Date
Page

Scanned by CamScanner

# SAMPLING dist & ESTIMATION

## Sampling

It is the process of selecting a part of pop$^n$ for the study of stastic in order to estimate about parameter.

## Statistic :-

It is the sample characteristic of a distribution. For eg :- sample size ($n$), Sample mean ($\bar{x}$), Sample S.D ($s$), sample prop ($p$), Sample corr$^n$ coeff ($r$)

## Parameter :-

It is the population characteristic of a distribution. For eg :- Pop$^n$ size ($N$), Pop$^n$ mean ($\mu$) Pop$^n$ S.D ($\sigma$), Pop$^n$ prop. ($P$), Pop$^n$ correl$^n$ coeff ($\rho$) etc

## Standard Error (S.E)

$\sigma$ is the standard deviation of sampling distribution of sample statistic.

For eg :- $S.E(\bar{x}) = \frac{\sigma}{\sqrt{n}}$ (For infinite $N$ or SRSWR)

$S.E(\bar{x}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ (For finite $N$ or SRSWOR)

$S.E(p) = \sqrt{\frac{pq}{n}}$ (For infinite $N$ or SRSWR)

$= \sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}}$ (For finite $N$ or SRSWOR)

where $\sigma$ = pop$^n$ S.D $\quad q = 1-p$

$p = \frac{x}{n}$ = sample pop$^n$

Scanned by CamScanner

## Estimation :-

n is the process of predicting the value of population parameter by the study of sample statistic

## Types OF Estimation

→ Point Estimation

If we estimate the value of pop$^n$ parameter by a single value then it is called point estimation

→ Interval Estimation

If we estimate the value of pop$^n$ parameter by a range of two values then it is called interval estimation. Since, probability is associated with interval estimation. So it is also known as confidence interval or judicial limits.

Confidence interval for pop$^n$ mean

* $(1-\alpha)\%$ C.I for $\mu = \bar{x} \pm Z_\alpha \cdot S.E(\bar{x})$
  (for large sample $n > 30$)
* $(1-\alpha)\%$ C.I for $\mu = \bar{x} \pm t_{\alpha(n-1)} S.E(\bar{x})$
  (for small sample $n \leq 30$)

where $Z_\alpha$ = Tab. value of $z$ at $\alpha\%$ level of significance
$t_{\alpha(n-1)}$ = Tab value of $t$ at $\alpha\%$ level of sig & $(n-1)$ degrees of freedom

Confidence interval for pop$^n$ prop

$(1-\alpha)\%$ C.I for $P = p \pm Z_\alpha \cdot S.E(p)$

Scanned by CamScanner

~~Conf~~ Sample size determination

For estimation of population mean

$$n = \left(\frac{Z_\alpha \cdot \sigma}{E}\right)^2 \quad \text{if } N \text{ is not given}$$

Note :- If $N$ is given then

$$n_0 = \left(\frac{Z_\alpha \cdot \sigma}{E}\right)^2$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

where $n$ = sample size

$\sigma$ = popⁿ S.D

$E = |\bar{x} - \mu|$ = Max Permissable error

For the estimation of popⁿ prop

$$n = \left(\frac{Z_\alpha}{E}\right)^2 \times P \times Q$$

where $Z_\alpha$ = Tab. value of $Z$ at $\alpha\%$ level of size

$E = |p - P|$ = Max Permissable Error

$P$ = Popⁿ prop

$Q = 1 - P$

Note :- For almost sure or certain we take $Z_\alpha = 3$

If value of $P$ & $Q$ are not given, we should take $P = Q = \frac{1}{2}$

Scanned by CamScanner

Date
Page

# Central Limit Theorem

let $x_i$ ; $i = 1, 2 \ldots n$ be i.i.d (Identically independently distributed) r.v which follows Normal dist with mean $\mu$ & variance $\sigma^2$ ie 1.)

$x_i \sim N(\mu, \sigma^2)$

Then i) $\sum x_i \sim N(n\mu, n\sigma^2) \Rightarrow Z = \frac{\sum x_i - n\mu}{\sigma\sqrt{n}}$

ii) $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

Scanned by CamScanner

# Numerical Examples

A random sample of 50 apples is the average weight of 200 gm with S.D 40 gm. Find
i) 95% C.I for popⁿ mean weight
ii) 99% " " " " " "

Here we have

Sample size $(n) = 50$
" mean $(\bar{x}) = 200$
" sd $(s) = 40$

Here $S.E(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{40}{\sqrt{50}}$

$\therefore S.E(\bar{x}) = 5.65$

i) Here $C.I\ (1-\alpha) = 95\%$
$\Rightarrow \alpha = 5\% \Rightarrow Z_\alpha = 1.96$

Now
95% C.I for popⁿ mean $= \bar{x} \pm Z_\alpha . SE(\bar{x})$
$= 200 \pm 1.96 \times 5.65$
$= 200 \pm 11.074$
$= (188.926 - 211.074)$

ii) Here $C.I\ (1-\alpha)\% = 99\%$
$\Rightarrow \alpha = 1\% \Rightarrow Z_\alpha = 2.58$

Now, 99% C.I for popⁿ mean $(\mu) = \bar{x} \pm Z_\alpha . S.E(\bar{x})$
$= 200 \pm 2.58 \times 5.65$
$= 200 \pm 14.577$
$= (185.423 - 214.577)$

Scanned by CamScanner

2.) In a random sample of 64 students from a college 8 student are found to be smokes. Find the i) 95% Confidence interval for pop$^n$ prop of st
ii) 99% " " " " " "

Sol$^n$

Here, we have
Sample size $(n) = 64$
No. of smoker $(x) = 8$
Sample prop $(p) = \frac{x}{n} = \frac{8}{64} = 0.125$

$\Rightarrow q = 1 - p = 0.875$

$S.E(p) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.875 \times 0.125}{64}} = 0.41$

i) Here $C.I\ (1-\alpha) = 95\%$
$\Rightarrow \alpha = 5\% \Rightarrow Z_\alpha = 1.96$

Now, 95% C.I for pop$^n$ prop of smoker
$= p \pm Z_\alpha \cdot S.E.(p)$
$= 0.125 \pm 1.96 \times 0.41$
$= 0.125 \pm 0.08$
$= (0.117 - 0.133)$

3)

4.)

Scanned by CamScanner

A researcher wishes to estimate average expenditure by sampling technique, Determine the sample size such that maximum permissable error be Rs 20 with probability 95%.

Sol^n (Given Pop s.d (σ) = 500)

Here we have

Max Permissable Error (E) = 20

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

$\alpha = 5\%$

$\Rightarrow Z_\alpha = 1.96$ (From table)

Pop^n S.D (σ) = 500

Sample size (n) = ?

We know

$$n = \left(\frac{Z_\alpha \cdot \sigma}{E}\right)^2 = \left(\frac{1.96 \times 500}{20}\right)^2$$

$\therefore n = 2401$

It is known that 30% of all junior executive change their first job within 5 years. Compute the appropriate sample size to update above into such that error being 10% is with probability 99%.

Sol^n

Here we have:

prob prop = 0.3 = p

q = 1 − 0.3 = 0.7

Max Permissable Error (E) = 10% = 0.1

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

$\Rightarrow \alpha = 1\%$

$\Rightarrow Z = 2.58$

Scanned by CamScanner

Sample size (n) = ?

We know;

$$n = \left(\frac{Z_\alpha}{E}\right)^2 p \times q$$

$$= \left(\frac{2.58}{0.1}\right)^2 \times 0.3 \times 0.7$$

$$= 139.78$$

$$n = 140$$

5) Determine sample size if risk of making an error of 5 is 0.0456.
(Given pop° S.D is 50)

Sol^n

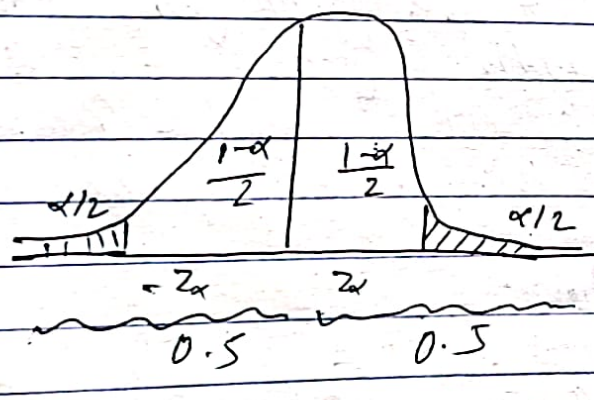
Here we have,

Max permissable Error (E) = 5

Risk ($\alpha$) = 0.0456

Pop^n S.D ($\sigma$) = 50,

Sample size (n) = ?




From graph,

$$P(0 < Z < Z_\alpha) = \frac{1-\alpha}{2} = \frac{1-0.0456}{2} = 0.4772$$

From SNT, $Z_\alpha = 2$

Now, $n = \left(\frac{Z_\alpha \cdot \sigma}{E}\right)^2 = \left(\frac{2 \times 50}{5}\right)^2 = 400$

Scanned by CamScanner

How does risk of making an error 4 with S.D 24 as sample varies from 36 to 144.

Sol$^n$

Here we have;

Max Permissable Error $(E) = 4$

Pop$^n$ S.D $(\sigma) = 24$

When, sample size $(n) = 36$

Risk $(\alpha) = ?$

We know,

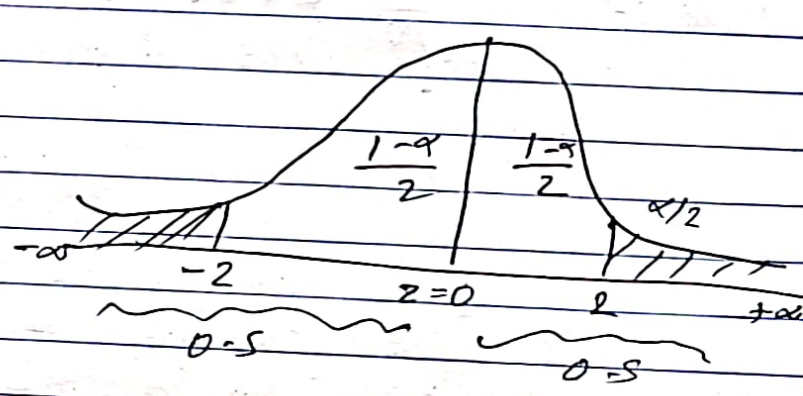
$$n = \left(\frac{Z_\alpha \, \sigma}{E}\right)^2$$

or, $36 = \left(Z_\alpha \times \frac{24}{4}\right)^2$

or $36 = 36 \, Z_\alpha^2$

or $1 = Z_\alpha^2$

$\therefore Z_\alpha = \pm 1$




From fig:- $P(0 < Z < Z_\alpha) = (1-\alpha)/2$

or, $P(0 < Z < 1) = \frac{1-\alpha}{2}$

or $0.3413 = \frac{1-\alpha}{2}$

$\alpha = 0.3174$ or $31.74\%$.

When sample size $(n) = 144$

Risk $(\alpha) = ?$

We know

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

or $144 = \left(Z_\alpha \times \frac{24}{4}\right)^2$

or, $Z_\alpha = \frac{144}{36} = \pm 2$

From graph

$P(0 < Z < Z_\alpha) = \frac{1-\alpha}{2}$

or $P(0 < Z < 2) = \frac{1-\alpha}{2}$

or $0.4772 = \frac{1-\alpha}{2}$

or $\alpha = 0.0456$ or $4.56\%$

Hence, risk reduced from $31.74\%$ to $4.56\%$

Scanned by CamScanner

## Characteristics of a Good Estimator

i) Unbiasedness
ii) Sufficiency
iii) Efficiency
iv) Consistency

Note :- If $\frac{n}{N} < 0.05$ we should neglect finite pop$^n$ correction fact$^r$ while calculating S.E.

Q no. 6 A random sample of 64 projector lamp indicated a sample average life of 3500 hours standard deviation of life is 200 hours. Then calculate the standard error of average projector lamp.

Sol$^n$

Here we have

Sample size $(n) = 64$

" mean $(\bar{x}) = 3500$

S.D $(s) = 200$

We know

$$S.E(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{s}{\sqrt{n}} = \frac{2400}{\sqrt{64}} = 25$$

Scanned by CamScanner

A simple random sample of size 20 is drawn without replacement from a finite population of 75 units, if the number of defective units in the population is 12. Assertain standard error of the sample proportion.
30

Here we have

Sample size $(n) = 20$

Popn size $(N) = 75$

No. of def. items $(x) = 12$

Sample prop $(p) = \frac{x}{n} = \frac{12}{20} = 0.6$

$\Rightarrow q = 1 - p = 0.4$

Now, $S.E.(p) = \sqrt{\frac{pq}{n}} \sqrt{\frac{N-n}{N-1}}$

$= \sqrt{\frac{0.6 \times 0.4}{20}} \sqrt{\frac{75-20}{75-1}}$

$= 0.0944$

Internet connections are often slowed by delays at nodes. Five hundred packets were sent through the same network between 5 h 6 pm and after four hours ie between 10 to 11 pm 300 packets were sent. The earlier sample showed mean delay time of 0.8 second with S.D of 0.1 second whereas the second sample showed mean delay time of 0.5 second with a S.D 0.08 sec. Calculate the standard error of difference between two sample means.
30

For $1^{st}$ sample

$n_1 = 500$

$\bar{x}_1 = 0.8$

$S_1 = 0.1$

For $2^{nd}$ sample

$n_2 = 300$

$\bar{x}_2 = 0.5$

$S_2 = 0.08$

$SE(\bar{x}_1 - \bar{x}_2) = ?$

We know,

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$= \sqrt{\frac{0.1^2}{500} + \frac{0.08^2}{300}}$$

$$= 0.0064$$

14)

11.) A random sample of size 64 has been drawn from a population with S.D 20. The mean of sample is 80. Calculate 95% confidence limit for the popⁿ mean. How does the width of the confidence interval change if sample size is 256 instead?

