4: α = 1% or 2%~~

step 4: Test statistic : $|z_{cal}| = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

step 5 $|z_{cal}| \leq |z_{tab}|$

Accept H_0.

→ type II (different of means)

level 1, 2, 3 same as type I

type 4 $z_{cal} = \dfrac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$

→ Type III

Z test for single propotion

1. Null hypothesis (H_0) $P = 0.5$
2. Alternative hypothesis (H_1) : $P \neq 0.5$ (two tailed)

   $P < 0.5$ or $P > 0.5$ (one tailed)
3. Level of significance (α) = 5% or 10% or 2% or 1%.
4. Test statistic = $z_{cal} = \dfrac{p-P}{\sqrt{PQ}}$ where $p = \frac{x}{n}$

5) Conclusion If $|z_{cal}| \leq |z_{tab}|$

we accept Null hypothesis H_0

Type IV z test for difference of propotions

1. Null hypothesis (H_0) = $P_1 = P_2$
2. Alternative hypothesis (H_1) H_1 = $P_1 \neq P_2$ (two tailed)

   $P_1 < P_2$ or $P_1 > P_2$ (one tailed)
3. level of significance (α) = 5% or 10% or 2% or 1%.
4. Test statistic : $z_{cal} = \dfrac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where $P_1 = \frac{x_1}{n_1}$, $P_2 = \frac{x_2}{n_2}$

   $P = \dfrac{x_1+x_2}{n_1+n_2}$
5. Conclusion If $|z_{cal}| \leq |z_{tab}|$

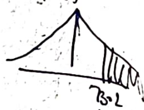
   we accept Null hypothesis H_0

Ex 1 :- Page 349

According to the norms establish for mechanical (or) an electrical ~~application~~ / appititude test, persons, who are 18 years old have an avg height of 73.2 with a standard deviation of 8.6. If 40 randomly selected persons of that age avg 76.7, test the hypothesis $\mu = 73.2$ against the alternative hypothesis $\mu > 73.2$ at the 0.01 level of significant. also find confidence interval, standard ERROR.

Sol: i) Null hypothesis (H_0) = $\mu = 73.2$

ii) Alternative hypothesis (H_1): $\mu > 73.2$ (one tailed)



iii) level of significance (α) = 0.01

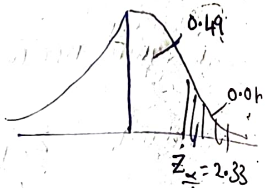
$(1-\alpha)100\% = 99\%$

$1-\alpha = \frac{99}{100}$

$1-\alpha = 0.99$

$\alpha = 1-0.99$

$\alpha = 0.01$



$\therefore |Z_{tab}| = 2.33$

iv) Test statistic: $Z_{cal} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$\bar{x}$ → sample mean, μ — popl mean, σ — popu s.d, n — sample size

$= \dfrac{76.7 - 73.2}{\frac{8.6}{\sqrt{40}}}$ $= Z_{cal} = 2.573$

v) Conclusion: $|Z_{cal}| \leq |Z_{tab}|$

$2.573 \nleq 2.33$

$\therefore$ reject Null hypothesis H_0

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

$= \left(76.7 - 2.33\left(\frac{8.6}{\sqrt{40}}\right), 76.7 + 2.33\left(\frac{8.6}{\sqrt{40}}\right)\right)$

$S.E = \frac{\sigma}{\sqrt{n}} = \frac{8.6}{\sqrt{40}} = 1.35$

→ A Sample of 64 students ~~with~~ have a mean weight of 70kgs. can this be regarded as a sample from a population with mean weight 56kgs and S.D 25kgs, find confident interval and Sandard error.

sol: Null hypothesis (H_0) = $\mu = 56$

2. Alternative hypothesis (H_1) = $\mu \neq 56$ (two tailed)

3. level of singnificance = 5% (Assumed)

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

$1-\alpha = 0.95$

$\alpha = 1-0.95$

$\alpha = 0.05$

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

$|Z_{tab}| = 1.96$

$0.500 - 0.025 = 0.475$

4) Test statistic = $Z_{cal} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$= \dfrac{70 - 56}{\frac{25}{\sqrt{64}}}$

$Z_{cal} = 4.48$

5) Conclusion: $|Z_{cal}| \leq |Z_{tab}|$

$4.48 \nleq 1.96$

$\therefore$ reject Null hypothesis H_0

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

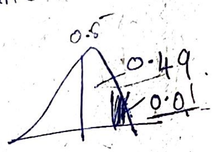
A simple sample of height of 6400 Englishmen has a mean of 67.85 inches and a S.D of 2.56 inches while a simple sample of heights of 1600 Australians has a mean of 68.55 inches and SD 2.52 inches do the data indicates the Australians are on the avg taller than the englishmen ($\alpha = 0.01$)

sol

1 Null hypothesis (H_0) = $\mu_1 = \mu_2$

2. A.H (H_1) = $\mu_1 < \mu_2$ (one tailed)

3 level of hypothesis (α) = 0.01



$(1-\alpha)100\% = 99\%$

$1-\alpha = 0.99$

$\alpha = 0.01$

$|z_{tab}| = 2.33$

4. Test statistic $z_{cal} = \dfrac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

$= \dfrac{67.85 - 68.55}{\sqrt{\frac{(2.56)^2}{6400} + \frac{(2.52)^2}{1600}}}$

$z_{cal} = |-9.9| \Rightarrow z_{cal} = |-9.9| = 9.9$

5 conclusion $|z_{cal}| \leq |z_{tab}|$

$\nleq 2.33$

reject Null hypothesis H_0.

Eg

In a big city 325 men out of 600 men were found to be smokers does this information support the conclusion, that the majority of this city are smokers.

Sol $n = 600 > 30$ (large sample)

$p = \frac{x}{n}$

$p = \frac{325}{600} = 0.5416$

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

1. Null hypothesis $H_0: P = 0.5$

2. Alternative hypothesis $H_1: P > 0.5$ (one tailed)

3. Level of significance $\alpha = 5\%$ (Assume)

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

$(1-\alpha)100 = 95$

$1-\alpha = 0.95$

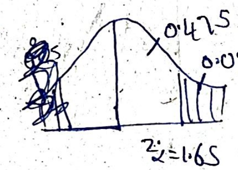
$\alpha = 0.05$

~~$\alpha = 0.025$~~

~~$z_{\alpha/2} = 1.96$~~

$\alpha = 0.45$

$z_{\alpha} = 1.65$



4. test statistic: $z_{cal} = \dfrac{p-P}{\sqrt{\frac{PQ}{n}}}$

$= \dfrac{0.5416 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{600}}}$

$|z_{cal}| = |2.037| = 2.037$

5. conclusion:

$|z_{cal}| \leq |z_{tab}|$

$2.037 \nleq 1.65$

$\therefore$ reject H_0

experience had soon that 20% of manufacture product is of top quality in one day production of 400 articles only 50 are top quality, test the hypothesis at 0.05 l. also find confidence interval.

Sol $n = 400 > 30$ (large sample)

$p = \frac{50}{400} = p = 20\% \Rightarrow 0.2 \quad Q = 0.8$

1. Null hypothesis H_0 $P = 0.2$
2. Alternative hypothesis H_1 $P \neq 0.5$ (two tailed)
3. level of significance . $\alpha = 0.05$

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

$Z_{\frac{\alpha}{2}} = 1.96$

4. test statistic . $Z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$

$= \frac{0.125 - 0.2}{\sqrt{\frac{(0.2)(0.8)}{400}}}$

$|Z_{cal}| = |-3.75|$

$= 3.75$

5. conclusion:

$|Z_{cal}| \leq |Z_{tab}|$

$3.75 \leq 1.96$

∴ reject H_0

confidence interval:

$\left(p - Z_{\frac{\alpha}{2}}\sqrt{\frac{PQ}{n}}, \quad p + Z_{\frac{\alpha}{2}}\sqrt{\frac{PQ}{n}}\right)$

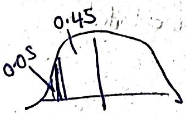
$= 0.125 - 1.96\sqrt{\frac{0.2 \cdot (0.8)}{400}}, \quad 0.125 + 1.96\sqrt{\frac{0.2(0.8)}{400}}$

→ A social worker believes that fewer than 25% of the couples in a certain area have ever used and form of birth control. a random sample of 120 couples was contacted, 20 of them said that their have used, test the belief of social work at 0.05 level