Sol/ Here we have,

sample size $(n) = 64$

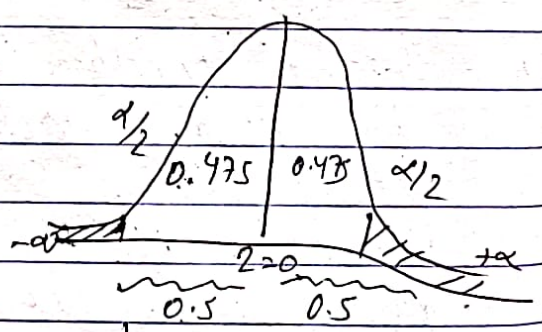
popⁿ S.D $(\sigma) = 20$

Samp. mean $(\bar{x}) = 80$

C.I $(1-\alpha)\% = 95\%$

$\Rightarrow \alpha = 5\%$

$\Rightarrow Z_\alpha = 1.96$ (From SNT)




Now, $S.E(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{64}} = 2.5$

95% C.I for popⁿ mean $(\mu)$

$= \bar{x} \pm Z_\alpha \cdot S.E(\bar{x})$

$= 80 \pm 1.96 \times 2.5$

$= 80 \pm 4.9$

$= (75.1 - 84.9)$

Scanned by CamScanner

Sample size $(n) = 256$

Here, $S.E(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{256}} = 1.25$

Now, 95% C.I for popⁿ mean $(\mu)$

$= \bar{x} \pm Z_\alpha \, S.E(\bar{x})$

$= 80 \pm 1.96 \times 1.25$

$= 80 \pm 2.45$

$= (77.55 - 82.45)$

In a random sample of 400 chips from a large consignment, 20 items were found 20 defective. Find 99% confidence limits for the percentage of defective chips in the consignment.

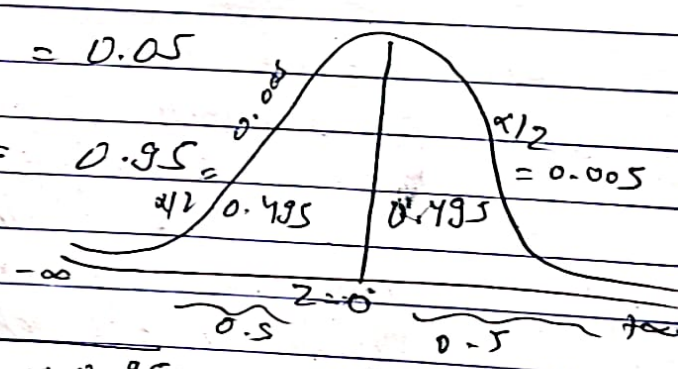
Sol:- Here we have

Sample size $(n) = 400$

No. of def. items $(x) = 20$

Sample prop $(p) = \frac{x}{n} = \frac{20}{400} = 0.05$

$\Rightarrow q = 1 - p = 0.95$




C.I $(1-\alpha)\% = 99\%$

$\Rightarrow \alpha = 1\%$

$Z_\alpha = 2.58$

Here, $S.E(p) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.05 \times 0.95}{400}} = 0.01$

Now, 99% C.I for popⁿ prop $(P)$

$= p \pm Z_\alpha \, S.E(p)$

$= 0.05 \pm 2.58 \times 0.01$

$= 0.05 \pm 0.0258$

$= 0.0242 - 0.0758$

$= 4.2\% - 7.58\%$

Scanned by CamScanner

20.) Assuming population standard deviation 3 how large should a sample to be estimate popⁿ mean with margin of error not exceeding 0.5 ?

Sol/ Here we have

Popⁿ S.D ($\sigma$) = 3

Sample size (n) = ?

Max Permissable Error (E) = 0.5

Here, $Z_\alpha$ is not given so we showed take $Z_\alpha = 1.96$ as risk level of sig ($\alpha$) is not given, $\alpha = 5\%$.

Now,

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

$$= \left(\frac{1.96 \times 3}{0.5}\right)^2 = 138.29$$

$$= 139$$

21.) The principle of a college wants to estimate the proportion of students who were interested to develop staking. What size of a sample should be select so as to have the diff of prop of interested students with true mean not too exceed by 10% with almost certainty? It is belived from previous records that the proportion of interested students was 0.30 ?

Sol/ Here we have

Max. Permissionable Error (E) = 10% = 0.1

Popⁿ prop (P) = 0.30

$\Rightarrow Q = 1 - P = 0.70$

For almost certain, we should take $Z_\alpha = 3$

$$\text{Samp size } (n) = \left(\frac{Z_\alpha}{E}\right)^2 \times P \times Q = \left(\frac{3}{0.1}\right)^2 \times 0.3 \times 0.7$$

23.)

Scanned by CamScanner

The average time taken by servers to execute an algorithm varies from time to time. From the past experience it is known that the time taken is normally distributed with standard deviation of 6.7 minutes. The IT manager wishes to estimate the average by drawing a random sample such that the probability is 0.95 that the mean of the sample will not.

Sol$^n$

Here we have

Pop$^n$ S.D ($\sigma$) = 6.7

Sample size (n) = ?

Prob $(1-\alpha) = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow Z_\alpha = 1.96$

Max Permissable Error (E) = 1

Now,

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

$$= \left(\frac{1.96 \times 6.7}{1}\right)^2$$

$$= 172.44$$

$$= 173$$

Scanned by CamScanner

Date
Page

i)

ii)

3)
i

Scanned by CamScanner

# Hypothesis Testing

Hypothesis :-
It is the assumption about statistic and parameter.

Types of Hypothesis

) Null Hypothesis
It is the assumptions of no difference between statistic and parameter. n is denoted by $H_0$.
Eg:- $H_0 : \mu = 50$ ie. there is no significance difference between population mean and it's specified value 50.

) Alternative Hypothesis
n is the assumption about statistic and parameter other than null hypothesis. It is denoted by $H_1$. Eg:- For $H_0 : \mu = 30$, alternative hypothesis may be
$H_1 : \mu \neq 30$ (Two tailed test)
or, $H_1 : \mu > 30$ (Right tailed test) } (One tailed test)
or, $H_1 : \mu < 30$ (Left tailed test) }

Procedure of testing of Hypothesis

) Setting of Hypothesis
~~Base~~ This is the first step in testing of Hypothesis, here based on sample size, no. of samples and parameter to test, we set up null and alternative hypothesis.
For eg:- $H_0 : \mu_1 = \mu_2$ ie there is no significant difference between two population means.
Vs $H_1 : \mu_1 \neq \mu_2$ i.e. there is significant difference between two pop$^n$ means

Scanned by CamScanner

or, $H_1 : \mu_1 > \mu_2$ (Right tailed) i.e. popⁿ average of 1st sample is greater than 2nd sample.

or, $H_1 : \mu_1 < \mu_2$ (Left tailed test) i.e. popⁿ mean of 1st sample is less than that of 2nd sample.

ii) Selection of Test Statistic

Based on sample size, no. of samples and parameters to test we select test statistic to test the hypothesis.

For eg:- for large sample $(n > 30)$ we select Z-statistic for testing the significance.

Test - statistic

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

In this step we get calculated value of test statistic.

iii) Level of significance

It is the risk or chance factor of making wrong decision about assumption made on statistic and parameters. It is denoted by $\alpha$. If the value of alpha($\alpha$) is not given in problem we should take $\alpha$ as 5% by default.

iv) Critical or table value

It is the table value of test statistic which is based on test statistic ~~and~~ no of sample and level of significance. Here we get tabulated value of test statistic

Date
Page

Decision making

Generally there are two approaches about making decision about hypothesis:

* Critical or table value approach
  - → If cal. value $\leq$ Tab. value, we accept $H_0$ & reject $H_1$
  - → If cal. value $>$ Tab. value, we reject $H_0$ & accept $H_1$
  - →

* p- value approach
  - → If ~~cal~~ p-value $\leq$ ~~Tab. value, we~~ $\alpha$, we reject $H_0$ & accept $H_1$
  - → If p-value $> \alpha$, we accept $H_0$ & reject $H_1$.

p-value

* For two tailed test
  p-value $= 2 \times p(Z > |Z_{cal}|)$

* For one-tailed test
  p-value $= p(Z > |Z_{cal}|)$

Z- test

(large sample test $n > 30$)

Case (I). Test of significance of single popⁿ mean

Test - statistic

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

If '$\sigma$' is not given, we use '$s$' instead of '$\sigma$'

Case II Test of significant difference bet two popⁿ means

Test - statistic

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

If $\sigma_1^2$ & $\sigma_2^2$ are not given we use $s_1^2$ & $s_2^2$ instead of $\sigma_1^2$ & $\sigma_2^2$

Scanned by CamScanner

Case III Test of significance of single popⁿ proportion

Test - statistic

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

where $p = \frac{x}{n}$ = sample proportion

$n$ = sample size

$x$ = No. of sample with certain attribute

$P$ = popⁿ proportion

$Q = 1 - P$

Case IV Test of significant difference betⁿ two popⁿ propⁿ

Test-statistic

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

where $p_1 = \frac{x_1}{n_1}$ = 1st sample prop

$p_2 = \frac{x_2}{n_2}$ = 2nd samp prop

$P$ = combined prop

$Q = 1 - P$

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

or $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

Scanned by CamScanner

d.f = degree of freedom

Date
Page

# t - Test

(Small - sample test $n \leq 30$).

Case (I) Test of significance of single pop$^n$ mean

Test - statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$ If s is given

or, $$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$ If data is given

Here $$s^2 = \frac{1}{n-1}\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right]$$

or, $$s^2 = \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right]$$ where $d = x - A$

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

Here d.f $= n - 1$

Case II :- Test of significance difference bet$^n$ two pop$^n$ means

Test - statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $$s_p^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$ If $s_1^2$ & $s_2^2$ are given

If data are given, then

$$S_p^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma x_1^2 - \frac{(\Sigma x_1)^2}{n_1} + \Sigma x_2^2 - \frac{(\Sigma x_2)^2}{n_2}\right]$$

or $$S_p^2 = \frac{1}{n_1 + n_2 - 2}\left(\Sigma d_1^2 - \frac{(\Sigma d_1)^2}{n_1} + \Sigma d_2^2 - \frac{(\Sigma d_2)^2}{n_2}\right]$$

where $d_1 = x_1 - A$, $d_2 = x_2 - B$

Here d.f $= n_1 + n_2 - 2$

Scanned by CamScanner

Case III Paired t-Test

Test-statistic

$$t = \frac{\bar{d}}{S_d/\sqrt{n-1}}$$ If 'Sd' is given

or $$t = \frac{\bar{d}}{S_d/\sqrt{n}}$$ If data are given

Here $d = x - y$

$x$ = score before

$y$ = score after

$$\bar{d} = \frac{\Sigma d}{n}$$

$$S_d^2 = \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right]$$

Here d.f $= n-1$

Case (iv) Test of significance of Sample corrn coeff

Test-statistic

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

where $r$ = corrn



Scanned by CamScanner

Ex - 2

) A sample of size 400 was drawn and the sample mean was found to be 99. Test whether this sample could have come from a normal population with mean 100 and standard deviation 8 at 5% level of significance.

Soln

Sample size $(n) = 400$

Sample mean $(\bar{x}) = 99$

S.D $(\sigma) = 8$

level of significance $(\alpha) = 5\%$

i) To test whether the sample could have come from popⁿ with mean 100 or not we set-up hypothesis as

$H_0 : \mu = 100$ i.e ~~pop~~ sample come from popⁿ with mean 100

Vs $H_1 : \mu \neq 100$ i.e. sample is not taken from popⁿ with mean 100.

ii) Under $H_0$ ; test - statistic

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

$$= \left| \frac{99 - 100}{8/\sqrt{400}} \right|$$

$\therefore |Z_{cal}| = 2.5$

Here ~~level of s~~ For two tailed test

$Z_{tab} = 1.96$

Decision

Since, $Z_{cal} > Z_{tab}$ we reject $H_0$ & accept $H_1$ with the conclusion that sample is not taken from popⁿ with mean 100.

Scanned by CamScanner

5) The mean life time of 400 laptop cells produced by a company is found to be 1570 hours with S.D of 150 hours. Test the hypothesis that the mean life time of the laptop cells produced by the company is 1600 hours against the alternative hypothesis that it is greater than 1600 hours at 1% level of significance.

Sol$^n$

Sample size $(n) = 400$

" mean $(\bar{x}) = 1570$

Pop$^n$ " $(\mu) = 1600$

Sample S.D $(s) = 150$

level of significance $(\alpha) = 1\%$

i) To test whether the mean lifetime of laptop cells is greater than 1600 hours or not we set up hypothesis as

$H_0 : \mu = 1600$ i.e. pop$^n$ average life of cells is 1600 hr

Vs $H_1 : \mu > 1600$ ie pop$^n$ average life of cells is greater than 1600 hrs.

ii) Under $H_0$, test-statistic is

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1570 - 1600}{150/\sqrt{400}} = -4.$$

$$|Z_{cal}| = 4$$

iii) level of significance $(\alpha) = 1\%$

For one-tailed test;

$$Z_{tab} = 2.33$$

Decision

Since, $Z_{cal} > Z_{tab}$, we reject $H_0$ and accept $H_1$ with conclusion that pop$^n$ average life of cells is

In a certain factory there are two independent process for manufacturing the same item. The average weight in a sample of 250 items produced from one process is found to be 120 gms with a S.D of ~~12.5 with a~~ 12 gms, while the other process are 124 and 14 in a sample of 400 items. Is the difference between the mean weights significant at 10% level of sig. Use p-value method.

Sol$^n$

For 1st Process

$n_1 = 250$

$\bar{x}_1 = 120, s_1 = 12$

For 2nd process

$n_2 = 400$

$\bar{x}_2 = 124, s_2 = 14$

level of sig ($\alpha$) = 10%

To test whether the difference in avg weight is sig. or no we set up hypothesis as

$H_0 : \mu_1 = \mu_2$ ie there is no sig. diff in avg wt of 2 process

Vs. $H_1 : \mu_1 \neq \mu_2$ ie there is sig diff in avg wt of 2 processes

Under $H_0$, test-statistic

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{120 - 124}{\sqrt{\frac{144}{250} + \frac{196}{400}}} = -3.87$$

$|Z_{cal}| = 3.87$

For two tailed test

$Z_{tab} = 1.645$

Decision

Since $Z_{cal} > Z_{tab}$, we reject $H_0$ & accept $H_1$ with the conclusion that there is significant diff in avg wt of 2 processes

Scanned by CamScanner

Using p - value method

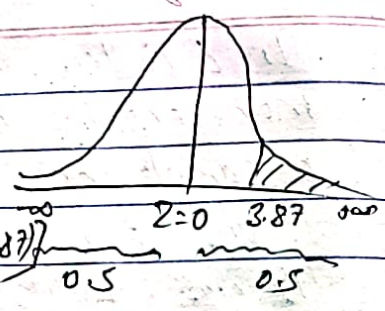
p-value $= 2 \times P(z > |z_{cal}|)$

$= 2 \times P(z > 3.87)$

$= 2[P(0<z<\infty) - P(0<z<3.87)]$

$= 2(0.5 - 0.4995)$

p-value $= 0.0001$




Decision

Since p-val $< \alpha$, we reject $H_0$ & accept $H_1$ - - - -

17) A sample of 400 male students is found to have a 171.38 cm. Can it be reasonably regarded as a sample from a large population with mean 171.17 cm and S.D of 3.30 cm. Use confidence limit to draw conclusion

Sol/ Sample size $(n) = 400$

" mean $(\bar{x}) = 171.38$

Popn " $(\mu) = 171.17$

" S.D $(\sigma) = 3.30$

i) To test whether the sample is taken from given popn or not we set up hypothesis as

$H_0 : \mu = 171.17$ i.e. sample is taken from given popn

Vs $H_1 : \mu \neq 171.17$ i.e. sample is not taken from given popn.

Under $H_0$ test statistic

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{171.38 - 171.17}{3.30/\sqrt{400}} = 1.27$$

Here level of sig($\alpha$) = 5% (Let)

For two tailed test

$$Z_\alpha = 1.96$$

Here, c.I for sample mean $= \mu \pm Z_\alpha \, S.E(\bar{x})$

$$= \mu \pm Z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$= 171.17 \pm 1.96 \times \frac{3.3}{\sqrt{400}}$$

$$= 171.17 \pm 0.3234$$

$$= (170.8466 - 171.4934)$$

Decision

Since $\bar{x} \in (170.8466 - 171.4934)$

So, we accept $H_0$ & reject $H_1$ with the conclusion that sample is taken from given popn.

Two research laboratories have independently produced drugs that provide relief to arthritis suffers. The first drug was tested on a group of 90 arthritis victims and produced an average of 8.5 hours relief with a SD of 1.6 hours. The second drug was tested on 80 arthritis victims producing an average of 7.9 hours of relief with a S.D of 2.1 hr. At 5% level of significance does the second drug provide a significantly shorter period of relief?

Soln For 1st drug

$n_1 = 90$

$\bar{x}_1 = 8.5$

For 2nd drug

$n_2 = 80$

$\bar{x}_2 = 7.9$

Scanned by CamScanner

Note :- To test about proportion, [illegible]
we should use z-test

$S_1 = 1.8$ $S_2 = 2.1$

level of sig ($\alpha$) = 5%

To test whether the second drug provides significantly shorter period of relief or not we set up hypothesis as,

$H_0: \mu_1 = \mu_2$ i.e. there is no sig diff in relief period of two drugs.

Vs $H_1: \mu_1 > \mu_2$ i.e. 2nd drugs provides significantly shorter period of relief.

i) Under $H_0$, test statistic

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{8.5 - 7.9}{\sqrt{\frac{1.8^2}{90} + \frac{2.1^2}{80}}} = 1.98$$

$Z_{cal} = 1.98$

Here, level of sig ($\alpha$) = 5%

For one tailed test

$Z_{tab} = 1.645$

Decision :-

Since, $Z_{cal} > Z_{tab}$, we reject $H_0$ & accept $H_1$ with the conclusion that second drug provide significantly shorter period of relief.

29) A dice was thrown 9000 times and of these 3220 yielded a 3 or 4. Is the consistent with the hypothesis that the dice is unbiased?

Sol/ Here, we have;

Sample size ($n$) = 9000

No. of times dice shows 3 or 4 ($x$) = 3220

Sample propⁿ ($p$) = $\frac{x}{n}$ = $\frac{3220}{9000}$ =

Scanned by CamScanner

Popⁿ prop $(P) = \frac{2}{6} = \frac{1}{3}$

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

To test dice is biased or not, we set up hypothesis as
$H_0 : p = 1/3$ i.e. dice is unbiased
Vs $H_1 : p \neq \frac{1}{3}$ (two tailed) ie dice was biased.

Under $H_0$ test statistic

$$Z = \frac{p - P}{\sqrt{PQ/n}} = \frac{\frac{322}{900} - 1/3}{\sqrt{\frac{1}{3} \times \frac{2}{3} \times \frac{1}{900}}} = 4.92$$

$Z_{cal} = 4.92$

Here level of sig $(\alpha) = 5\%$.
For two tailed test, $Z_{tab} = 1.96$

Decision
Since $Z_{cal} > Z_{tab}$, we reject $H_0$ & accept $H_1$ with the conclusion that Dice was biased.

The coordinator in a college claimed that at least 98% of the students submit their assignment on time. Taking the sample of 250 students, 15 were not submitting assignment in whole semister. Test the claim at 10% level of significance.

Sol$^n$/
Here, we have,
Popⁿ prop $(P) = 98\% = 0.98$
$\Rightarrow Q = 1 - 0.98 = 0.02$
Sample size $(n) = 250$

Scanned by CamScanner

No. of students who submit assignment $(x) = 250 - 15 = 235$

Sample prop $(p) = \frac{x}{n} = \frac{235}{250} = 0.94$ i)

i) To test whether the co-ordinator's claim is verify or not we setup hypothesis as

$H_0 : p = 0.98$ ie claim is verified ii)

Vs $H_2 : p < 0.98$ (one tailed) ie claim is not verified.

ii) ~~Here~~ Under $H_0$, test-statistic is

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91 - 0.98}{\sqrt{\frac{0.98 \times 0.02}{252}}} = -4.5$$

$|Z_{cal}| = 4.5$ iii)

iii) level of sig $(\alpha) = 10\%$

For one tailed test,

$Z_{tab} = 1.28$

Decision

Since $Z_{cal} > Z_{tab}$; we reject $H_0$ and accept $H_1$, with the conclusion that claim is not verified. 38)

34) A machine produces 20 defective articles in a batch of 400. After overholding it produces 10 defectives in a batch of 300. Has the machine improved?

Sol/

Befor overhouling

$n_1 = 400$

$x_1 = 20$

$P_1 = \frac{x_1}{n} = \frac{20}{400} = 0.05$

After overhouling

$n_2 = 300$

$x_2 = 10$

$P_2 = \frac{x_2}{n} = \frac{10}{} = ($

Scanned by CamScanner

) To test whether the machine has been improved or not we set up hypothesis as

$H_0 : p_1 = p_2$ i.e. machine has not been improved

Vs $H_1 : p_1 > p_2$ i.e. machine has been improved

Under $H_0$, test-statistic is

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

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

$= \frac{20+10}{400+300} = \frac{30}{700} = \frac{3}{70}$

$Q = 1 - P = 1 - \frac{3}{70} = \frac{67}{70}$

$$= \frac{\frac{1}{20} - \frac{1}{30}}{\sqrt{\frac{3}{70} \times \frac{67}{70}\left(\frac{1}{400} + \frac{1}{300}\right)}}$$

$Z_{cal} = 1.1$

Here, level of sig ($\alpha$) = 5%

For one tailed test

$Z_{tab} = 1.645$

Decision : Since, $Z_{cal} < Z_{tab}$ we reject $H_1$ & accept $H_0$ with the conclusion that machine has not been improved.

In a sample of 600 students of a certain college 400 are found to use dot pens. In another college from a sample of 900 students 450 were found to use dot pens. Test whether the two colleges are significantly diff w.r.t the habit of using the dot pens at 1% level of sig? Use p value method.

Sol/

P.T.O

For 1st college

$n_1 = 600,\ x_1 = 400$

$p_1 = \frac{x_1}{n_1} = \frac{400}{600} = \frac{2}{3}$

For 2nd college

$n_2 = 900,\ x_2 = 450$

$p_2 = \frac{x_2}{n_2} = \frac{450}{900} = \frac{1}{2}$

$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{400+450}{600+900} = \frac{850}{1500} = \frac{17}{30}$

$Q = 1 - P = 1 - \frac{17}{30} = \frac{13}{30}$

i) To test whether there is sig. diff in dot pen users of two colleges or not, we setup hypothesis as

$H_0 : P_1 = P_2$ ie there is no sig diff in dot pen users of 2 colz

Vs $H_1 : P_1 \neq P_2$ ie there is sig diff in dot pen users of 2 colz

ii) Unda $H_0$, test-statistic;

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

$$Z_{cal} = \frac{2/3 - 1/2}{\sqrt{\frac{17}{30} \times \frac{13}{30} \times \left(\frac{1}{600} + \frac{1}{900}\right)}} = 6.38$$

iii) level of sig $(\alpha) = 1\%$.

We know for two tailed test,

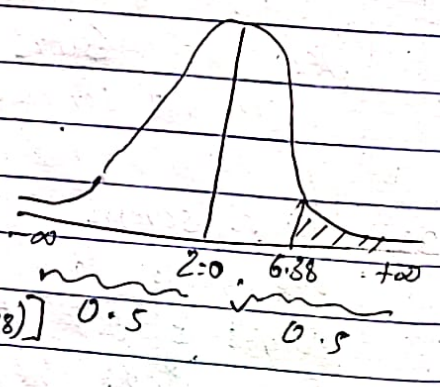
$p\text{-value} = 2 \times P(Z > |Z_{cal}|)$

$= 2 \times P(Z > 6.38)$

$= 2[P(0 < Z < \infty) - P(0 < Z < 6.38)]$

$= 2(0.5 - 0.5)$

$= 0$




Decision

Since, p-value < $\alpha$, we reject $H_0$ and accept $H_1$ with the conclusion that there is sig. diff in dot pens users in two colz.

Scanned by CamScanner

Rainfall records of a particular place for last 12 years for the month of July showed that average rainfall was 50 mm and standard deviation of 30 mm. Do you agree that the average rainfall at the place was less than 51.2 mm Use 10% level of significance?

Sol$^n$/ Here we have

Sample size $(n) = 12$

$\bar{x} = 50$

$s = 30$

Pop$^n$ mean $(\mu) = 51.2$

To test whether the population mean is less than 51.2 mm or not we set up hypothesis as

i) $H_0 : \mu = 51.2$ mm ie pop$^n$ mean rainfall is 51.2 mm

Vs $H_1 : \mu < 51.2$ ie pop$^n$ " " is less than 51.2 mm

ii) Under $H_0$, test - statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{50 - 51.2}{30/\sqrt{12-1}} = -0.132$$

$|t_{cal}| = 0.132$

Here, level of sig. $(\alpha) = 10\%$ d.f $= n-1 = 12-1 = 11$

For, one tailed test,

$t_{tab} = 1.363$

Decision

Since $t_{cal} < t_{tab}$ we accept $H_0$ & reject $H_1$ with the conclusion that pop$^n$ mean rainfall is 51.2 mm.

Scanned by CamScanner

40) Ten patients are selected at random from a popⁿ of patients and their blood pressure recorded are as follows
125, 147, 118, 145, 140, 128, 155, 150, 160, 149
Do the data support the hypothesis that the popⁿ average blood pressure of patients is 135 ? Use 5% level of significance.

Sol$^n$ Here we have

Sample size $(n) = 10$

Popⁿ mean $(\mu) = 135$

i) To test whether popⁿ average blood pressure is 135 or not we set up hypothesis as

$H_0 : \mu = 135$ ie popⁿ average blood pressure is 135

Vs $H_1 : \mu \neq 135$ ie " " " " is not 135

ii) Calculation table for $\bar{x}$ & s

| $x$ | $d = x - A$ | $d^2$ |
|---|---|---|
| 125 | -15 | 225 |
| 147 | 7 | 49 |
| 118 | -22 | 484 |
| 145 | 5 | 25 |
| 140 | 0 | 0 |
| 128 | -12 | 144 |
| 155 | 15 | 225 |
| 150 | 10 | 100 |
| 160 | 20 | 400 |
| 149 | 9 | 81 |
| | $\Sigma d = 17$ | $\Sigma d^2 = 1733$ |

let $A = 140$

For $\bar{x}$

$$\bar{x} = A + \frac{\Sigma d}{n} = 140 + \frac{17}{10} = 141.7$$

Note:-

Scanned by CamScanner

For $s^2$

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

$$= \frac{1}{10-1}\left[1733 - \frac{17^2}{10}\right]$$

$$s = 13.76$$

Now,

$$t_{cal} = \frac{141.7 - 135}{13.76/\sqrt{10}}$$

$$\therefore t_{cal} = 1.54$$

Here level of sig $(\alpha) = 5\%$

$$d.f = n-1 = 10-1 = 9$$

For two tailed test

$$t_{tab} = 2.262$$

Decision

Since, $t_{cal} < t_{tab}$ we accept $H_0$ & reject $H_1$ with the conclusion that popⁿ average B.P is 135.

$\therefore$ $(1-\alpha)\%$ C.I for popⁿ mean for small sample

$$= \bar{x} \pm t_{\alpha, n-1} \; S.E(\bar{x})$$

where

$$S.E(\bar{x}) = \frac{s}{\sqrt{n}} \quad \text{If data are given}$$

$$= \frac{s}{\sqrt{n-1}} \quad \text{If 's' is given}$$

Scanned by CamScanner

45) A random sample of 10 bulbs has the following life in months ; 24, 26, 32, 28, 20, 20, 23, 34, 30 and 43. Obtain the 95% confidence limit for the popn mean life of bulbs.

Soln/ Here we have

Sample size $(n) = 10$

C.I $(1-\alpha) = 95\% \Rightarrow \alpha = 5\%$

$df = n-1 = 10-1 = 9$

$t_{tab} = 2.262$

Calⁿ table for $\bar{x}$ & $s$.

| $x$ | $d = x - A$ | $d^2$ |
|---|---|---|
| 24 | -6 | 36 |
| 26 | -4 | 16 |
| 32 | 2 | 4 |
| 28 | -2 | 4 |
| 20 | -10 | 100 |
| 20 | -10 | 100 |
| 23 | 7 | 49 |
| 34 | 4 | 16 |
| 30 | 0 | 0 |
| 43 | 13 | 169 |
| | $\Sigma d = -20$ | $\Sigma d^2 = 494$ |

For $\bar{x}$

$$\bar{x} = A + \frac{\Sigma d}{n} = 30 + \left(\frac{-20}{10}\right) = 28$$

For $s^2$

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

$$= \frac{1}{10-1}\left[494 - \frac{(-20)^2}{10}\right]$$

$$= 7.1024$$

Scanned by CamScanner

Now, $S.E(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{7.1024}{\sqrt{10}} = 2.246$

Now, 95% ~~of~~ C.I for pop$^n$ mean $= \bar{x} \pm t_{\alpha, n-1} S.E(\bar{x})$

$= 28 \pm 2.262 \times 2.246$

$= 28 \pm 5.08$

$= (22.92 - 33.08)$

The average number of articles produced by two machines per day are 200 and 250 with s.d 20 and 25 respectively on the basis of records of 25 days production. Can you regard both the machines equally efficient at 1% level of significance.