Sol $p = \frac{20}{120} = 0.16 \quad P = 25\% \Rightarrow 0.25$

$Q = 1 - P = 0.75$

1. Null hypothesis H_0 $P = 0.25$
2. Alternative hypothesis H_1 $P < 0.25$ (one tailed)
3. Level of significance $\alpha = 0.05$



$Z_\alpha = 1.65$

$|Z_{tab}| = 1.65$

4. test statistic = $Z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$

$= \frac{0.16 - 0.25}{\sqrt{\frac{(0.25)(0.75)}{120}}}$

$= -2.27$

$|Z_{cal}| = |-2.27| = 2.27$

5. Conclusion:

$|Z_{cal}| \leq |Z_{tab}|$

$2.27 \nleq 1.65$

reject H_0

→ Random Sample 400 men and 600 women were asked wheath they would like to have a flyover near their residence: 200 men and 325 women were in favour of proposal test the hypothesis that proposition of mean and women are proposal are same at 5% level

Sol $n_1 = 400 > 30$, $n_2 = 600 > 30$ (large samples)

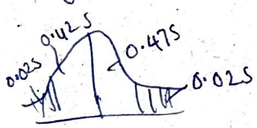
$p_1 = \frac{200}{400} \frac{x_1}{n_1}$, $P = \frac{325}{600} \frac{-x_2}{-n_2}$

$= 0.5$ $\qquad$ $= 0.541$

1. null hypothesis H_0 $P_1 = P_2$

2. Alternative hypothesis H_1 $P_1 \neq P_2$ (two tailed).

3. Level of significance $\alpha = 5\%$



$|z_{tab}| = 1.96$

4. test statistic

$$z_{cal} = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

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

$$= \frac{200 + 325}{400 + 600}$$

$P = 0.525$

$q = 0.475$

$$|z_{cal}| = \frac{0.5 - 0.541}{\sqrt{(0.525)(0.475)\left(\frac{1}{400} + \frac{1}{600}\right)}}$$

$z_{cal} = 1.3$

5) Conclusion: $|z_{cal}| \leq |z_{tab}|$

$1.3 \leq 1.96$

Accept H_0,,

→ on the basis of their total scores, 200 Candidates of a civil service examination are divided into two groups, the upper 30% and the remaining 70%.

4 - Test of Hypothesis of Small Sample

Degree of Freedom :- The no. of independent values (or) quantities which can be asign to the statical distribution

— If the size of the sample less than 30, then that sample is called samll sample or exact sample.

They are 3 kinds of test for small samples

1. t-test
   - t-test for single mean
   - t-test for difference of means
   - paired t-test (x)

2. F-test - for equallity of variance

3. χ^2-test (khi square test)
   - χ^2 test for goodness of fit (x)
   - χ^2 test for independent of attribute

→ **t-test for single mean** :- (one sample with mean and S.D)

testing small samples using ~~key di~~ t-distribution is called t-test.

– single mean test procedure

1. Null hypothesis $H_0 = \mu = \mu_0$

2. Alternative hypothesis $H_1 : \mu \neq \mu$ (two tailed)

   $\mu < \mu_0$ or $\mu > \mu_0$ (one tailed)

3. α : 1% ; 2% (or) 5% (or) 10%.

4. Test statistic : $t_{cal} = \dfrac{\bar{x} - \bar{\mu}}{\frac{s}{\sqrt{n-1}}}$ (or) $\dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$ — $\bar{x}$ – sample mean, μ – popleⁿ mean, s – sample s.d, n – sample size → raw data when mean is no given

where $s^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2$

5. Conclusion : t_{tab} value for d.F $(\nu) = n-1$

   at α (two tailed) or $\frac{\alpha}{2}$ (one tailed)

   if $|t_{cal}| \leq |t_{tab}|$

   Accept H_0.

Confidence Interval :- $\left(\bar{x} \pm z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n-1}}\right)$

→ **difference of mean test procedure**

1. $H_0 : \mu_1 = \mu_2$

2. H_1 $\mu_1 \neq \mu_2$ (two tailed)

   $\mu_1 < \mu_2$ or $\mu_1 > \mu_2$ (one tailed)

3. α : 1% or 2% or 5% or 10%.

4. Test statistic

$$t_{cal} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $s^2 = \frac{1}{n+n-2}\left[\Sigma (x_i - \bar{x})^2 + \Sigma (y_i - \bar{y})^2\right]$ (or)

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

5. Conclusion :

t_{tab} value for d.f $(\nu) = n_1 + n_2 - 2$ at α (two tailed) or $\alpha/2$ (one tailed). If $|t_{cal}| \leq |t_{tab}|$

Accep H_0.

... population showed a mean of 53 and A sum of squares of diviation from the mean ~~is eq~~ = 150, can this Sample be regarded as taken from the popution ~~org~~ having 56 as mean? uptain 95% confidence limits of the mean of the population.

Sol $n = 16 < 30$ (small sample)

$\bar{x} = 53$

$\mu = 56$

$\Sigma(x_i - \bar{x})^2 = 150$

$s^2 = \frac{1}{n-1}\Sigma(x_i - \bar{x})^2$

$= \frac{1}{16-1}(150)$

$= \frac{150}{15}$

$s^2 = 10$

$\boxed{S = 3.16}$

1. $H_0: \mu = 56$
2. $H_1: \mu \neq 56$ (two talled)
3. $\alpha = 5\%$

   $\alpha = 0.05$

   $\frac{\alpha}{2} = 0.025$
4. test Statistic: $t_{cal} = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n-1}}} = \frac{53-56}{\frac{3.16}{\sqrt{16-1}}} = -3.79$

   $|t_{cal}| = |-3.79| = 3.79$
5. Conclusion: $|t_{tab}| = 2.13$ For d.f.$(v) = n-1 = 16-1 = 15$ at $\frac{\alpha}{2} = 0.025$ level of significant

if $|t_{cal}| \leq |t_{tab}|$

$3.79 \nleq 2.13$

Reject H_0

confidence interval

$\left(\bar{x} - Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n-1}}, \bar{x} + Z_{\frac{\alpha}{2}}\frac{S}{\sqrt{n-1}}\right)$

~~(64.97, 64.98)~~

- producer of Gutka claims that the nicotine contain in his Gutka on the average is 1.83 Mg. can this claim t accepted if a random sample of 8 Gutka of this kind of the nicotine contain 2.0, 1.7, 2.1, 1.9, 2.2, 2.1, 2.0, 1.6 mg? used a 0.05 level of significent.

Sol $n = 8 < 30$ (small sample)

x	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
2.0	2-1.95=0.05	
1.7	1.7-1.95=-0.25	
2.1	2.1-1.95=0.15	
1.9	1.9-1.95=-0.05	
2.2	2.2-1.95=0.25	
2.1	2.1-1.95=0.15	
2.0	2.0-1.96=0.05	
1.6	1.6-1.96=-0.35	
Σx 15.6		

$\bar{x} = \frac{\Sigma x}{n} = \frac{15.6}{8} = 1.95$

→ The means of 2 random samples of sizes 9 and 7 are 196.42 and 198.82 respectively. The sum of the squares of the deviation from the mean are 26.94 and 18.73 respectively. Can the sample be considered to have been drawn from the same normal population.

Sol: $n_1 = 9 < 30$, $n_2 = 7 < 30$ (both are small samples).

$\Sigma(x_i - \bar{x})^2 = 26.94$, $\Sigma(y_i - \bar{y})^2 = 18.73$

$$s^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2\right]$$

$$= \frac{1}{9+7-2}\left[26.94 + 18.73\right]$$

$s^2 = 3.26$

1. $H_0 : \mu_1 = \mu_2$
2. $H_1 : \mu_1 \neq \mu_2$ (two tailed)
3. $\alpha = 5\%$

$\alpha = 0.05$

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

4. test statistics : $t_{cal} = \dfrac{\bar{x}-\bar{y}}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

$$= \frac{196.42 - 198.62}{\sqrt{3.26\left(\frac{1}{9}+\frac{1}{7}\right)}}$$

$t_{cal} = -2.6367$

$|t_{cal}| = |-2.6367|$

$= 2.6367$

5. conclusion:

$|t_{tab}| = 2.145$ value for d.f $(\nu) = n_1+n_2-2$

$= 9+7-2$

$= 14$ at $\frac{\alpha}{2} = 0.025$ level of significance

$|t_{cal}| \leq |t_{tab}|$

$2.6367 \nleq 2.145$

Reject H_0

→ Samples of two types of electric light bulbs were tested for length of life and following data were obtained

Type I	Type II
Sample number, $n_1 = 8$	$n_2 = 7$
Sample mean, $\bar{x} = 1234$ hrs	$\bar{y} = 1036$ hrs
Sample S.D, $s_1 = 36$ hrs	$s_2 = 40$ hrs

Is the difference in the means sufficient to warrant that type I is superior to type II regarding length of life.