Sol$^n$

Here we have,

For 1st machine

$n_1 = 25, \bar{x}_1 = 200, s_1 = 020$

For 2nd machine

$n_2 = 25, \bar{x}_2 = 250, s_2 = 25$

To test whether the both machines are equally efficient or not. We set up hypothesis as.

$H_0 : \mu_1 = \mu_2$ ie two machines are equally efficient

Vs $H_1 : \mu_1 \neq \mu_2$ ie two machines are not equally efficient.

Under $H_0$ test - statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $s_p^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{25 \times 20^2 + 25 \times 25^2}{25 + 25 - 2}$

$= 533.854$

$$|t_{cal}| = \frac{200 - 250}{\sqrt{533.854\left(\frac{1}{25} + \frac{1}{25}\right)}}$$

$= 7.65$

Here, level of sig $(\alpha) = 1\%$

Scanned by CamScanner

$d.f = n_1 + n_2 - 2 = 25 + 25 - 2 = 48$

From table, $t_{tab} = 2.66$

Decision :-

Since, $t_{cal} > t_{tab}$, we reject $H_0$ & accept $H_1$ with the conclusion that two machine are not equally efficient.

50) Two new drugs A and B are given to two independent groups of 10 and 12 patient with heart disease respectively. The reduction of blood pressure due to the two new drugs A and B are given below:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Drug A | 7 | 16 | 14 | 9 | 10 | 11 | 6 | 8 | 10 | 9 | | |
| Drug B | 10 | 12 | 16 | 14 | 11 | 12 | 13 | 8 | 12 | 15 | 9 | 12 |

50) ~~Here~~

To test whether the drug A is less effective than drug B or not, we set up hypothesis as

$H_0 : \mu_1 = \mu_2$ ie two drugs are equall efficient

Vs $H_1 : \mu_1 < \mu_2$ ie two drugs are not equally efficient.

Under $H_0$ - test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{sp^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Calculation table for $\bar{x}_1, \bar{x}_2$ & $sp^2$

| $x_1$ | $x_2$ | $d_1 = x_1 - A$ | $d_1^2$ | $d_2 = x_2 - B$ | $d_2^2$ |
|---|---|---|---|---|---|
| 7 | 10 | -3 | 9 | -2 | 4 |
| 16 | 12 | 6 | 36 | 0 | 0 |
| 14 | 16 | 4 | 16 | 4 | 16 |
| 9 | 14 | -1 | 1 | 2 | 4 |

| | | | | | |
|---|---|---|---|---|---|
| 10 | 11 | 0 | 0 | -1 | 1 |
| 11 | 12 | 1 | 1 | 0 | 0 |
| 6 | 13 | -4 | 16 | 1 | 1 |
| 8 | 8 | -2 | 4 | -4 | 16 |
| 10 | 12 | 0 | 0 | 0 | 0 |
| 9 | 15 | 1 | 1 | 3 | 9 |
| | 9 | | | -3 | 9 |
| | 12 | | | 0 | 0 |
| | | $\Sigma d_1 = 0$ | $\Sigma d_1^2 = 83$ | $\Sigma d_2 = 0$ | $\Sigma d_2^2 = 60$ |

Here, $A = 10$, $B = 12$

Here, $\bar{x}_1 = A + \frac{\Sigma d_1}{n_1} = 10 + \frac{0}{10} = 0$

$\bar{x}_2 = B + \frac{\Sigma d_2}{n_2} = 12 + \frac{0}{12} = 12$

& $S_p^2 = \frac{1}{n_1+n_2-2}\left[\Sigma d_1^2 - \frac{(\Sigma d_1)^2}{n_1} + \Sigma d_2^2 - \frac{(\Sigma d_2)^2}{n_2}\right]$

$= \frac{1}{10+12-2}(84 - 0 + 60 - 0)$

$\therefore S_p^2 = 7.2$

Now

$t_{cal} = \frac{10-12}{\sqrt{7.2\left(\frac{1}{10}+\frac{1}{12}\right)}} = -1.74$

$|t_{cal}| = 1.74$

Here, level of sig $(\alpha) = 5\%$

d.f $= n_1+n_2-2 = 10+12-2 = 20$

For one tailed test

$t_{tab} = 1.725$

Decision: Since $t_{cal} > t_{tab}$, we reject $H_0$ & accept $H_1$, with the conclusion that Drug A is less efficient than Drug B.

Scanned by CamScanner

5b) Memory capacity of 10 students was tested before and after training state whether the training was effective or not from the following scores.

| Roll no | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Before training | 12 | 14 | 11 | 8 | 7 | 10 | 3 | 0 | 5 | 6 |
| After training | 15 | 16 | 10 | 7 | 5 | 12 | 10 | 2 | 3 | 8 |

Sol/

To test whether the training was efficient or not we setup hypothesis as

$H_0 : \mu_x = \mu_y$ ie training wasn't effective

Vs $H_1 : \mu_x < \mu_y$ ie training was effective

Under $H_0$, test-statistic

$$t = \frac{\bar{d}}{S_d/\sqrt{n}}$$

Calc° table for $\bar{d}$ & $S_d$

| X | Y | $d = Y - X$ | $d^2$ |
|---|---|---|---|
| 12 | 15 | 3 | 9 |
| 14 | 16 | 2 | 4 |
| 11 | 10 | -1 | 1 |
| 8 | 7 | -1 | 1 |
| 7 | 5 | -2 | 4 |
| 10 | 12 | 2 | 4 |
| 3 | 10 | 7 | 49 |
| 0 | 2 | 2 | 4 |
| 5 | 3 | -2 | 4 |
| 6 | 8 | 2 | 4 |
| | | $\Sigma d = 12$ | $\Sigma d^2 = 84$ |

Here, ~~x~~ = For $\bar{d}$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{12}{10} = 1.2$$

& $$S_d^2 = \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right] = \frac{1}{10-1}\left[84 - \frac{12^2}{10}\right] = 2.781$$

Now,

$$t_{cal} = \frac{1.2}{2.781/\sqrt{10}} = 1.36$$

$\therefore |t_{cal}| = 1.36$

Here, level of sig $(\alpha) = 5\%$ (lel)

d.f & $n-1 = 10-1 = 9$

For one tailed test,

$t_{tab} = 1.83$

Decision

Since $t_{cal} < t_{tab}$, we accept $H_0$ & reject $H_1$ with the conclusion that training was not effective

Ten students were given intensive training on python for a month. The scores obtained in test I and V are given below

| of stds | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| in 1st test | 50 | 52 | 53 | 60 | 65 | 67 | 48 | 69 | 72 | 80 |
| in 5th test | 65 | 55 | 65 | 65 | 60 | 75 | 82 | 49 | 74 | 86 |

So, Does the score from test I to V show improvement. ($\alpha = 5\%$)

To test whether the test I to V shown an improvement or not we setup hypothesis as

$H_0 : \mu_x = \mu_y$ i.e. there is no improvement

$H_1 : \mu < \mu_y$ i.e there is an improvement

Under $H_0$, test statistic

$$t = \frac{\bar{d}}{S_d/\sqrt{n}}$$

Scanned by CamScanner

Calculation table for $\bar{d}$ and $S_d$

| X | Y | $d = Y - X$ | $d^2$ |
|---|---|---|---|
| 50 | 65 | 15 | 225 |
| 52 | 55 | 3 | 9 |
| 53 | 65 | 12 | 144 |
| 60 | 65 | 5 | 25 |
| 65 | 60 | -5 | 25 |
| 67 | 67 | 0 | 0 |
| 48 | 49 | 1 | 1 |
| 69 | 82 | 13 | 169 |
| 72 | 74 | 2 | 4 |
| 80 | 86 | 6 | 36 |
| | | $\Sigma d = 52$ | $\Sigma d^2 = 638$ |

For $\bar{d}$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{52}{10} = 5.2$$

& $$S_d^2 = \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right]$$

$$= \frac{1}{10-1}\left[84 - \frac{12^2}{10}\right]$$

$$= \frac{1}{10-1}\left[638 - \frac{(52)^2}{10}\right]$$

$$= 40.84$$

$$S_d = 6.39$$

Now

$$t = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{5.2}{6.39/\sqrt{10}} = 2.57$$

$$|t_{cal}| = 2.57$$

Scanned by CamScanner

Here $\alpha = 5\%$

$d.f = n-1 = 10-1 = 9$

For one tailed test

$t_{tab} = 1.833$

Decision

Since $t_{cal} > t_{tab}$ we reject $H_0$ & accept $H_1$ with the conclusion that there is improvement.

Scanned by CamScanner

Date
Page

Scanned by CamScanner

# Unit - 3

## Non Parametric Test

Run :- It is the sequence of same or identical sign. It is used to test the uniformity or randomness of given data series.

### Run Test

* For small sample ($n_1, n_2 \leq 20$)

Test - statistic

test - statistic = No. of runs ($r$)

where, $n_1$ = no. of data $\leq M_0$

$n_2$ = . . . . . $> M_0$

Table $r$ critical value

For given level of significance ($\alpha$), value of $n_1$ & $n_2$ we obtain $\underline{r}$ (lower value of $r$) & $\bar{r}$ (upper value of $r$) from respective table.

Decision

If $r \in (\underline{r}\ \bar{r})$, we accept $H_0$ & reject $H_2$ otherwise reject $H_0$.

* For large sample ($n_1$ or $n_2 > 20$)

Test - statistic

$$Z = \frac{r - \mu_r}{\sigma_r}$$

where $\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1$

& $\sigma_r^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$

Decision * If $Z_{cal} \leq Z_{tab}$, we accept $H_0$ & reject $H_1$

* If $Z_{cal.} > Z_{tab}$, we reject $H_0$ & accept $H_1$.

Scanned by CamScanner

9.) In 30 toss of a coin the following sequence of heads (H) and tails (T) is obtained

H T T H T H H H T H H T H T H H T H T T H T H H T H T

Test at 0.05 level of significance level whether the sequence is random.

Sol/

To test whether the sequence is random or not we set up hypothesis as

$H_0$ : sequence is random

Vs $H_1$ : sequence is not random

Here we have

No of heads $(n_1) = 16$

No of tails $(n_2) = 14$

No. of runs $(r) = 22$

For sample (Sman)

Test-statistic = No of runs $(r) = 22$

For table value

Here $\alpha = 5\%$.

$\underline{r} = 10$ $\bar{r} = 22$

Decision

Since $r \in (\underline{r}\ \bar{r})$ we reject $H_0$ & accept $H_1$ with the conclusion that sequence is not random.

If $\alpha$- value is others than 5% we must use Z-test for run-test.

12.)

Scanned by CamScanner

) The following are the number of II students absent from a college on 24 consecutive colz days.
29, 25, 31, 28, 30, 28, 33, 31, 35, 27, 31, 33, 35, 28, 36, 30, 33, 26, 30, 28, 32, 31, 38 and 27. Test for randomness. ($\alpha = 0.01$)

Sol$^n$

To test whether the given sequence of number is random or not we setup hypothesis as

$H_0$ : sequence of number is random

Vs $H_1$ : sequence of number is not random

Arranging the data in ascending order.

25, 26, 27, 28, 28, 28, 28, 29, 29, 30, 30, 30, 31, 31, 31, 31, 32, 33, 33, 33, 35, 35, 36, 38

$N = 24$, Median $= \left(\frac{N+1}{2}\right)^{th} = \frac{25}{2} = 12.5^{th} = \frac{30+31}{2}$
$= 30.5$

Now, let, A = value $< M_d$, B = value $> M_d$

Here, we have

A AB AAA BBB A BBB A BAB AAA BBB A

Here, $n_1$ = no. of A = 12

$n_2$ = no. of B = 12

$r$ = no. of runs = 13

For $\alpha = 0.01$ we use Z-test

Test statistic

$$Z = \frac{r - \mu_r}{\sigma_r}$$

where $\mu_r = \frac{2n_1 n_2}{n_1 + n_2} + 1 = \frac{2 \times 12 \times 12}{12 + 12} + 1 = 13$

Scanned by CamScanner

$$\sigma_r = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1+n_2)^2 (n_1+n_2-1)}}$$

$$= 2.395$$

Now

$$Z_{cal} = \frac{13-13}{2.395} = 0$$

For $\alpha = 1\%$ & for two tailed test

$$Z_{tab} = 2.575$$

Decision

Since, $Z_{cal} < Z_{tab}$, we accept $H_0$ & reject $H_1$ with the conclusion that given sequence is random.

13.) The height (in inches) of 15 football players are as follows 65.3, 67.7, 68.4, 71, 70.2, 67, 69.8, 64, 67.6, 71.5, 64.4, 63.6, 66.8, 70.4, 69.

Test whether the order of the heights is random or not. Use 1% level of significance.

Sol$^n$

Date
Page

Scanned by CamScanner

# Binomial Test

It is used to test whether the two groups of data are equally probable or not.

For small sample $(n \leq 25)$

Test - statistic

$x_0 = \min\{n_1, n_2\}$

where $n_1$ = no. of symbol of 1st kind

$n_2$ = " " " " 2nd kind

Then, p-value is computed as

$$p = \text{prob}\ (x \leq x_0) = \sum_{x=0}^{x_0} p(x) = \sum_{x=0}^{x_0} {}^nC_x p^x q^{n-x}$$

Decision :-

For two tailed test

$2p > \alpha$, we accept $H_0$ & reject $H_1$

For one tailed test

$p > \alpha$, we accept $H_0$ & reject $H_1$.

For large sample $(n > 25)$

we use z-test as

$$Z = \frac{x_0 - np}{\sqrt{npq}}$$

we adjust Z as

$$Z = \frac{(x_0 + 0.5) - np}{\sqrt{npq}} \quad \text{If } x_0 < np$$

& $$Z = \frac{(x_0 - 0.5) - np}{\sqrt{npq}} \quad \text{If } x_0 > np$$

where 0.5 is continuity correction factor.