Sol: 1. $H_0 : \mu_1 = \mu_2$

2. $H_1 : \mu_1 \neq \mu_2$ (two tailed)

3. $\alpha = 5\%$

$\alpha = 0.05$

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

4. test statistics $t_{cal} = \dfrac{\bar{x}-\bar{y}}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}$$

$$= \frac{8(36)^2 + 7(40)^2}{8+7-2}$$

$s^2 = 1659.08$

$$t_{cal} = \frac{1234 - 1036}{\sqrt{1659.08\left(\frac{1}{8} + \frac{1}{4}\right)}}$$

$t_{cal} = 9.39$

5. conclusion

$|t_{tab}| = 2.160$ values for $d.f(v) = n_1 + n_2 - 2$
$= 8 + 7 - 2$
$= 13$ at $\frac{\alpha}{2} = 0.025$ lead of significant

$|t_{cal}| \leq |t_{tab}|$

$9.39 \nleq 2.160$

Reject H_0

→ Two independent samples of 8 and 7 items respectively had the following values

Sample I	11	11	13	11	15	9	12	14
Sample II	9	11	10	13	9	8	10	-

Is the difference b/w the means of samples significant?

Sol: $\bar{x} = \frac{\Sigma x}{n} = \frac{96}{8} = 12$

$\bar{y} = \frac{\Sigma y}{n} = \frac{70}{7} = 10$

x	y	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
11	9	-1	1	-1	1
11	11	-1	1	1	1
13	10	1	1	0	0
11	13	-1	1	3	9
15	9	3	9	-1	1
9	8	-3	9	-2	4
12	10	0	0	0	0
14		2	4		
$\Sigma x_i = 96$	$\Sigma y_i = 70$		26		16

1. $H_0: \mu_1 = \mu_2$
2. $H_1: \mu_1 \neq \mu_2$ (two tailed)
3. α : 5%

$\alpha = 0.05$

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

4. test statistics $t_{cal} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2\right]$$

$$= \frac{1}{8 + 7 - 2}\left[26 + 16\right]$$

$s^2 = 3.23$

$$t_{cal} = \frac{12 - 10}{\sqrt{3.23\left(\frac{1}{8} + \frac{1}{7}\right)}} = 2.15$$

5. conclusion

$|t_{tab}| = 2.16$ value for $d.f(v) = n_1 + n_2 - 2 = 8 + 7 - 2$
$= 13$ at $\frac{\alpha}{2} = 0.025$ lead of significant

$|t_{cal}| \leq |t_{tab}|$

$2.15 \leq 2.160$

Accept H_0

F-test for equality of variance :-

1. Null hypothesis $H_0: \sigma_1^2 = \sigma_2^2$
2. Alternative hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$
3. level of significant $\alpha = 5\%$ or 1%
4. Test statistics $F_{cal} = \frac{S_1^2}{S_2^2}$ $(S_1^2 > S_2^2)$ where $S_1^2 = \frac{1}{n_1-1}\Sigma(x_i-\bar{x})^2$

(OR) $S_2^2 = \frac{1}{n_2-1}\Sigma(y_i-\bar{y})^2$

$= \frac{\text{Greater variance}}{\text{smaller variance}}$.

5) F_{tab} value for d.f $(v) = (\underset{v_1}{n_1-1}, \underset{v_2}{n_2-1})$ at α - level of significance.

IF $|F_{cal}| \leq |F_{tab}|$

Accept H_0.

→ In 1 sample of 8 observations from a normal population the sum of the squares of deviations of the sample values from the sample mean is 84.4 and in another sample of 10 observations it was 102.6. Test at 5% level wheather the populations have the same variance.

Sol · $n_1 = 8$ $\Sigma(x_i-\bar{x})^2 = 84.4$

$n_2 = 10$ $\Sigma(y_i-\bar{x})^2 = 102.6$

$S_1 = \frac{1}{n_1}\Sigma(\bar{x}_i-\bar{x})^2$

$S_1^2 = \frac{1}{8-1}(84.4)$

$= 12.057$

$S_2^2 = \frac{1}{n_2-1}\Sigma(y_i-\bar{y})^2$

$= \frac{1}{10-1}(102.6)$

$S_2^2 = 11.4$

1. $H_0: \sigma_1^2 = \sigma_2^2$
2. $H_1: \sigma_1^2 \neq \sigma_2^2$
3. $\alpha = 5\%$

$\alpha = 0.05$

4. $F_{cal} = \frac{S_1^2}{S_2^2}$ $(S_1^2 > S_2^2)$

OR

$\frac{\text{Greater variance}}{\text{Smaller variance}} = \frac{12.057}{11.4} = 1.05)$

5. $[F_{tab}] = 3.29$ value for d.f $(v) = (n_1-1, n_2-1) = (8-1, 10-1))$ α. level of significance $= \underset{v_1\ v_2}{(7,9)}$

IF $|F_{cal}| \leq |F_{tab}|$

$1.057 \leq 3.29$

Accept H_0 ✓

→ Two random samples reveal the follow results:

Sample	Size	Sample mean	Sum of Square of Deviations from Mean
1	10	15	90
2	12	14	108

Test wheather the samples came from the same normal population.

Sol here we have to take 2 test

1. F-test for equality of variance
2. T-test for equality of means

F-test for equality of variance

1. $H_0 : \sigma_1^2 = \sigma_2^2$

2. $H_1 \quad \sigma_1^2 \neq \sigma_2^2$

3. Level $\alpha = 5\%$ @ 1%.

4. $F_{cal} = \frac{S_1^2}{S_2^2} \quad (S_1^2 > S_2^2)$

$= S_1^2 = \frac{1}{n_1 - 1} \Sigma (x_i - \bar{x})^2$

$S_2^2 = \frac{1}{n_2 - 1} \Sigma (x_i - \bar{x})^2$

$= S_1^2 = \frac{1}{10-1} \cdot 90 = \frac{90}{9} = 10$

$S_2^2 = \frac{1}{12-1}(108) = \frac{108}{11} = 9.82$

$= \frac{10}{9.82} = 1.018$

5. value of d.f(v) $= ((n_1 - 1), (n_2 - 1))$

$= (10-1), (12-1)$

$= (9, 11)$

$\alpha \cdot F(v_1, v_2)$

$0.05 \cdot (9, 11)$

$= 2.90$

$1.018 \leq 2.90$

Accept H_0

t-test for equality of mean

1. $H_0 : \mu_1 = \mu_2$

2. $H_1 : \mu_1 \neq \mu_2$ (two tailed.)

3. level of significan $= 5\% = 0.05$

$= (1-\alpha)100 = 5$

$\alpha = 0.05$

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

$2\frac{\alpha}{2} = 1.90$

4. $t_{cal} = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$S^2 = \frac{1}{n_1 + n_2 - 2} \left[\Sigma (x - \bar{x})^2 + \Sigma (y - \bar{y})^2 \right]$

$S^2 = \frac{1}{10+12-2} [90 + 108] = \frac{1}{20} [198] = 9.9$

$S = 3.15$

$t_{cal} = \frac{15 - 14}{3.15 \sqrt{\frac{1}{10} + \frac{1}{12}}} = 0.7437$

$|t_{cal}| = 0.7437$

5. conclusion: $d.f(v) = n_1 + n_2 - 2$

$= 10 + 12 - 2 = 20$ at 0.025

$= 2.086$

$0.7437 \leq 2.086$

accept H_0

→ ~~Find~~

→ Time taken by the workers in performing a job by method I and method II is given below.

Method I	20	16	26	27	23	22	-
Method 2	27	33	42	35	32	34	38

Do the data show that the variance of time distribution from population from which these samples are drawn do not differ significantly?

Sol

1. H_0 $\sigma_1^2 = \sigma_2^2$

2. H_1 $\sigma_1^2 \neq \sigma_2^2$

x	y	$(x-\bar{x})$	$(x-\bar{x})^2$	$(y-\bar{y})$	$(y-\bar{y})^2$
20	27	-2.3	5.29	-7.4	54.76
16	33	-6.3	39.69	-1.4	1.96
26	42	3.7	13.69	7.6	57.76
27	35	4.7	22.09	0.6	0.36
23	32	0.7	0.49	-2.4	5.76
22	34	-0.3	0.09	-0.4	0.16
	38			3.6	12.96

Σ

$\bar{x} = \frac{\Sigma xi}{n_1} = \frac{134}{6} = 22.3$

$\bar{y} = \frac{\Sigma y}{n_2} = \frac{241}{7} = 34.4$

$\Sigma(x-\bar{x})^2 = 81.34$

$\Sigma(y-\bar{y})^2 = 133.72$

$S_1^2 = \frac{1}{n_1-1}\Sigma(x-\bar{x})^2 = \frac{1}{6-1}(81.34)^2 = 16.36$

$S_2^2 = \frac{1}{n_2-1}\Sigma(y-\bar{y})^2 = \frac{1}{7-1}(133.72)^2 = 22.29$

$S_2^2 > S_1^2$

$F_{cal} = \frac{\text{Greater variance}}{\text{smaller varian}} = \frac{22.29}{16.36} = 1.3699 \simeq 1.37$

F_{tab}: d.F $(v) = (n_1-1, n_2-1)$

$= (5, 6)$ at 5%

$= 4.39$

$1.37 \leq 4.39$

Accept H_0

→ Pumpkins were grown under two experimental conditions. Two random samples of 11 and 9 pumkins, show that sample standard deviation of their weight as 0.8 and 0.5 respectively. Assuming that the weight distributions are normal, test hypothesis that the weight distribution are normal, test hypothesis that the true variance are equal.

Sol H_0: $\sigma_1^2 = \sigma_2^2$

$H_1 = \sigma_1^2 \neq \sigma_2^2$

$n_1 = 11$, $n_2 = 9$, $S_1 = 0.8$, $S_2 = 0.5$

$S_1^2 = \frac{n_1 s_1^2}{n_1-1} = \frac{11 \times (0.8)^2}{10} = 0.704$

$S_2^2 = \frac{n_2 s_2^2}{n_2-1} = \frac{9 \times (0.5)^2}{8} = 0.281$

iv $F_{cal} = \frac{S_1^2}{S_2^2} = \frac{0.704}{0.281}$

$= 2.5$

5. $F_{tab} = f(v)(n_1-1, n_2-1)$

$(0, 8)$ at 5%

$F_{tab} = 3.35$

$F_{cal} \leq F_{tab}$

$2.5 \leq 3.35$

Accept H_0

— χ^2-test for independent of attributes :-

literally an attribute mean's a quality (or) characteristi

Ex drinking, smoking, honesty, handsome, beauty..

- an attribute may be mark. by its presence (or) absences in a no. of given population.

- Test procedure

① Null hypothesis (attributes are independent.)

② Alternative hypothesis (attribute's are dependent.)

③ level of significant 1%

③ Test statistic $\chi^2_{cal} = \sum \frac{(O_i - E_i)^2}{E_i}$ observed freq, Expected freq

= Expected freq (E_i) = row total × column total / Grand total

5) Conclusion

χ^2_{tab} value for $df(v) = (\text{no. of rows} - 1)(\text{no. of columns} - 1)$

If $|\chi^2_{cal}| \leq |\chi^2_{tab}|$, then accept H_0

→ Ex Given the following contigency table for hair colour and eye colour. Find the value of χ^2. Is there good association b/w the two?

		Hair colour			
		Fair	Brown	Black	Total
Eye colour	Blue	15	5	20	40
	Grey	20	10	20	50
	Brown	25	15	20	60
	Total	60	30	60	150

Sol

$\frac{40\times60}{150} = 16$	$\frac{40\times30}{150} = 8$	$\frac{40\times60}{150} = 16$
$\frac{50\times60}{150} = 20$	$\frac{50\times30}{150} = 10$	$\frac{50\times60}{150} = 20$
$\frac{60\times60}{150} = 20$	$\frac{60\times30}{150} = 12$	$\frac{60\times60}{150} = 24$

O_i		$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
15	16	-1	1	$\frac{1}{16} = 0.0625$
5	8	-3	9	$\frac{9}{8} = 1.125$
20	16	4	16	$\frac{16}{16} = 1$
20	20	0	0	0
10	10	0	0	0
20	20	0	0	0
25	24	1	1	$\frac{1}{24} = 0.042$
15	12	3	9	$\frac{9}{24} = 0.75$
20	24	-4	16	$\frac{16}{24} = 0.6$

1) Null hypothesis (H_0) = Hair colour and Eye colour are independent

2) Alternative hyp (H_1) not independent

3 $\alpha = \alpha = 5\%$

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

4. test statistic $X^2_{cal} = \sum \frac{(O_i - E_i)^2}{E_i}$

$= 3.6458$

5. Conclusion

$X^2_{tal} = 11.14$ For $d.f(v) = (\text{no. of rows} - 1)(\text{no. of column} - 1)$

$|X^2_{cal}| \leq |X^2_{tab}|$ $= (3-1)(3-1) = 4$

$3.6458 \leq 11.14$

Accept H_0

→ Two researchers adopted different sampling techniques while investigating some group of students to find the no. of students falling into different intelligence level.

Researchers	Below Average	Average	Above Average	Genius	Total
X	86	60	44	10	200
Y	40	33	25	2	100
Total	126	93	69	12	300

would you say that the sampling techniques adopted by the two researchers are significantly different. [Given 5% value of X^2 for 2 degrees and 3 degrees of freedoms are 5.991 and 7.82 respectively].

Sol Table of expected frequencies.

$\frac{126 \times 200}{300} = 84$	$\frac{93 \times 200}{300} = 62$	$\frac{69 \times 200}{300} \simeq 46$	$\frac{12 \times 200}{300} = 8$	200
$\frac{126 \times 100}{300} = 42$	$\frac{93 \times 100}{300} = 31$	$\frac{69 \times 100}{300} = 23$	$12 \times 100 = 4$	100
126	93	69	12	300

calculation of X^2

O.F (O_i)	E.F (E_i)	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
86	84	4	0.0476
60	62	4	0.0645
44	46	4	0.0869
10	8	4	0.5
40	42	4	0.0952
33	31	4	0.129
25	23	4	0.1739
2	4	4	1

i) H_0: reachers are independent
~~H_1~~ reachers are dependent
iii) level of significant = 5% = 0.05
Test statistic $\chi^2_{cal} = \sum \frac{(O_i - E_i)^2}{E_i}$

$= 2.0971$

v) conclusion χ^2_{tab} = for d.f.(v) = (no. of rows −1) + (no. of colum −1).

$= (4-1) + (2-1)$

$= 3$

$|\chi^2_{cal}| \leq |\chi^2_{table}|$

$2.0971 \leq 7.82$

Accept H_0

~~→ on the~~ basic o

→ Find maximum difference that we can except with probability 0.95 b/w the means of sample of sizes 10 and 12 from a normal population if there S.D is are found to be 2 and 3 respectively

$H_0 = \mu_1 = \mu_2$

$H_1 - \mu_1 \neq \mu_2$ (two tailed

level of significant 0.05

$(1-\alpha)100 = 0.05$

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

$t_{cal} = \frac{\bar{x} - \bar{y}}{\sqrt{\cancel{\frac{s_1^2}{n}} + \frac{s^2}{n}}}$ $\frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$t_{cal} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$\bar{x} - \bar{y} = t_{cal} \sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

$\bar{x} - \bar{y} = |t_{cal}| \sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

$\leq |t_{tab}| \sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{10(4) + 12(9)}{10 + 12 - 2} = 7.4$

$\bar{x} - \bar{y} \leq |2.086| \sqrt{7.4\left(\frac{1}{10} + \frac{1}{12}\right)} = 1.35$

$\bar{x} - \bar{y} = 1.35$

Corelation and Regression :-

A distribution involving 2 varibles is known as bivarial distribution. if these 2 varibles vary such that changing one varible affects the changk in other varible, the varible are said to be co-related.

Ex1) the exist a relation b/w the height and weight of a person.

Ex2 the prise of a community and its demand

Ex3 rainfall and production of rice -etc.

→ The degree of relationship b/w the varible under consideration is measure, through the co-relation analysis. The measure of co-relation is called as the corelation co-efficient (or) co-relation index.