Decision: * If $Z_{cal} \leq Z_{tab}$ we accept $H_0$ & reject $H_1$

* If $Z_{cal} > Z_{tab}$ we reject $H_0$ & accept $H_0$

Scanned by CamScanner

The following are defective (D) and non-defective (N) electric cable produced in the given order by a certain machine in a manufacturing industry :- D N D D N D N D N N D N N D D N N D N D D N N N D. Test whether defective & ~~cab~~ non defective cables are equally produced. Use binomial test at the 5% level of sig.

Sol:

To test whether the defective and non defective cables are equally produced or not, we setup hypothesis as

$H_0$ : equally produced

$H_1$ : not equally produced

Here. $n = 26$, $p = 1/2$, $q = 1 - 1/2 = 1/2$

$n_1$ = no of def = 12, $n_2$ = no of non-defective = 14

$x_0$ = min of $\{n_1, n_2\}$ = min of $\{12, 14\}$ = 12

For large sample, test statistic

$$Z = \frac{x_0 - np}{\sqrt{npq}}$$

Here $np = 26 \times 1/2 = 13$

$npq = 26 \times 1/2 \times 1/2 = 6.5$

Since $x_0 < np$, So adjust Z is

$$Z = \frac{(x_0 + 0.5) - np}{\sqrt{npq}}$$

$$= \frac{(12 + 0.5) - 13}{\sqrt{6.5}}$$

$\therefore |Z_{cal}| = 0.196$

Here, level of sig $(\alpha) = 5\%$

For two tailed test

$Z_{tab} = 1.96$

Decision:

Since $Z_{cal} < Z_{tab}$ we accept $H_0$ & reject $H_1$ with the conclusion that defective & non defective cables are equally produced.

Scanned by CamScanner

5) 100 workers of sales house willing to express opinion of cell phones, 40 expressed preferences to brand Nokia and 60 for brand Oppo. Are both brand of cell phone equally popular at 5% level of significance. Use Binomial test.

16)

Scanned by CamScanner

# Kolmogrov Smirnov Test

It is used to test the goodness of fit for observed data

## Test Statistic

* For small sample

$D_0 = \text{Max } |F_0 - F_e|$

where $F_0$ = observed relative cum. freq

$F_e$ = Expected " " "

## Table Value (Critical Value)

Based on $\alpha$ & sample size we obtain value of $D_{n,\alpha}$ from respective table.

## Decision

If $D_0 \geq D_{n,\alpha}$ Reject $H_0$ & accept $H_1$ otherwise accept $H_0$

) In a certain computer hardware manufacturing industry six different types of machines are working to cut pieces of wires. The number of wires of unequal length recorded in a day is as follows:

| Machine | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| No. of wire | 2 | 0 | 4 | 8 | 5 | 11 |

Scanned by CamScanner

Date
Page

17)

Vs

Scanned by CamScanner

The number of virus infected computers of 5 different capacity of hard disk is given below

| Cap of hd (GB) | 500 | 320 | 1000 | 2000 | 4000 |
|---|---|---|---|---|---|
| No. of virus inf. | 11 | 15 | 20 | 3 | 1 |

Test whether the computers of five harddisk are uniformly infected using Kolmogrov Smirov test.

Sol

To test whether the computers of five harddisk are uniformly infected, ~~we~~ or not, we set up hypothesis as

$H_0$ : All harddisk are uniformly infected

$H_1$ : " " " " not infected.

Under $H_0$, test statistic

$$D_0 = \text{Max} \, |F_0 - F_e|$$

Calc. table for $F_0$ & $F_e$

| Cap | $f_o$ | $c/f_o$ | $F_o$ | $f_e$ | $c/f_e$ | $F_e$ | $\lvert F_o - F_e \rvert$ ~~where~~ |
|---|---|---|---|---|---|---|---|
| 500 | 11 | 11 | 11/50 | 10 | 10 | 10/50 | 1/50 |
| 320 | 15 | 26 | 26/50 | 10 | 20 | 2/5 | 6/50 |
| 1000 | 20 | 46 | 46/50 | 10 | 30 | 3/5 | 16/50 |
| 2000 | 3 | 49 | 49/50 | 10 | 40 | 4/5 | 9/50 |
| 4000 | 1 | 50 | 1 | 10 | 50 | 1 | 0 |
| | N=50 | | | N=50 | | | |

where $F_o = \frac{c/f_o}{N}$

$F_e = \frac{c/f_e}{N}$

Now

$D_0 = 16/50 = 0.32$

Here, level of sig $(\alpha) = 5\%$.

For two tailed test

$$D_{n,\alpha} = \frac{1.36}{\sqrt{n}} = \frac{1.36}{\sqrt{50}} = 0.192$$

Decision :- Since $D_0 > D_{n,\alpha}$ we accept $H_1$ & reject $H_0$ with the conclusion that all harddisk are not uniformly infected.

Scanned by CamScanner

18 Sol:

To test whether the chimpanzees recognize colours or not we setup hypothesis as

$H_0$ : chimpanzees recognize color

vs $H_1$ : " donot recognize color

Under $H_0$, test statistic

$D_0 = \text{Max} |F_0 - F_e|$

Calc. table for $F_0$ & $F_e$

| Matched set | $f_0$ | $cf_0$ | $F_0$ | $f_e$ | $cf_e$ | $F_e$ | $\lvert F_0 - F_e \rvert$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 1/20 | 4 | 4 | 4/20 | 3/20 |
| 1 | 0 | 1 | 1/20 | 4 | 8 | 8/20 | 7/20 |
| 2 | 5 | 6 | 6/20 | 4 | 12 | 12/20 | 6/20 |
| 3 | 7 | 13 | 13/20 | 4 | 16 | 16/20 | 3/20 |
| 4 | 7 | 20 | 1 | 4 | 20 | 1 | 0 |
| $N = 20$ | | | | $N = 20$ | | | |

$\therefore D_0 = 7/20 = 0.35$

$\alpha = 5\% = 0.05$

For two tailed test

$D_{n,\alpha} = \frac{1.36}{\sqrt{n}} = \frac{1.36}{\sqrt{20}} = 0.304$

Decision

Since $D_0 > D_{n,\alpha}$

Scanned by CamScanner

# Two independent Sample Test

## Median Test

It is a non-parametric test used to test the significant difference between two independent distribution

* For small sample $(n_1 \le 10, n_2 \le 10)$

Test - statistic

$$P(A = a) = \frac{{}^{n_1}C_a \cdot {}^{n_2}C_{k-a}}{{}^{n}C_k}$$

where $n = n_1 + n_2$

$k = \frac{n_1 + n_2}{2} = \frac{n}{2}$

Then,

Test - statistic $= P(A \ge a)$

$a$ = no of samples $\le M_d$ for 1st samples

$M_d$ = Median of combined sample

* For large sample $(n_1 > 10, n_2 > 10)$

We use $\chi^2$ - test statistic as

$$\chi^2 = \frac{N(ad - bc)^2}{(a+c)(a+b)(b+d)(c+d)}$$

where $N = a+b+c+d$

$a$ = no of sample $\le M_d$ for 1st sample

$b$ = " " " $> M_d$ " 1st "

$c$ = no of sample $\le M_d$ " 2nd "

$d$ = " " $> M_d$ " " "

Decision

* For small sample

p-value $\le \alpha$ we reject $H_0$ & accept $H_1$ otherwise accept $H_0$

* For large sample

$\chi^2_{cal} \le \chi^2_{tab}$, we accept $H_0$ & reject $H_1$ otherwise reject $H_0$.

Scanned by CamScanner

7.

Sol/

To test whether two training equally effective or not we set up hypothesis as

$H_0$ : two trainings are equally effective

$H_1$ : two training are not equally effective

For median

Arranging the data in ascending order

30, 32, 34, 39, 40, 42, 43, 44, 45, 46, 47, 47, 48, 49, 50, 51, 55, 55, 59, 71

$N = 20$

$M_d = \left(\frac{N+1}{2}\right)^{th}$ value

$= \frac{21}{2} = 10.5^{th}$ value

$M_d = \frac{46+47}{2} = 46.5$

Here, $n_1 = 10$ $n_2 = 10$ $k = \frac{n_1+n_2}{2} = \frac{10+10}{2} = 10$

$n = n_1 + n_2 = 10+10 = 20$

$a$ = no. of samples $\leq M_d$ of 1'st sample

$\therefore a = 7$

Now, test statistic

$p\text{-value} = P(A \geq a) = P(A \geq 7)$

$= P(7) + P(8) + P(9) + P(10)$

$= \frac{{}^{10}C_7 \times {}^{10}C_3 + {}^{10}C_8 \times {}^{10}C_2 + {}^{10}C_9 \times {}^{10}C_1 + {}^{10}C_{10}\, {}^{10}C_0}{{}^{20}C_{10}}$

p-value =

8)

Scanned by CamScanner

Here, level of sig ($\alpha$) = 5% = 0.05

Decision

Since p-value > $\alpha$ we accept $H_0$ & reject $H_1$ with the conclusion that there is no significant difference between two trainings.

50)

To test whether there is significant difference betn two types of tubes or not we set of hypothesis as

$H_0$ : there is no sign diff betn two tubes

vs $H_1$ : there is sig diff betn two tubes

For Median

7, 11, 17, 20, 24, 25, 25, 28, 30, 34, 39, 41, 45, 47, 59, 61, 68, 75, 76, 83, 96, 99, 101, 108, 125, 140, 165, 178, 200, 283

$M_d = \left(\frac{N+1}{2}\right)^{th}$ value $= \left(\frac{30+1}{2}\right)^{th}$ value $= 15.5^{th}$ value

$= \left(\frac{59+61}{2}\right) = 60$

Here, $n_1 = 14$, $n_2 = 16$

For large sample

| Sample | Sample $\leq M_d$ | $> M_d$ | Total |
|---|---|---|---|
| I | 5 | 9 | 14 |
| II | 10 | 6 | 16 |
| Total | 15 | 15 | 30 |

Here, test-statistic

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

$$\therefore \chi^2_{cal} = \frac{30(5\times 6 - 9\times 10)^2}{14\times 15\times 15\times 16} = 2.14$$

Scanned by CamScanner

$d.f = (r-1)(c-1) = (2-1)(2-1) = 1$

Here level of sig $(\alpha) = 5\% = 0.05$

$\chi^2_{tab} = 3.841$

Decision :-

Since $\chi^2_{cal} < \chi^2_{tab}$ we accept $H_0$ & reject $H_1$ with the conclusion that ....

# KOLMOGOROV SMIRNOV TEST

→ For two sample ←

It is used to test whether two samples are from same population or not.

* For small sample $n_1 = n_2 < 40$, $n_1$ & $n_2 \leq 20$ $n_1 + n_2$

Test statistic

$D_0 = \text{Max of } \{ F(x) - F(y) \}$

Where $F(x)$ = Relative cum. freq of 1st sample

$F(y)$ = " " " " 2nd "

Critical value

Based on $\alpha$, $n_1$, $n_2$, we obtain $D_{(n_1, n_2), \alpha} = D_\alpha$

Decision

If $D_0 \geq D_\alpha$, we reject $H_0$ & accept $H_2$ otherwise accept $H_0$

*) For large sample

$\chi^2 = 4D_0^2 (n_1 n_2 / n_1 + n_2)$ & Critical value $= \chi^2_{(2)}$ d.f for one-tail

For two tail :- $D_\alpha = 1.38 \times \sqrt{(n_1+n_2)/(n_1 n_2)}$

Decision :- i) For one-tail :- Reject $H_0$ if $\chi^2_{cal} > \chi^2_{(2)}$ otherwise accept $H_0$

ii) For two-tail :- Reject $H_0$ if $D_0 \geq D_\alpha$ otherwise accept $H_0$

Scanned by CamScanner

) Sol$^n$

To test whether there is two types of motor or not we set up hypothesis as

$H_0$ : There is no significance difference between two types of motors

$H_1$ : There is significant difference difference bet$^n$ two types of motor ~~under~~

Under $H_0$, test - Statistic

$$D = \text{Max of } |F(x) - F(y)|$$

Cal$^n$ table for D

| Life | Fo Motor M f | cf | F(x) | Fo Motor N f | cf | F(y) | $\lvert F(x) - F(y)\rvert$ |
|---|---|---|---|---|---|---|---|
| 2 | 0 | 0 | 0 | 1 | 1 | 1/9 | 1/9 |
| 3 | 3 | 3 | 3/9 | 2 | 3 | 3/9 | 0 |
| 4 | 2 | 5 | 5/9 | 2 | 5 | 5/9 | 0 |
| 5 | 2 | 7 | 7/9 | 2 | 6 | 6/9 | 1/9 |
| 6 | 1 | 8 | 8/9 | 2 | 8 | 8/9 | 0 |
| 7 | 1 | 9 | 1 | 1 | 9 | 1 | 0 |

N =

Now, $D = \text{Max of } |F(x) - F(y)| = 1/9$

Here level of sig $(\alpha) = 0.05$

From table,

$$D_\alpha = 5/9$$

Decision :- $D < D_\alpha$ accept $H_0$ and reject $H_1$ with the conclusion that there is no significant difference between two motors.

Scanned by CamScanner

Pn 5

13 Sol

To test whethere there is significant differene betn junior and senior programmers or not we set up hypothesis as

$H_0$ : There isnt significant diff betn them

Vs $H_1$ : There is " " " "

Calⁿ table for $D_0$

| No. of sig a | For Junior f | c.f | f(x) | For senior f | c.f | f(x) | \|F(x) - F(y)\| |
|---|---|---|---|---|---|---|---|
| 0-2 | 7 | 7 | 7/26 | 5 | 5 | 5/25 | 0.063 |
| 2-4 | 6 | 13 | 13/26 | 4 | 9 | 9/25 | 0.14 |
| 4-6 | 4 | 17 | 17/26 | 6 | 15 | 15/25 | 0.053 |
| 6-8 | 2 | 19 | 19/26 | 3 | 18 | 18/25 | 0.011 |
| 8-10 | 2 | 21 | 21/26 | 4 | 22 | 22/25 | 0.072 |
| 10-12 | 3 | 24 | 24/26 | 2 | 24 | 24/25 | 0.037 |
| 12-14 | 2 | 26 | 1 | 1 | 25 | 1 | 0 |
| | 26 | | | 25 | | | |

Now

$D_0$ = Max of $|F(x) - F(y)|$

$\therefore D_0 = 0.14$

For two tailed test

$\alpha = 5\%$, $n_1 = 26$, $n_2 = 25$

$D_\alpha = 1.36 \times \sqrt{\frac{n_1 + n_2}{n_1 n_2}} = 1.36 \times \sqrt{\frac{26+25}{26 \times 25}}$

$\therefore D_\alpha = 0.38$

Decision

Since $D_0 < D_\alpha$ we accept $H_0$ & reject with the conclusion that .....

Scanned by CamScanner

# MANN WHITNEY U - TEST

It is used to test whether the two samples have been drawn from population with same distribution or not.

* Small sample $(n_1 \le 10, n_2 \le 10)$

Test - statistic

$$U_0 = \text{Min}\{U_1, U_2\}$$

where

$R_1$ = Sum of ranks of 1st sample

$R_2$ = Sum of " " 2nd "

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

Such that $U_1 + U_2 = n_1 n_2$

Then, obtain p-value as

$$p = \text{prob}\{U \le U_0\}$$

Decision :-

* If p-value $\le \alpha$, we reject $H_0$ & accept $H_1$
* If p-value $> \alpha$, we accept $H_0$ & reject $H_0$

* For large samples $(n_1 > 10, n_2 > 10)$

Test - statistic

$$Z = \frac{U_0 - \mu_u}{\sigma_u} \quad \text{where } \mu_u = \frac{n_1 n_2}{2} \ \& \ \sigma_u^2 = n_1 n_2$$

For repeated rank case $\& \ \sigma_u^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$

$$\sigma_u^2 = \frac{n_1 n_2}{n(n-1)\left\{\frac{n^3 - n}{12} - \sum \frac{t_i(t_i^2 - 1)}{12}\right\}}$$

$t_i$ = No. of times a rank repeats

Decision

* If $Z_{cal} \leq Z_{tab}$, we accept $H_0$ & reject $H_1$
* If $Z_{cal} > Z_{tab}$ we reject $H_0$ & accept $H_1$

14) 50%

To test the significance difference between two types of manages we set ~~we~~ up hypothesis

$H_0$ : There isn't any significant difference

Vs $H_1$ : There is significant difference

Here, we have

$n_1 = 5$ , $n_2 = 4$

Cal$^n$ table for $R_1$ & $R_2$

| | | | | | | |
|---|---|---|---|---|---|---|
| Trained rank | 78 | 64 | 75 | 45 | 82 | |
| ~~Untrained rank~~ | 7 | 4 | 6 | 1 | 8 | $R_1 = 26$ |
| Untrained rank | 110 | 70 | 53 | 51 | | |
| | 9 | 5 | 3 | 2 | | $R_2 = 19$ |

Now

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 5\times4 + \frac{5\times6}{2} - 26$$

$\therefore U_1 = 9$

&

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 5\times4 + \frac{4\times5}{2} - 19 = 11$$

Now

$U_0 = \min\{U_1, U_2\} = 9$

From table :- p - value = 0.4524

level of sig ($\alpha$) = 5% = 0.05

Decision :-

Since p-value > $\alpha$ ; we accept $H_0$ and reject $H_1$

Scanned by CamScanner

Date
Page

Scanned by CamScanner

Sol/

To test whether two samples come from identical popⁿ or not

$H_0$ : Two sample are come from identical popⁿ

$H_1$ : Two sample don't come from " "

Calculation table for $R_1$ & $R_2$

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 16.5 | 20 | 17 | 19.8 | 18.5 | 19.2 | 19 | 18.2 | 20.8 | 18.7 | 16.7 | 18.1 | 17.9 | |
| 2 | 23 | 5 | 21 | 13 | 18 | 17 | 10 | 25 | 15 | 4 | 9 | 8 | 170 |
| 18.6 | 17.8 | 18.3 | 16.6 | 20.5 | 16.3 | 19.3 | 18.4 | 19.7 | 18.8 | 19.1 | 17.6 | | |
| 14 | 7 | 11 | 3 | 24 | 1 | 19 | 12 | 20 | 16 | 22 | 6 | | 155 |

Here $n_1 = 13$ , $n_2 = 12$

Now,

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 13 \times 12 + \frac{13 \times 14}{2} - 170 = 77$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 13 \times 12 + \frac{12 \times 13}{2} - 155 = 79$$

Now, $U_0$ = Min of $\{U_1, U_2\}$ = Min of $\{77, 79\}$

$\therefore U_0 = 77$

For large sample

$$\mu_u = \frac{n_1 n_2}{2} = \frac{13 \times 12}{2} = 78$$

$$\sigma_u^2 = \frac{n_1 n_2 (n_1+n_2+1)}{12} = \frac{12 \times 13 \ (12+13+1)}{12} = 338$$

$\therefore \sigma_u = 18.38$

Under $H_0$ - test - statistic

$$Z = \frac{U_0 - \mu_u}{\sigma_u} = \frac{77 - 78}{18.38} = -0.054$$

$\therefore Z_{cal} = 0.054$

$\therefore$ Since $Z_{cal} < Z_{tab}$, we accept $H_0$ & reject $H_1$ with the ....

Scanned by CamScanner

# $\chi^2$ - Test (chi-square Test)

Applications :
Test of randomness or uniformity
Goodness of fit Test
Independence or association of attributes

Test of randomness or uniformity
Test - statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$ where $O$ = observed freq
$E$ = Expected $= \frac{\sum O}{n}$

$$d.f = n-1$$

Goodness of fit Test
Test - statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where $E = N \times$ prob
For Binomial distn
$$E = N \times {}^nC_r\, p^r q^{n-r}$$
For Poisson distn
$$E = N \times e^{-\lambda} \frac{\lambda^r}{r!}$$

Here $d.f = n - 1 - k_1 - k_2$
$k_1$ = No of parameters
$k_2$ = No of expected freq less than 5

Scanned by CamScanner

Test of Independence or association of attributes

Test - statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Where $E = \frac{RT \times CT}{GT}$

RT = Row total
CT = Column total
GT = Grand total

$$d.f = (r-1)(c-1)$$

r = no. of column , c = no. of column

Note :- For 2x2 contigency table if any cell frequency is less than 5, we use Yate's correction to compute $\chi^2$ - statistic

$$\chi^2 = \frac{N[|ad-bc| - N/2]^2}{(a+b)(a+c)(b+d)(c+d)}$$

Scanned by CamScanner

i) Soln

To test whether the dice was biased or not, we setput hypothesis as

$H_0$ : Dice was unbiased

Vs $H_1$ : Dice was biased

Under $H_0$, test statistic :

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

Now, Calculation table for $\chi^2$ -statistic

| O | E | $(O-E)^2/E$ |
|---|---|---|
| 8 | 10 | 0.4 |
| 9 | 10 | 0.1 |
| 13 | 10 | 0.9 |
| 7 | 10 | 0.9 |
| 15 | 10 | 2.5 |
| 8 | 10 | 0.4 |
| 60 | 60 | 5.2 |

Here,

$\chi^2_{cal} = 5.2$

level of sig $(\alpha) = 5\%$ d.f $= n-1 = 6-1 = 5$

Now,

$\chi^2_{tab} = 11.070$

Decision :-

Since $\chi^2_{cal} < \chi^2_{tab}$, we accept $H_0$ & reject $H_2$ with the conclusion that dice was unbiased.

Scanned by CamScanner

22.) To test whether the genetic theory supports observed data or not, we setup hypothesis as;

$H_0$ : observed data supports genetic theory

Vs $H_1$ : " " does not support genetic theory

Under $H_0$ test statistics

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Here, $E(A) = 1/4 \times 300 = 75$

$E(AB) = 2/4 \times 300 = 150$

$E(B) = 1/4 \times 300 = 75$

Calculation table for $\chi^2$-statistics

| For | O | E | $\left(\frac{O-E}{E}\right)^2$ |
|---|---|---|---|
| A | 90 | 75 | 3 |
| AB | 135 | 150 | 1.5 |
| B | 75 | 75 | 0 |
| | 300 | 300 | 4.5 |

Here, $\chi^2 = 4.5$

$\alpha = 5\%$, $d.f = n-1 = 3-1 = 2$

Now,

$\chi^2_{tab} = 5.999$

Since $\chi^2_{cal} < \chi^2_{tab}$ we accept $H_0$ & reject $H_1$ with the conc.

23)

| $x = r$ |
|---|
| 0 |
| 1 |
| 2 |
| 3 |
| 4 |
| 5 |

Scanned by CamScanner

To test whether the data ~~set is~~ ~~phase~~ fits poisson distribution or not we set up hypothesis as;

$H_0$ : data fits poisson distⁿ

Vs $H_1$ : data doesnt fit poisson distⁿ

Under $H_0$ test-statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where E = Expected freq = $N \times e^{-\lambda} \frac{\lambda^r}{r!}$

Since, $\lambda$ is not given, we estimate it through given data. Here,

$$\lambda = \bar{x} = \frac{\sum fx}{N}$$

or, $$\lambda = \frac{0 \times 221 + 1 \times 167 + 2 \times 70 + 3 \times 30 + 4 \times 7 + 5 \times 5}{221 + 167 + 70 + 30 + 7 + 5}$$

$\therefore \lambda = 0.9$ $\quad N = 500$

Now, $E = 500 \times e^{-0.9} \times \frac{0.9^r}{r!}$

Now, calⁿ table for $\chi^2$- statistic

| O | E | $\frac{(O-E)^2}{E}$ |
|---|---|---|
| 221 | 203.285 | 1.5444 |
| 167 | 182.956 | 1.392 |
| 70 | 82.33 | 1.847 |
| 30 | 24.69 | 1.138 |
| 7 } 12 | 5.557 } 6.557 | 4.518 |
| 5 | | |
| $\Sigma O = 500$ | $\Sigma E = 500$ | $\Sigma \frac{(O-E)^2}{E} = 10.439$ |

Here,

$\chi^2_{cal} = 10.439$ , $\chi^2_{tab} = 7.815$

Since, $\chi^2_{cal} > \chi^2_{tab}$ we reject $H_0$ & accept $H_1$

Scanned by CamScanner

24) To test whether the data fits binomial distribution or not, we setup hypothesis as

$H_0$ : Data fits binomial distribution

Vs $H_1$ : Data doesnot fit binomial distribution.

Under $H_0$ test statistics;

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where $E = N \times {}^nC_r \, p^r q^{n-r}$

Here

$N = 50$, $n = 5 \Rightarrow$ no. of trials

$p = 0.30$, $q = 1-p = 1-0.30 = 0.70$

Now, $E = 50 \times {}^5C_r \, p^5 q^{5-r}$

$= 50 \times {}^5C_r \, (0.3)^5 \, (0.7)^{5-r}$

Calcn table for $\chi^2$ statistic

| $X = r$ | $O$ | $E$ | $(O-E)^2/E$ |
|---|---|---|---|
| 0 | 4 | $8.4035 = 8$ | 2.3075 |
| 1 | 13 | $18.0075 = 18$ | 1.3929 |
| 2 | 17 | $15.435 = 15$ | 0.1587 |
| 3 | 12 | 6.615 | |
| 4 | 3 } 16 | 1.4175 } 8.154 | } 7.5496 |
| 5 | 1 | 0.1215 | |
| | 50 | 50 | 11.4032 |

$\chi^2_{cal} = 11.4082$

$d.f = n-1-k_1-k_2 = 6-1-0-2 = 3$

$\chi^2_{tab} = 7.815$

Decision

Since $\chi^2_{cal} > \chi^2_{tab}$ we reject $H_0$ & accept $H_1$ with the conclusion that,

Scanned by CamScanner

To test whether the no. of children is independent of level of education attended by father or not, we set up hypothesis as :

$H_0$ : No of children and level of education attended by father is independent

$H_1$ :- dependent

Under $H_0$, test - statistics

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad \text{Here, } E = \frac{RT \times CT}{GT}$$

Calculation table for expected freq

| Edu | No of children 0-1 | 2-3 | Over 3 | Total |
|---|---|---|---|---|
| Ele | 14 | 37 | 32 | 83 |
| Sec | 19 | 42 | 17 | 78 |
| College | 12 | 17 | 10 | 29 |
| Total | 45 | 96 | 55 | 190 |

$E(14) = \frac{83 \times 45}{190} = 19.69$

$E(37) = \frac{83 \times 96}{190} = 41.94$

$E(32) = \frac{83 \times 59}{190} = 25.77$

$E(19) = 18.47$, $E(42) = 39.41$

$E(17) = 24.22$, $E(12) = 6.86$

$E(17) = 14.65$, $E(10) = 9.01$

| O | E | $(O-E)^2/E$ |
|---|---|---|
| 14 | 19.69 | 1.6421 |
| 37 | 41.94 | 0.582 |
| 32 | 25.77 | 1.506 |
| 19 | 18.47 | 0.015 |
| 42 | 39.41 | 0.170 |
| 17 | 24.22 | 2.15 |
| 12 | 6.86 | 3.85 |
| 17 | 14.65 | 0.377 |
| 10 | 9.01 | 0.108 |
| $\Sigma O = 190$ | $\Sigma E = 190$ | 10.4 |

$= 10.4$, d.f $= (r-1)(c-1)$

$\chi^2_{tab} = 9.488 \quad = (3-1)(3-1) = 4$

Ans :- Since $\chi^2_{cal} > \chi^2_{tab}$, reject $H_0$ and accept $H_1$ with the conclusion that no. of children & education level of father are dependent.

Scanned by CamScanner

31) To test whether the flower color is independent of flatness of leaf or not we setup hypothesis as

$H_0$ : Independent Vs $H_1$ : Dependent

Under $H_0$ test - statistics

Since for 2x2 contingency table cell freq is less than 5 then

$$\chi^2 = \frac{N\left[|ad-bc| - N/2\right]^2}{(a+b)(a+c)(b+d)(c+d)}$$

$N = a+b+c+d = 3+9+22+11 = 45$

Here,

$a+b = 3+22 = 25$ $\quad b+d = 22+11 = 33$

$a+c = 3+9 = 12$ $\quad c+d = 9+11 = 20$

$$\chi^2 = \frac{45\left[|3\times11 - 22\times9| - 45/2\right]^2}{25\times12\times33\times20}$$

$\chi^2 = 7.97$.

$d.f = (r-1)(c-1) = (2-1)(2-1) = 1$

$\chi^2_{tab} = 3.841$

Decision :

Since $\chi^2_{cal} > \chi^2_{tab}$ ; we reject $H_0$ & accept $H_1$ with the conclusion that flower color is dependent of flatness of leaf.

Scanned by CamScanner

Wilcoxon - Matched —
— Pair signed rank test

It is non-parametric test used to compare two populations for paired observations. It is based upon magnitude & sig.