→ Types of co-relation

1. +ve & -ve co-relation
2. Simple and multiple co-relation
3. partial and total co-relation
4. linear and non-linear (curved linear) co-relation

→ +ve & -ve co-relation :- If the two varibles are deviate in the same direction, i.e. if the increase (decrease) in 1 varible results in a corresponding increase (decrease) in the other, then the corelation is said to be direct (or) corelate relation

Ex1) The relationship b/w height and weight of a person

Ex2) rainfall and production of rice -etc are +ve corelation

→ -ve corelation :- if 2 varibles are constantly deviate in opposite direction i.e if the increase (decrease) in 1 varible results in a corresponding decrease (increase) in the other, the corelation is said to inverse (or) -ve

Ex: the corelation b/w prise of a community & its demand.

2. Simple & multiple corelation :-

when we study only 2 varibles, the relationship is discribe as simple corelation.

Ex: demand and prise

when we study more than 2 varible simultanous, the relationship is discribe as multiple corelation.

Ex: relationship of price, demand and supply of commodity

3. partial and total corelation :-

The study of 2 varibles exculding some other varibles partial co-relation.

Ex:- we study price and demand exculding supply.

In total corelation all facts are taken into consideration

(4) linear & non-linear co-relation :-

If the ratio of change b/w 2 varible is uniform then there will be linear co-relation b/w them.

Ex.

x	1	2	3	4	5
y	5	10	15	20	25

In a curve linear or non-linear co-relation, the amount of change in one varible does not be a constant ratio of the amount of change in the other varible.

Methods of Co-relation :-

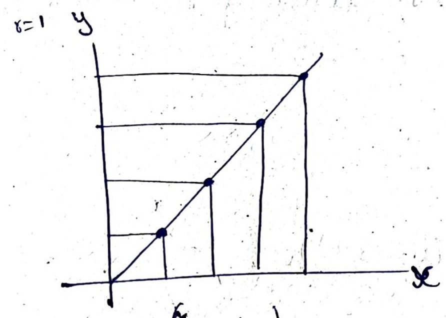
① scatter diagram method
② graphic method
~~③ curve~~
③ carl pearson's coefficient corelation
④ Rank method (spearman's rank corelation)
⑤ Concurrent Deviation method
⑥ method of least squares

→ Scatter diagram method :-

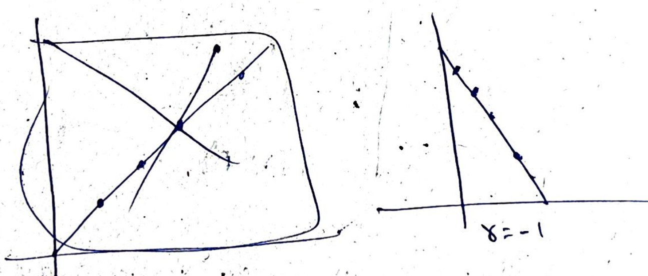
In the bivariate distribution, the values of varible's x and y are plotted in the xy plane, the diagram of dots obtained is called a scatter diagram.

- If all the points lie on a st-line ~~catti~~ falling from the lower left end corner to upper right end corner, correlation is said to be perfectly +ve (corelation co-efficient $r=1$)

(i)

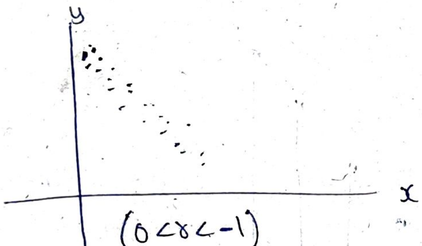


ii) If the the points lie on a st-line raising from upper left hand corner to lower right hand corner, correlation is said to perfectly -ve



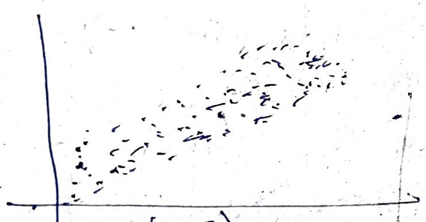
iii) If the plotted points fall in a narrowing bound, and there show a raising tendency from a lower left hand corner to upper right end corner, there would be high degree of +ve co-relation

④ If the plotted points fall in a narrow bound, and there show a declaring tendency from the upper left end corner to lower right hand corner, there would be high degree of -ve correlation.



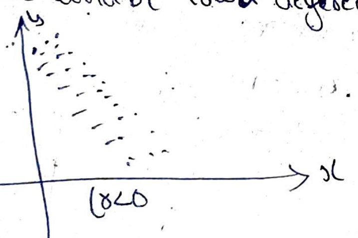
$(0 < r < -1)$

→ If the points are widely scatted over the diagram and the points are raising from the lower left end corner to upper right hand cornert, there would be lower degree of +ve correlation



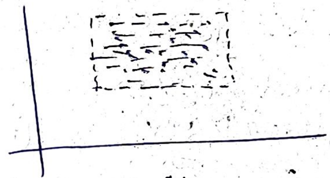
$(r > 0)$

→ If the points are widely scattered and the points are running from the upper left hand side to lower right hand side there would be lower degeree of -ve co-relation



$(r < 0$

→ If the plotted points lie on a ~~st~~ line, parallel to x-axis and parallel to y-axis, It shows ~~options~~ absents of corelation b/w 2 varibles



Advantages of ~~standard~~ setter diagrams:-

Setter diagram is simple and attrative method, to find out the nature of correlation.

→ It is easy to understand a glance.

- A rough Idea is ~~got~~ droped wheather a +ve (or) -ve corelation.

Carl pearson's coefficient correlation:-

as a measure of degree of linear relationship b/w 2 varibles, carl pearson's developed a formula called co-relation coefficient, The co-relation coefficient b/w 2 random varibles x and y usually denoted by r or r_{xy}. Is a measure of linear relationship b/w them. and its defined as

$$r = \frac{\Sigma dx\,dy}{\sqrt{\Sigma (dx)^2 \, \Sigma (dy)^2}}$$

(or)

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \, \Sigma Y^2}}$$

Spearman's rank corelation :- (Rank ~~metho~~ co-relation)

Suppose, $x_1, x_2, x_3 \ldots x_n$ and $y_1, y_2, y_3, \ldots y_n$ are the ranks of 2 varibles x and y in the order of merit the co-relation b/w this ends pairs of ranks

① rank correlation coefficient

$$\gamma = 1 - \frac{6 \Sigma d^2}{n^3 - n}$$

d^2 → difference is ranks [rank are not ocpresent]

$n^3 - n$ → no. of paired observations

ii) rank correlation cofficien

$$1 - \frac{6[\Sigma d^2 + C.F]}{n^3 - n}$$

C.F → correction factor

[ranks are represent]

$$C.F = \frac{m^3 - m}{12}$$

→ Calculate coefficient of following data

x	12	9	8	10	11	13	7
y	14	8	6	9	11	12	3

Ans Karl pearson's coefficient of correlation

$$\gamma = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$

x	y	$X = x - \bar{x}$	$Y = y - \bar{y}$	XY	X^2	Y^2
12	14	2	5	10	4	25
9	8	-1	-1	1	1	1
8	6	-2	-3	6	4	9
10	9	0	0	0	0	0
11	11	1	2	2	1	4
13	12	3	3	9	9	9
7	7	-3	-6	18	9	36
				$\Sigma XY = 46$	$\Sigma X^2 = 28$	$\Sigma Y^2 = 84$

$\Sigma x = 70$

$\bar{x} = \frac{\Sigma x}{n} = \frac{70}{7} = 10$ $\quad \bar{y} = \frac{\Sigma y}{n} = \frac{63}{7} = 9$

$$\gamma = \frac{46}{\sqrt{28 \cdot 84}}$$

$\gamma = 0.948$

② Find Karl pearson's coefficient of correlation from the following data

wages	100	101	102	102	100	99	97	98	96	95
cost of learning	98	99	99	97	95	92	95	94	90	91

→ calculate carl pearson's and coefficient correlation for the following data.

x	28	41	40	38	35	33	40	32	35	33
y	23	34	33	34	30	26	28	31	36	38

4) obtain Rank correlation for the following data.