For small sample ($n \leq 25$)
Test - statistic

$$T = \text{Min}\{ s(+), s(-) \}$$

where $s(+)$ = Sum of all +ve ranks of difference
$s(-)$ = " " " -ve " " "

Critical Value
Table a critical value is obtained from respective table for corrected sample size (n)
Corrected sample is obtained by omitting for zero difference

Decision: Reject $H_0$ if $T \leq T_{\alpha, n}$ otherwise accept $H_0$.

For large sample ($n > 25$)

$$Z = \frac{T - \mu_T}{\sigma_T}$$

where $\mu_T = \frac{n(n+1)}{4}$

$$\sigma_T^2 = \frac{n(n+1)(2n+1)}{24}$$

Scanned by CamScanner

7.) 50%

To test whether given up smoking increase one weight we set of hypothesis as

$H_0: M_{d_1} = M_{d_2}$ i.e. there is no effect of smoking of weight

Vs $H_1: M_{d_1} < M_{d_2}$ (one tailed test) i.e. giving up smoking increase one weight

Calculation table for test statistic

| Before | After | d | Rank |
|---|---|---|---|
| 66 | 71 | 5 | 5 |
| 80 | 82 | 2 | 2.5 |
| 69 | 68 | -1 | -1 |
| 52 | 56 | 4 | 4 |
| 75 | 73 | -2 | -2.5 |

Here, $s(+) = 11.5$, $n = 5$

$s(-) = 3.5$

Now, Test statistic

$T = \min\{s(+), s(-)\}$

$T = \min\{11.5, 3.5\}$

$= 3.5$

Here level of sig $(\alpha) = 5\%$, $n = 5$

Table or critical value

$T_{\alpha,n} = 1$

Decision :-

Since, $T_{cal} > T_{\alpha,n}$, we accept $H_0$ & reject $H_1$ with the conclusion that smoking has no effect on weight.

10)

Vs

Date
Page

Q07

To test whether the archeologist $x$ is better than $y$ or not. We set up hypothesis as

$H_0$ : $M_{d_1} = M_{d_2}$ i.e there is no sig diff betn two archeologist

$H_1$ : $M_{d_1} > M_{d_2}$ i.e ~~there is~~ archeologist $x$ is greater than $y$.

Calⁿ table for Test-statistic

| X | Y | d | Rank |
|---|---|---|---|
| 1 | 0 | 1 | 7 |
| 0 | 0 | 0 | — |
| 2 | 1 | -1 | 7 |
| 3 | 0 | 3 | ~~22~~ 28 |
| 1 | 2 | -1 | -7 |
| 0 | 0 | 0 | — |
| 2 | 0 | 2 | 17 |
| 2 | 1 | 1 | 7 |
| 3 | 1 | 2 | 17 |
| 0 | 2 | -2 | -17 |
| 1 | 0 | 1 | 7 |
| 1 | 1 | 0 | — |
| 4 | 2 | 2 | 17 |
| 1 | 1 | 0 | — |
| 2 | 1 | 1 | 7 |
| 1 | 0 | 1 | 7 |
| 3 | 2 | 1 | 7 |
| 5 | 2 | 3 | 22 |
| 2 | 6 | -4 | 25 |
| 1 | 0 | 1 | 7 |
| 3 | 2 | 1 | 7 |
| 2 | 3 | -1 | -7 |
| 4 | 0 | 4 | |
| 1 | 2 | -1 | -7 |
| 3 | 1 | 2 | 17 |
| 2 | 0 | 2 | 17 |
| 0 | 1 | -1 | -7 |
| 2 | 0 | 2 | |
| 4 | 1 | 3 | 22 |
| 1 | 5 | -4 | 25 |

Now,

$S(+) = 256$

$S(-) = 95$

Here, $T = \text{Min}\{S(+), S(-)\}$

$= \text{Min}\{256, 95\}$

$\therefore T = 95$

For $\mu_T$

$$\mu_T = \frac{n(n+1)}{4} = \frac{26 \times 27}{4} = 175.5$$

$$\sigma_T^2 = \frac{n(n+1)(2n+1)}{24} = \frac{26 \times 27 \times 53}{24}$$

$$\sigma_T^2 = 1550.25$$

Now, Test-statistic

$$Z = \frac{T - \mu_T}{\sigma_T} = \frac{95 - 175.5}{\sqrt{1550.25}}$$

$|Z_{cal}| = 2.07$

Here $\alpha = 5\%$.

For one tailed test

$Z_{tab} = 1.645$

Decision

$Z_{cal} > Z_{tab}$, reject $H_0$

accept $H_1$

Scanned by CamScanner

# COCHRAN Q - Test

It is used for test of significant difference in frequencies of 3 or more related samples.

Test - statistic

$$Q = \frac{(k-1)\{k \sum R_i^2 - (\sum R_i)^2\}}{k \sum C_j - \sum C_j^2}$$

Critical or table value

Table value is obtained from $\chi^2$ - table at $\alpha$% level of sig & $(k-1)$ d.f

Decision

Reject $H_0$ if $Q > \chi^2_{\alpha(k-1)}$ otherwise accept $H_0$.

Scanned by CamScanner

) Sol/

To test whether the three diets are equally effective or not, we set up hypothesis as

$H_0$ : Three diets are equally effective

Vs $H_1$ : As least one pair of diets are not equally effective

Under $H_0$ test - statistic,

$$Q = \frac{(k-1)\{k\Sigma R_i^2 - (\Sigma R_i)^2\}}{k\Sigma C_j - \Sigma C_j^2}$$

Calⁿ table for Q

| Diet | Buffaloes 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | $R_i$ | $R_i^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| P | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 5 | 25 |
| Q | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 6 | 36 |
| R | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 5 | 25 |
| $C_j$ | 3 | 1 | 2 | 0 | 3 | 2 | 1 | 2 | 2 | $\Sigma R_i = 16$ | $\Sigma R_i^2 = 86$ |
| $C_j^2$ | 9 | 1 | 4 | 0 | 9 | 4 | 1 | 4 | 4 | | |

$\Sigma C_j = 16$ $\Sigma C_j^2 = 36$ , $k = 3$

Now

$$Q = \frac{(3-1)\{3 \times 86 - 16^2\}}{3 \times 16 - 36}$$

$\therefore Q = 0.33$

Here, level of sig $(\alpha) = 5\%$.

d.f $= k-1 = 3-1 = 2$

$\chi^2_{tab} = 5.991$

Decision :-

Since $Q_{cal} < \chi^2_{tab}$, we accept $H_0$ & reject $H_1$, with the conclusion that three diets are equally effective.

Scanned by CamScanner

# KRUSKAL WALLIS H-Test

It is also known as Kruskal Wallis one way ANOVA
It is used to test the significant difference between locations of three or more independent populations.

Test - statistic

$$H = \frac{12}{n(n+1)} \sum \frac{R_i^2}{n_i} - 3(n+1)$$

(For simple case)

* For tied rank case

$$H = \frac{\frac{12}{n(n+1)} \sum \frac{R_i^2}{n_i} - 3(n+1)}{1 - \sum \frac{t_i(t_i^2-1)}{n(n^2-1)}}$$

where $t_i$ = no of repeatation of $i^{th}$ rank

Critical or table value

For $n_i \leq 5$ & $k=3$ p-value is obtained from k-w table

For $n_i > 5$ & $k > 3$, critical value is obtained from $\chi^2$ table at $(k-1)$ d.f.

Scanned by CamScanner

13 - 17)

50%

) To test whether the operating times for all three calculators are equal, we set up hypothesis as

$H_0$ :

Vs $H_1$ : Operating time for at least one pair of calculator is different

Calⁿ Table

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Cal A | 4.9 | 6.1 | 4.3 | 4.6 | 5.3 | | | $R_i$ | $R_i^2/n_i$ |
| Rank | 4 | 12 | 1 | 2 | 6 | | | 25 | 125 |
| Cal B | 5.5 | 5.4 | 6.2 | 5.8 | 5.5 | 5.2 | 4.8 | | |
| Rank | 8.5 | 7 | 13 | 11 | 8.5 | 5 | 3 | 56 | 448 |
| Cal C | 6.4 | 6.8 | 5.6 | 6.5 | 6.3 | 6.6 | | | |
| Rank | 15 | 18 | 10 | 16 | 14 | 17 | | 90 | 1350 |
| | | | | | | | | $\sum \frac{R_i^2}{n_i}$ | 1923 |

Here, $n = n_1 + n_2 + n_3 = 5 + 7 + 6 = 18$

$t_1 = 2$

$$\frac{\sum t_i (t_i^2 - 1)}{n(n^2-1)} = \frac{2(2^2-1)}{18(18^2-1)} = \frac{1}{3 \times 323}$$

Now, test-statistic,

$$H = \frac{\frac{12}{n(n+1)} \sum \frac{R_i^2}{n_i} - 3(n+1)}{1 - \frac{\sum t_i (t_i^2 - 1)}{n(n^2-1)}}$$

$$= \frac{\frac{18}{18 \times 19} \times 1923 - 3 \times 19}{1 - \frac{1}{3 \times 323}}$$

$\therefore H = 10.4$

Here, level of sig $(\alpha) = 5\%$. d.f $= k - 1 = 3 - 1 = 2$

$\chi^2_{tab} = 5.99$

Conclusion :- Since $H_{cal} > \chi^2_{tab}$ we reject $H_0$ & accept $H_1$ with ...

Scanned by CamScanner

# Friedman - F - test

i) It is used to test the significant difference in yield due to three or more than three characters

ii) It is also known as two way ANOVA test

Test statistics

a) For non-repeated case

$$F = \frac{12}{nk(k+1)} \sum_{i=1}^{k} R_i^2 - 3n(k+1)$$

b) For tied case

$$F = \frac{\frac{12}{nk(k+1)} \sum_{i=1}^{k} R_i^2 - 3n(k+1)}{1 - \frac{\sum (t_i^3 - t_i)}{n(k^3 - k)}}$$

*) Critical value

i) For $2 \le n \le 9$ and $k = 3$ and $2 \le n \le 4$ & $k = 4$ p-value is obtained from table

ii) For $n \ge 5$ and $k > 3$
Table value of $\chi^2_{cal}(k-1)$ is obtained from $\chi^2_{table}$.

Scanned by CamScanner

18, 19, 20)

To test whether there is significant difference in variation between matched groups or not, we setup hypothesis as;

$H_0$ : There is no significant difference in variation betn matched group

Vs $H_1$ : There is significant difference in variation betn at least one pair of group

Calculation table for Friedman F-test

| Group | I | Rank | II | Rank | III | Rank | IV | Rank | V | Rank |
|---|---|---|---|---|---|---|---|---|---|---|
| A | 9 | 3 | 5 | 3 | 2 | 1 | 5 | 3 | 6 | 2 |
| B | 6 | 2 | 4 | 2 | 3 | 2.5 | 4 | 2 | 6 | 2 |
| C | 5 | 1 | 1 | 1 | 3 | 2.5 | 3 | 1 | 6 | 2 |

| $N_i$ | Rank | $R_i$ | $R_i^2$ |
|---|---|---|---|
| 7 | 3 | 15 | 225 |
| 5 | 1.5 | 12 | 144 |
| 5 | 1.5 | 9 | 81 |
| | | | 450 |

Here, $k = 3$, $n = 6$, $t_1 = 2$, $t_2 = 3$, $t_3 = 2$

Now

$$\sum \frac{t_i(t_i^2-1)}{nk(k^2-1)}$$

$$= \frac{2(2^2-1) + 3(3^2-1) + 2(2^2-1)}{6 \times 3(3^2-1)}$$

$$= (6 + 24 + 6) \div 144$$

$$= \frac{1}{4}$$

Scanned by CamScanner

$$F = \frac{\frac{12}{nk(k+1)} \{R_i^2 - 3n(k+1)}{1 - \sum \frac{t_i(t_i^2-1)}{nk(k^2-1)}}$$

$$= \frac{\frac{12}{6\times3\times4} \times 450 - 3\times6\times4}{1 - 1/4}$$

$$= 4$$

Here, $\alpha = 5\% = 0.05$

From table

p- value $= 0.184$

Decision :-

Since p-value $> \alpha$, we accept $H_0$ and reject $H_1$ with the conclusion that there is no significant difference in variation of matched group.

19) Here, To test whether the birth rate is constant over four seasons or not we setup hypothesis as

$H_0$ :- The birth rate is constant over all seasons

Vs $H_1$ : The birth rate is constant at least on pair of season

Calculation table for F-test

| Hospitals | Winter | Rank | Spring | Rank | Summer | Rank | Fall | Rank | $R_i$ | $R_i^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| A | 92 | 4 | 112 | 4 | 94 | 4 | 77 | 4 | 16 | 256 |
| B | 9 | 1 | 11 | 1 | 10 | 1 | 15 | 1 | 4 | 16 |
| C | 58 | 3 | 71 | 3 | 57 | 3 | 62 | 3 | 12 | 144 |
| D | 19 | 2 | 26 | 2 | 19 | 2 | 18 | 2 | 8 | 64 |
| | | | | | | | | | | 480 |

Scanned by CamScanner

Here, $n = 4$, $k = 4$

Test - statistics

$$F = \frac{12}{nk(k+1)} \sum_{i=1}^{k} R_i^2 - 3n(k+1)$$

$$= \frac{12}{4 \times 4 \times 5} \times 480 - 3 \times 4 \times 5$$

$$= 72 - 60$$

$$= 12$$

level of sig $(\alpha) = 5\% = 0.05$

p-value $= 0.158$

Scanned by CamScanner

# 4.) Multiple Correlation and Regression

## 2.) Partial Correlation

It is the correlation coefficient between two variables keeping the effect of all other variables as constant.

Let us consider 3 variables $x_1, x_2$ & $x_3$ then

$r_{12.3}$ = Partial correlation coeff between $x_1$ & $x_2$ keeping effect of $x_3$ as constant

$r_{13.2}$ = Partial correlation coeff bet $x_1$ & $x_3$ keeping effect of $x_2$ as constant

$r_{23.1}$ = " " " $x_2$ & $x_3$ " " of $x_1$ " const

*) Computational Formula

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

$$r_{23.1} = \frac{r_{23} - r_{12} \cdot r_{13}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{13}^2}}$$

$$r_{13.2} = \frac{r_{13} - r_{12} \cdot r_{23}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{23}^2}}$$

where

$r_{12}$ = simple correl$^n$ bet $x_1$ & $x_2$

$r_{13}$ = " " bet $x_2$ & $x_3$

$r_{23}$ = " " " $x_2$ & $x_3$

Scanned by CamScanner

Date
Page

) Multiple correlation coefficient

It is the correlation coefficient between one dependent and two or more than two independent variables. It is also known as correlation betn one dependent and combined effect of all other independent variables.