x	68	64	75	50	64	80	75	40	55	64
y	62	58	68	45	81	60	68	48	50	70

Sol By spearman's rank corelation

$$r = 1 - \frac{6\left[\Sigma d^2 + C.F\right]}{n^3 - n}$$

where $CF = \frac{m^3 - m}{12}$

correction factor

x	y	rank of $x = x_1$	rank of $y = y_1$	$d = x_1 - y_1$	d^2
68	62	4	5	-1	1
64	58	6	7	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1
64	81	6	1	5	25
80	60	1	6	-5	25
75	68	2.5	3.5	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
					$\Sigma d^2 = 72$

For x series 75 is repeated 2 times (rank $= \frac{2+3}{2} = 2.5$)

for x series 64 is repeated 3 times (rank $= \frac{5+6+7}{3} = 6$)

for y series 68 is repeated 2 times (rank $= \frac{3+4}{2} = 3.5$)

Now $r = 1 - \frac{6\left[\Sigma d^2 + \frac{m^3-m}{12} + \frac{m^3-m}{12} + \frac{m^3-m}{12}\right]}{n^3 - n}$

$= 1 - \frac{6\left[72 + \frac{2^3-2}{12} + \frac{3^3-3}{12} + \frac{2^3-2}{12}\right]}{10^3 - 10}$

$\boxed{r = 0.545}$ +ve correlational

→ calculate spearman's coefficient of rank correlation for the following data

x	58	98	95	81	75	61	59	55
y	49	25	32	37	30	40	39	45

→ Find the rank correlation for the following indices of supply and price of an article.

Employ index	124	100	105	112	102	93	99	115	123	104	99	113	121	[illegible]	101
price index	80	104	102	91	100	111	109	100	89	104	111	102	98	111	[illegible]

① A sample of 12 Fathers and their elderest sons gave the following data about their height in inches calculate the coefficient of rank correlation

Father	65	63	67	64	68	62	70	66	68	67	69	71
son	68	66	68	65	69	66	68	65	71	67	68	70

Ssol

x	y	$x = x_1$	$y = y_1$	$d = x_1 - y_1$	d^2
53	47	8	1	7	49
98	25	1	8	-7	49
95	32	2	6	-4	16
81	37	3	5	-2	4
75	30	4	7	-3	9
61	40	5	3	2	4
59	39	6	4	2	4
55	45	7	2	5	25
					160

$$r = 1 - \frac{6[160 + 0]}{8^3 - 8}$$

$$r = 1 - \frac{960}{504}$$

16 × 6 = 960

64 × 8 = 512

Regression :-

The study of corelation measures the direction & strength of relationship b/w 2 varibles, in co-relation, we can estimate the value of 1 varible when the value of other varible is event.

- Since price and supply are corelated, we can findout the expected amount of supply for a given price a we can findout the required price, the a given amount of supply.
- In regression we can estimate the value of varible with the value of other varible, which is known other varible the statistical method which helps us to estimate the unknown value of 1 varible from the known value of related varible is called regression.
- The line describe in the avg relationship b/w 2 varibles is known as line of regression.

Uses

i) It is used to estimate the relation b/w 2 economic varible like income & expendature
ii) It is highly valuble tool in economics and business
iii) widely used for production perpuse
iv) we can calculate co-efficient of co-relation and co-efficient of determination with the help of regression coefficient
v) It is useful statistical estimation of demand curves, supply curves, production function, costfunction & consuction func

Comparison b/w co-relation & regression

co-relation-cofficient	regression
i) It is a measure of degree of variblity b/w 2 varibles	i) A regression establises a function relation b/w dependent or Independent varia
ii) both x & y are random varibles	ii) x is random varible y is fixed varible.
iii) It is a relative measure	iii) It is a absolute figure

methods of studing regression

2 methods.

1. Graphic method
2. Algebraic method

→ Graphic method: In this method the point representing the pairs of values of the varibles are ploted in a graph the Independent varibles are taken on x-axis & dependent varible on Y-axis.

- These point from a scattered diagram. A regression line is draw b/w this points by free hand

→ Algebraic method:

regression line: It is a st. line fited to the data by the method least squares it indicates the best posible mean values of 1 varible corresponding to the mean value of other

there are always 2 regression lines.

1. constructed for the relationship b/w 2 varibles x & y

Thus one, regression line shows the regression x upon y and other shows the regression on Y upon X

→ regression equations:

1. regression equation X on Y is $X-\bar{x} = b_{xy}(Y-\bar{y})$

where $b_{xy} = \frac{\Sigma XY}{\Sigma Y^2}$

2. regression equation on Y on X is $Y-\bar{y} = b_{xy}(X-\bar{x})$

where $b_{xy} = \frac{\Sigma XY}{\Sigma X^2}$

- from the following data obtain 2 regression eqns

x	6	2	10	4	8
y	9	11	5	8	7

also ① find y when x = 10,

② find x when y = 5

- regression eqn X on Y and y on x.

X on Y = $\bar{x}-\bar{x} = bxy\,(Y-\bar{y})$

Y on X = $Y-\bar{y} = b_{xy}(X-\bar{x})$

x	y	$X = x - \bar{x}$	$Y = y - \bar{y}$	XY	X^2	Y^2
6	9	0	1	0	0	1
2	11	-4	3	-12	16	9
10	5	4	-3	-12	16	9
4	8	-2	0	0	4	0
8	7	2	-1	-2	4	1
				$\Sigma XY = 26$	$\Sigma X^2 = 40$	$\Sigma Y^2 = 20$

$\bar{x} = \frac{\Sigma x_i}{n} = \frac{30}{5} = 6$

$\bar{y} = \frac{\Sigma y_i}{n} = \frac{40}{5} = 8$

Regression eqn X on Y

$x - \bar{x} = b_{xy}(y - \bar{y})$ — (1)

$b_{xy} = \frac{\Sigma XY}{\Sigma Y^2} = \frac{26}{20} = \frac{-13}{10}$

From (1)

$x - 6 = \frac{-13}{10}(y - 8)$

Find x when y = 5

$x - 6 = \frac{-13}{10}(5 - 8)$

$\begin{array}{r} 39 \\ 60 \\ \hline 9.9 \end{array}$

$x = \frac{39}{10} + 6$

$\boxed{x = \cancel{4.9}}$ $\boxed{x = 10}$

Regression eqn Y on X

$y - \bar{y} = b_{xy}(x - \bar{x})$ where $b_{xy} = \frac{\Sigma XY}{\Sigma X^2} = \frac{-\cancel{26}\,13}{\cancel{40}\,20} = \frac{-13}{20}$

$y - 8 = \frac{-13}{20}(x - 6)$

i) Find y when x = 10

$y - 8 = \frac{-13}{20}(10 - 6)$

$y = \frac{-\cancel{52}}{20} + 8$

$y = \frac{\cancel{108}\,54}{\cancel{20}\,10}$

$\boxed{y = 5.4}$ $\boxed{y = 5}$

→ prince indices of cotton and wool are given below for the 12 months of a year. Obtain the equality of lines of regressions.

price index of cotton	78	77	85	89	87	82	81	77	76	83	97	93
price index of wool	84	82	82	85	89	90	88	92	83	89	98	99

→ The heights of mothers' and daughters are given below. Find the equation of lines of regression.

Mother	62	63	62	64	65	66	68	70
Daughter	64	65	61	69	69	68	71	65

Sol

x	y	$X = x-\bar{x}$	$Y = y-\bar{y}$	XY	X^2	Y^2
78	84	-6	-5	30	36	25
77	82	-7	-7	49	49	49
85	82	1	-7	-7	1	49
88	85	4	-4	-16	16	16
87	89	3	0	0	9	0
82	90	-2	1	-2	4	1
81	88	-3	-1	3	9	1
77	93	-7	4	-28	49	16
76	83	-8	-6	48	64	36
83	89	1	0	1	1	0
97	98	13	9	117	169	81
93	99	9	10	90	81	100
				287	488	365

$\bar{x} = \frac{\Sigma x_i}{n}$ $\frac{1004}{12} = 84$

$\bar{y} = 88$

- Regression eqn X on Y

$x - \bar{x} = b_{xy}(y - \bar{y})$

$= X - 84 = 0.78(Y - 88)$ $b_{xy} = \frac{\Sigma xy}{\Sigma y^2} = \frac{287}{365}$

- Regression eqn Y on X

$Y - \bar{y} = b_{xy}(X - \bar{x})$

$Y - 88 = \frac{287}{488}(X - 84)$

$Y - 88 = (0.58X) - 48.72$

$Y = (0.58X) + 39.82$

→ Unit-V

Stochastic process & Markov chain

(or)

Random

- In probability theory a Markov model is a stochastic model used to model randomly changeing system over time

Ex: no. of students enter in a class before time t

let X_0 be a no. of students, entered in a class at $t = t_0$.

- X_1 be a no. of students, entered in a class at $t = t_0 + 1$, ----

- X_1 be a no. of students, entried in a class at time $t = t_n$

let $\{X_n ; n = 0, 1, 2, ---\}$ be a stochastic process that takes on a finite or countable number of possible values.

- notation: $X_n = 1$, it means process is said to be in state I at time n

In Markov model two terms play an important model.

① states

② transition probability

Markov Chain :-

It is the process x_0, x_1, x_2 ---- here system states are observable and fully autonomous and simplest of all markov model

Markov property :-

The basic property of markov chain is that X_{n+1}, depends on X_n but it doesnot depend upon $X_{n-1}, X_{n-2}, X_{n-3}, \ldots, X_1, X_0$

mathematically :-

The markov property is stated

$$P_{ij} = P\left\{ X_{n+1} = j \,/\, X_n = i, X_{n-1} = i_1, X_{n-2} = i_2 \ldots X_0 = i_0 \right\}$$

given

for all $n = 1, 2, \ldots$

and for all states

- Transition probabilities :-

Classification of Stochastic process :-

i) Continous Stochastic process :-

→ If x & t both are continous, the stochastic process is known as CSP.

ii) Discrete Stochastic Process :-

If x is continous & t is discrete the stochastic process is called DSP.

- If x is discrete & t is continous then the stochastic process is called DSP.

- If both x & t are discrete then the stochastic process is DSP.

iii) Deterministic stochastic process :-

- A random process is called discrete deterministic Stochastic process if all the future values can be predicted from past observation.

iv) Non-deterministic stochastic process :

- A stochastic process is known as NDSP.

Probability Transition.

- probabilities from after one step time period denoted by P_{ij} is defined by

$$P_{ij} = P\{x_{n+1} = j \mid x_n = i\}$$

Ex $P_{12} = \{x_3 = 2 \mid x_2 = 1\}$ ii) $P_{21} = \{x_1 = 1 \mid x_0 = 2\}$

$$P_{12} = \{x_3 = 2 \mid x_2 = 1\}$$

n-step transition probability:

- probabilities from state i to state j after n-step time period denoted by $P_{ij}(n)$ (or) P_{ij}

$P_{ij}^{n+1} = P\{x_{n+1}$

An urn initially contains five black balls & five white balls. The following experiment is repeated indefinitely. A ball is drawn from the urn; if the ball is white it is put back in the ~~urn~~ urn, otherwise it left out.

let x_n be no. of black balls remains.

a) Find apropirate transition probabilities

b) Find the one-step tp matrix P for x_n.

c) Find the two-step TP matrix P^2 by matrix multiplica

d) what happens to x_n as n approaches infinity? use your answer to guess the limit of P^n as $n \to \infty$.

Sol $P_{00} = P\{x_n = 0 | x_{n-1} = 0\} = \frac{5c_1}{5c_1} = 1$

$P_{01} = 0, P_{02} = 0, P_{03} = 0, P_{04} = 0, P_{05} = 0$

$P_{10} = P\{x_n = 0 | x_{n-1} = 1\} = \frac{1c_1 \times 5c_0}{6c_1} = \frac{1}{6}$

$P_{11} = P\{x_n = 1 | x_{n-1} = 1\} = \frac{1c_0 \times 5c_1}{6c_1} = \frac{5}{6}$

$P_{12} = 0, P_{13} = 0, P_{14} = 0, P_{15} = 0$

$P_{20} = 0$

$P_{21} = P\{x_n = 1 | x_{n-1} = 2\} = \frac{2c_1 \times 5c_0}{7c_1} = \frac{2}{7}$

$P_{22} = P\{x_n = 2 | x_{n-1} = 2\} = \frac{2c_0 \times 5c_1}{7c_1} = \frac{5}{7}$

$P_{23} = 0, P_{24} = 0, P_{25} = 0$

$P_{30} = 0, P_{31} = 0$

$P_{33} = P\{x_n = 3 | x_{n-1} = 3\} = \frac{3c_0 \times 5c_1}{8c_1} = \frac{5}{8}$

$P_{32} = P(x_n = 2 | x_{n-1} = 3) = \frac{3c_1 \times 5c_1}{8c_1} = \frac{3}{8}$

$P_{34} = 0, P_{35} = 0$

$P_{40} = 0, P_{41} = 0, P_{42} = 0$

$P_{43} = P\{x_n = 3 | x_{n-1} = 4\} = \frac{4c_1 \times 5c_0}{9c_1} = \frac{4}{9}$

$P_{44} = P\{x_n = 4 | x_{n-1} = 4\} = \frac{4c_0 \times 5c_1}{9c_1} = \frac{5}{9}$

$P_{45} = 0$

$P_{50} = 0, P_{51} = 0, P_{52} = 0, P_{53} = 0$

$P_{54} = P\{x_n = 4 | x_{n-1} = 5\} = \frac{5c_1 \times 5c_0}{10c_1} = \frac{5}{10}$

$P_{55} = P\{x_n = 5 | x_{n-1} = 5\} = \frac{5c_0 \times 5c_1}{10c_1} = \frac{5}{10}$

ii) TPM:

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{5}{6} & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{7} & \frac{5}{7} & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{8} & \frac{5}{8} & 0 & 0 \\ 0 & 0 & 0 & \frac{4}{9} & \frac{5}{9} & 0 \\ 0 & 0 & 0 & 0 & \frac{5}{10} & \frac{5}{10} \end{bmatrix} \end{array}$$

iii) $$P^2 = P \cdot P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{5}{6} & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{7} & \frac{5}{7} & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{8} & \frac{5}{8} & 0 & 0 \\ 0 & 0 & 0 & \frac{4}{9} & \frac{5}{9} & 0 \\ 0 & 0 & 0 & 0 & \frac{5}{10} & \frac{5}{10} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{5}{6} & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{7} & \frac{5}{7} & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{8} & \frac{5}{8} & 0 & 0 \\ 0 & 0 & 0 & \frac{4}{9} & \frac{5}{9} & 0 \\ 0 & 0 & 0 & 0 & \frac{5}{10} & \frac{5}{10} \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{11}{36} & \frac{25}{36} & 0 & 0 & 0 & 0 \\ \frac{1}{21} & \frac{65}{144} & \frac{25}{49} & 0 & 0 & 0 \\ 0 & \frac{3}{28} & \frac{225}{448} & \frac{25}{64} & 0 & 0 \\ 0 & 0 & 0 & \frac{95}{192} & \frac{25}{81} & 0 \\ 0 & 0 & 0 & \frac{2}{9} & \frac{19}{36} & \frac{1}{4} \end{bmatrix}$$

& $n \to \infty$

$$P^n = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The transaction probability matrix of a Marcov chain $\{x_n\}$ $n=1,2,3 \ldots$ having 3 states 1,2,3 is

$$P = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$$

and the intial distribution is $p^{(0)} = (0.7, 0.2, 0.1)$ find

i) $P\{x_2 = 3\}$

ii) $P\{x_3 = 2, x_2 = 3, x_1 = 3, x_0 = 2\}$

~~$P^2 = P.P$~~

Sol $P(0.7, 0.2, 0.1)$

$P(x_0=0) \quad P(x_0=2) \quad P(x_0=3)$

$$P^2 = P.P = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{bmatrix} = \begin{bmatrix} 0.43 & 0.31 & 0.26 \\ 0.24 & 0.42 & 0.34 \\ 0.36 & 0.35 & 0.29 \end{bmatrix}$$

i) $P\{\underset{\text{step}}{x_{\underset{\text{state}}{2}}} = 3\} = P\{x_2 = 3 | x_0 = 1\} \times P(x_0 = 1) + P\{x_2 = 3 | x_0 = 2\} \times P(x_0 = 2)$

$+ P\{x_2 = 3 | x_0 = 3\} \times P(x_0 = 3)$

$$\boxed{\text{step size } P_{ij}^{(n+1-1)} = P\{x_{n+1} = j | x_1 = i\} \quad \text{States}}$$

$= P_{13}^{2-0} \times P(x_0 = 1) + P_{23}^{(2-0)} \times P(x_0 = 2) + P_{33}^{(2-0)} \times P(x_0 = 3)$

$P_{13}^{(2)} \times 0.7 + P_{23}^{(2)} \times 0.2 + P_{33}^{(2)} \times 0.1$

$= 0.26 \times 0.7 + 0.34 \times 0.2 + 0.29 \times 0.1 = 0.271$

ii) $P(X_3=2, X_2=3, X_1=3, X_0=2\}$

$P(X_1=3, X_0=2) = P\{X_1=3 \mid X_0=2\} \times P(X_0=2)$

$= P_{23}^{(1-0)} \times P(X_0=2)$

$= P_{23} \times P(X_0=2)$

$= 0.2 \times 0.2 = 0.04$

$P(X_2=3, X_1=3, X_0=2\} = P\{X_2=3 \mid X_1=3, X_0=2\} = P\{X_2=3 \mid X_1=3\}$
$\times P\{X_1=3, X_0=2\}$

$= P_{33}^{(2-1)} \times 0.04$

$= P_{33} \times 0.04$

$= 0.3 \times 0.004.$

$P(X_3=2, X_2=3, X_1=3, X_0=2\} = P(X_3=2 \mid X_2=3, X_1=3, X_0=2)$

$= P(X_3=2 \mid X_2=3) \times P(X_2=3, X_1=3, X_0=2)$

$= P_{32}^{3-2} \times 0.012$

$= P_{32} \times 0.012$

$= 0.4 \times 0.012$

$= 0.0048$

→ Consider markov chain

$$P = \begin{bmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 3/4 & 1/4 \end{bmatrix}$$

Find $P_{01}^{(2)}$ and $P(X_2=1, X_0=0)$ where $P(X_0=0) = \frac{1}{3}$

sol $P^2 = P.P$

$$= \begin{bmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 3/4 & 1/4 \end{bmatrix} \begin{bmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 3/4 & 1/4 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 10/16 & 5/16 & 1/16 \\ 5/16 & 8/16 & 3/16 \\ 3/16 & 9/16 & 4/16 \end{bmatrix}$$

i) $P_{01}^{(2)} = \frac{5}{16}$

ii) $P(X_2=1, X_0=0) = P\{X_2=1 \mid X_0=0\} \times P(X_0=0)$

$= P_{0}^{2-0} \times P(X_0=0)$

$= P_{01}^{2} \times P(X_0=0) = \frac{5}{16} \times \frac{1}{3} = \frac{5}{48}$

A fair die ~~tossed~~ toss if $f(n)$ denoted the maximum of number occuring in the n tooses find the transaction probability matrix P of the Markov chain $\{X_n\}$ find also P^2 and $P(X_2=6)$

Notes

A Transction matix P has zero entries and P^2 also conta 'O' entries, we may wonder how far shall we compute P^k to be certain ~~to be~~ that the matrix is not regular

- The answer is that if zero's occur in the identical places work P^k and P^{k+1} for any k they will appear In those places for higher powers of P. so P is not regular (or) when 1 occurs principle diagonal.

- Which of the strocatoic matrixs are regular

(i)
$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

ii) $$C = \begin{bmatrix} 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$$

iii) $$A = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.5 & 0.5 \\ 0.6 & 0.4 & 0 \end{bmatrix}$$

iv) $$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

v) $$\begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

vi) $$\begin{bmatrix} \frac{1}{3} & 0 \\ \frac{1}{3} & 1 \end{bmatrix}$$

vii) $$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

viii) $$\begin{bmatrix} \frac{1}{2} & \frac{1}{4} & 1 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$C^2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$$C^3 = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.5 & 0.5 \end{bmatrix}$$

$$C^4 = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.25 & 0.25 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}$$

$$C^5 = \begin{bmatrix} 0.25 & 0.5 & 0.25 \\ 0.125 & 0.25 & 0.375 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$$

A Finite irreducible markov chain is called Normal.

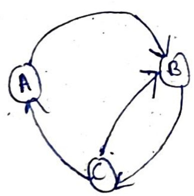
Note: written state: A state i of a markov chain is called written state if $(P_{ij})^n > 0 \ \forall F_n \geq 1$

- if there exist atleast one path can not be written to the orginal state then that state is called transnent start
- ~~s~~ if a state is not-transnent, then it is a recorrent state

- 3 Boys A, B, C are throwing a ball to each other if A always throws the ball to B, B always throws the Ball to C, but C is just as likely to throw the Ball B as to A. show that the process is Markovbian. Find the TM (TPM) and clasify the states do all the states are ergadic?

Sol

$$\begin{array}{c|ccc} & A & B & C \\ A & 0 & 1 & 0 \\ B & 0 & 0 & 1 \\ C & 0.5 & 0.5 & 0 \end{array}$$



- states of X_n depend only on states of X_{n-1} but not on states X_{n-2}, X_{n-3} ---

$\therefore$ X_n is markovbian

state A is accessible from states B and C

state C is " " " A and C

state C " " " A and B

- all states are accessible for all other states

$\therefore$ all states are communicate

- here the no. of states in markov chain are finite.

since all state communicate each other Therefor A, B and C irreducible.

- all states are irreducible and it is finite markov chain

$\therefore$ all the states are Nonnull ~~persistent~~ persient

- period of A $(d_i) = \gcd\{3, 5, 7, ---\}$

$= 1$

$\therefore$ State A is aperiodic

period of B $(d_i) = \gcd\{2, 3\}$

$= 1$

$\therefore$ states B is aperiodic

period of C $(d_i) = \gcd\{2, 3\} = 1$

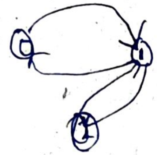
$\therefore$ states C is aperiodic

$\therefore$ all states are aperiodic since A, B and C states are non-null persisted Hence all states are eregodic

- 3 states are not absorbing states (i.e. No state has one diagonal elements hence it is not absorbing.

→ Find the nature of states of markov chain with TPM

$$P = \begin{array}{c c} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \begin{bmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix} \end{array}$$



state 0 is accessible from state 1 and 2

state 1 " " 0 and 2

state 2 " " 0 and 1

0, 1, and 2 all accessible from all other states

∴ 0, 1, and 2 all communicate.

hence the no. of states in markov chain is finite since all the states are communicate each other

∴ 0, 1, 2 are irreducible

- all states are irreducible and it is a finite markov chain

∴ all states are normal persistent.

- to find period:

period of '0' (d_i): $\gcd\{2, 4, 6, \ldots\} = 2$

∴ state 0 is periodic

Period of 1 $(d_i) = \gcd\{2, 2\}$

$= 2$

∴ state 1 is periodic

period of 2 $(d_i) = \gcd(2, 4, 6, \ldots\}$

$= 2$

state is periodic

- All states are periodic

0, 1, 2 states are non-null persistent but not aperiodic

Hence 0, 1 and 2 are not ergodic

→ Steady State Condition

- stable probability in many markov chains, the probability for a particular state will approach a limiting value as time goes to infinity. In otherwords in the four future the probability won't be changing from 1 transaction to another fraction, these limiting values are called stable probability.

if a system is such that each state as probability equal to its stable probability, the probability will persist for all the time then the system is said to be steady state condition

- Steady state vector (or) equillibrium vector (or) fixed vector

- if a markov chain with transition matrix P is regular then there is a unique vector V such that for any probability vector V for all large values of

$vp^n = v$ where v is called steady state vector.

Note :- a probability vector is a matrix of only one row, having non-negative entires with the sum of entires = 1

→ Find the equilibrium vector for the transition matrix

$$P = \begin{bmatrix} 0.5 & 0.28 & 0.07 \\ 0.15 & 0.67 & 0.18 \\ 0.12 & 0.36 & 0.52 \end{bmatrix}$$

– we observed that all entires in ~~markov~~ TPM are +ve so, this regular matrix then there exist steady state vector v such that

$vp^n = v$

$vp = v$ let $v = [v_1\ v_2\ v_3]$

$$[v_1\ v_2\ v_3] \begin{bmatrix} 0.65 & 0.28 & 0.07 \\ 0.15 & 0.67 & 0.18 \\ 0.12 & 0.36 & 0.52 \end{bmatrix} = v_1\ v_2\ v_3$$

$= [0.65v_1 + 0.15v_2 + 0.12v_3 \quad 0.28v_1 + 0.67v_2 + 0.36v_3 \quad 0.07 + 0.18v_2 + 0.52v_3]$

$= [v_1\ v_2\ v_3]$