Let us consider $x_1$, $x_2$ and $x_3$ then,

$R_{1.23}$ = Multiple correlⁿ coeff among $x_1, x_2$ & $x_3$ taking $x_1$ as dependent

$R_{2.13}$ = " " " " " taking $x_2$ as independent

$R_{3.12}$ = " " " $x_3$ as dependent

) Computational formula

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2 r_{12} r_{13} r_{23}}{1 - r_{23}^2}}$$

$$R_{2.13} = \sqrt{\frac{r_{23}^2 + r_{12}^2 - 2 r_{12} r_{13} r_{23}}{1 - r_{13}^2}}$$

$$R_{3.12} = \sqrt{\frac{r_{13}^2 + r_{23}^2 - 2 r_{12} r_{13} r_{23}}{1 - r_{12}^2}}$$

Scanned by CamScanner

3.) Multiple linear Regression Equation

It is the mathematical relationship between one dependent & two or more than two independent variables. It is used to estimate the value of dependent variable with know values of independent variables.

Let us consider 3 variables $y, x_1$ & $x_2$ then multiple regression equation of $y$ on $x_1$ and $x_2$ is written as;

$$y = a + b_1 x_1 + b_2 x_2 \quad \text{—— (i)}$$

where,

$y$ = dependent variable

$x_1, x_2$ = independent variables

$a$ = y-intercept or value of y independent of change in $x_1$ and $x_2$

$b_1$ = Rate of change in y w.r.t $x_1$

$b_2$ = Rate of " " y " $x_2$

Normal eqn to solve $a, b_1$ & $b_2$ are

$$\Sigma y = na + b_1 \Sigma x_1 + b_2 \Sigma x_2 \quad \text{—— (2)}$$

$$\Sigma y \cdot x_1 = a \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 \quad \text{—— (3)}$$

$$\Sigma y \cdot x_2 = a \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2 \quad \text{—— (4)}$$

Now

We use Cramer's rule to solve normal equation

Scanned by CamScanner

Standard Error of Estimate

It is the standard deviation of estimated value from regression equation

→ Computational formulae

$$S.E. = \sqrt{\frac{\Sigma y^2 - a\Sigma y - b_1 \Sigma x_1 y - b \Sigma x_2 y}{n-k-1}}$$

where, $n$ = no. of data

$k$ = no. of independent variables

It is always positive

Coeff of Multiple Determination

It is the square of correlation coeff & represents percentage of total variation of dependent variable explained by independent variables used.

For example if $R_{1.23} = 0.9$ coeff of multiple determination ~~variation~~ $(R_{1.23})^2 = 0.9^2 = 0.81 = 81\%$ which means 81% of total variation of dependent variation is explained by independent variables

Its value lies between 0 & 1

→ Computational formula:

$$R^2_{y.12} = \frac{a\Sigma y + b_1 \Sigma x_1 y + b_2 \Sigma x_2 y - n\bar{y}^2}{\Sigma y^2 - n\bar{y}^2}$$

Scanned by CamScanner

Note i) The value of partial correlation coeff lies betn -1 and +1

ii) Value of multiple correlation coeff lies betn 0 & 1

iii) For the test of significance of single regression coeff, we use t-test

iv) For the overall test of regression model, we use F-test.

Numerical

Q.9) Here,

$r_{12} = 0.5$, $r_{23} = 0.1$ and $r_{13} = 0.4$

i) $$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} = \frac{0.5 - 0.4 \times 0.1}{\sqrt{1-0.4^2}\sqrt{1-0.1^2}}$$

$$= \frac{0.5 - 0.04}{\sqrt{1-0.16}\sqrt{1-0.01}}$$

$$= 0.5$$

11) Here $r_{23} = 0.8$, $r_{31} = 0.5$, $r_{12} = 0.6$

Now

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

$$= \frac{0.6 - (0.5 \times 0.8)}{\sqrt{1-0.25}\sqrt{1-0.64}}$$

$$= 1.92$$

Since, it lies beyond the range of partial correlation coeff, So, the data is inconsistent.

Scanned by CamScanner

Date
Page

) Sol$^n$

$\Sigma x_1 = 10$, $\Sigma x_2 = 20$, $\Sigma x_3 = 30$, $\Sigma x_1^2 = 20$, $\Sigma x_2^2 = 68$
$\Sigma x_3^2 = 170$, $\Sigma x_1 x_2 = 10$, $\Sigma x_1 x_3 = 15$, $\Sigma x_2 x_3 = 64$
$n = 10$

Here, we need simple correl$^n$ coeff

$$r_{12} = \frac{n \Sigma x_1 x_2 - \Sigma x_1 \Sigma x_2}{\sqrt{n \Sigma x_1^2 - (\Sigma x_1)^2}\sqrt{n \Sigma x_2^2 - (\Sigma x_2)^2}}$$

$$= \frac{10 \times 10 - 10 \times 20}{\sqrt{10 \times 20 - (10)^2}\sqrt{10 \times 68 - (20)^2}}$$

$$= -0.5977$$

$$r_{13} = \frac{n \Sigma x_1 x_3 - \Sigma x_1 \Sigma x_3}{\sqrt{n \Sigma x_1^2 - (\Sigma x_1)^2}\sqrt{n \Sigma x_3^2 - (\Sigma x_3)^2}}$$

$$= \frac{10 \times 15 - 10 \times 30}{\sqrt{10 \times 20 - 10^2}\sqrt{10 \times 170 - 30^2}} = -0.53$$

$$r_{23} = \frac{n \Sigma x_2 x_3 - \Sigma x_2 \Sigma x_3}{\sqrt{n \Sigma x_2^2 - (\Sigma x_2)^2}\sqrt{n \Sigma x_3^2 - (\Sigma x_3)^2}}$$

$$= \frac{10 \times 64 - 20 \times 30}{\sqrt{10 \times 68 - 20^2}\sqrt{10 \times 170 - 30^2}}$$

$$= 0.08$$

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}} = \frac{-0.59 - (-0.53) \times 0.08}{\sqrt{1 - (-0.53)^2}\sqrt{1 - (-0.08)^2}}$$

$$= -0.65$$

Scanned by CamScanner

Sol^n

Here, we have

$n = 20,\ b_1 = 4,\ b_2 = 3,\ S_{b_1} = 1.2,\ S_{b_2} = 0.8$

For 1st explanatory variable

To test whether the 1st explanatory variable is significant or not, we setup hypothesis as;

$H_0 : B_1 = 0$ ie 1st explanatory variable is insignificant

Vs $H_1 : B_1 \neq 0$ ie " " " " significant

Under $H_0$ test statistics

$$t = \frac{b_1}{S_{b_1}} = \frac{4}{1.2} = 3.33$$

level of sig $(\alpha) = 5\% = 0.05$

$d.f = n-k-1 = 20-2-1 = 17$

$t_{table} = 2.11$

Decision :-

Since, $t_{cal} > t_{tab}$, we reject $H_0$ & accept $H_1$ with the conclusion that 1st explanatory variable is significant.

For 2nd

To test whether 2nd explanatory variable is significant or not we setup hypothesis as

$H_0 : B_2 = 0$ ie 2nd explanatory variable is insignificant

Vs $H_1 : B_2 \neq 0$ ie " " " " significant

Under $H_0$, test statistics

$$t = \frac{b_2}{S_{b_2}} = \frac{3}{0.8} = 3.75$$

level of sig $(\alpha) = 5\% = 0.05$

$d.f = n-k-1 = 20-2-1 = 17$

Scanned by CamScanner

$t_{tab} = 2.11$

Decision :-

Since, $t_{cal} > t_{tab}$ we reject $H_0$ and accept $H_1$ with the conclusion that 2nd explanatory variable is significant.

) Sol

$n = 10$

Total sum of square (TSS) $= \Sigma(y-\bar{y})^2 = 297.8$

Sum of square of errors (SSE) $= \Sigma(y-\hat{y})^2 = 23.5$

$k = 2$

To test whether the regression model is significant or not, we setup hypothis as ;

$H_0 : B_i = 0$ ie regⁿ model is insigⁿ

$H_1 : B_i \neq 0$ ie " " " sigⁿ

Under $H_0$ test statistics.

$$F = \frac{MSSR}{MSSE} \quad \text{——— (i)}$$

Here,

$$MSSR = \frac{SSR}{k} = \frac{SSE - TSS}{k} = \frac{374.1}{2} = 187.05$$

$$MSSE = \frac{SSE}{n-k-1} = \frac{23.5}{7} = 3.357$$

Now, $F_{cal} = \frac{187.05}{3.357} = 55.2$

Now

level of sig $(\alpha) = 5\% = 0.05$ d.f $= (k, n-k-1)$

$F_{tab} = 4.74$ $= (2, 7)$

Decision :- Since $F_{cal} > F_{tab}$ we reject $H_0$ & accept $H_1$ with the conclusion that regression model is signⁿ

i) $TSS = SSR + SSE$

ii) $F = \frac{MSSR}{MSSE}$

iii) $MSSR = \frac{SSR}{d.f} = \frac{SSR}{k}$

iv) $MSSE = \frac{SSE}{d.f} = \frac{SSE}{n-k-1}$

TSS = Total sum of square

SSR = Sum of square due to regression

SSE = " " error

MSSR = Mean sum of square due to reg

MSSE = Mean sum of square due to error

Scanned by CamScanner

Date
Page

a) Regn eqn of $x_3$ on $x_1$ & $x_2$

$$x_3 = a + b_1 x_1 + b_2 x_2 \quad \text{—— (i)}$$

Normal eqns to solve for $a$, $b_1$ & $b_2$ are

$$\Sigma x_3 = na + b_1 \Sigma x_1 + b_2 \Sigma x_2 \quad \text{—— (2)}$$

$$\Sigma x_1 x_3 = a \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 \quad \text{—— (3)}$$

$$\Sigma x_2 x_3 = a \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2 \quad \text{—— (4)}$$

Now, substituting the values of various sum

$$10a + 13b_1 + 11b_2 = 51$$

$$13a + 63b_1 - 240b_2 = 77$$

$$11a - 240b_1 + 95b_2 = 136$$

Here, det of coefficients

$$\Delta = \begin{vmatrix} 10 & 13 & 11 \\ 13 & 63 & -240 \\ 11 & -240 & 95 \end{vmatrix} =$$

For a

$$\Delta_1 = \begin{vmatrix} 51 & 13 & 11 \\ 77 & 63 & -240 \\ 136 & -240 & 95 \end{vmatrix} =$$

For $b_1$

$$\Delta_2 = \begin{vmatrix} 10 & 51 & 11 \\ 13 & 77 & -240 \\ 11 & 136 & 95 \end{vmatrix} =$$

For $b_2$

$$\Delta_3 = \begin{vmatrix} 10 & 13 & 51 \\ 13 & 63 & 77 \\ 11 & -240 & 136 \end{vmatrix} =$$

Now, $a = \frac{\Delta_1}{\Delta} = 5.6688$

$b_1 = \frac{\Delta_2}{\Delta} = -0.3485$

$b_2 = \frac{\Delta_3}{\Delta} = -0.1053$

Scanned by CamScanner

Now, fitted regⁿ eqⁿ

$$x_3 = 5.6688 - 0.3485\, x_1 - 0.1053\, x_2$$

ii) When $x_1 = 1$ & $x_2 = 4$, predicted value of $x_3$

$$x_3 = 5.6688 - 0.3485 \times 1 - 0.1053 \times 4$$

$\therefore x_3 =$

iii)

iv) For S.E

$$S.E = \sqrt{\frac{\Sigma x_3^2 - a\Sigma x_3 - b_1 \Sigma x_3 x_1 - b_2 \Sigma x_3 x_2}{n-k-1}}$$

$$= \sqrt{\frac{450 - 5.6688 \times 51 + 0.3485 \times 77 + 0.1053 \times 136}{10-2-1}}$$

$\therefore S.E =$

v) Coeff of multiple determination

$$R^2 = \frac{a\Sigma x_3 + b_1 \Sigma x_1 x_3 + b_2 \Sigma x_2 x_3 - n\bar{x}_3^2}{\Sigma x_3^2 - n\bar{x}_2^2}$$

Here, $\bar{x}_3 = \frac{\Sigma x_3}{n} = \frac{51}{10} = 5.1$

$$= \frac{5.6688 \times 51 + 0.3485 \times 77 - 0.1053 - 10 \times 5.1^2}{450 - 10 \times 5.1^2}$$

$=$

Scanned by CamScanner

~~From~~ Calc$^n$ table for various sum

| $\Sigma y =$ | $\Sigma x_1 =$ | $\Sigma x_2 =$ | $\Sigma y x_1 =$ | $\Sigma y x_2 =$ | $\Sigma x_1 x_2 =$ | $\Sigma x_1^2 =$ | $\Sigma x_2^2$ | $\Sigma y^2 =$ |
|---|---|---|---|---|---|---|---|---|
| 51 | 13 | 11 | 77 | 136 | -24 | 63 | 95 | 487 |

Here, $n = 6$

3) Req$^n$ eq$^n$ of $y$ on $x_1$ & $x_2$ is written as

$$y = a + b_1 x_1 + b_2 x_2 \quad \text{—— (1)}$$

Normal eq$^n$s to solve for $a$, $b_1$ & $b_2$ are

$$\Sigma y = na + b_1 \Sigma x_1 + b_2 \Sigma x_2$$

$$\Sigma y x_1 = a \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2$$

$$\Sigma y x_2 = a \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2$$

Now substituting the values of various sum

$$6a + 13 b_1 + 11 b_2 = 51 \quad \text{—— (2)}$$

$$13a + 63 b_1 - 24 b_2 = 77 \quad \text{—— (3)}$$

$$11a + 63 b_1 + 99 b_2 = 136 \quad \text{—— (4)}$$

Here, coefficient of

| $a$ | $b_1$ | $b_2$ | constant |
|---|---|---|---|
| 6 | 13 | 11 | 51 |
| 13 | 63 | -24 | 77 |
| 11 | -24 | 99 | 136 |

Scanned by CamScanner

Date
Page

Scanned by CamScanner

# DESIGN OF EXPERIMENT

Terminology

Principle of design of experiment
→ Replication
→ Randomization.
→ Local control

Precision of Design

## CRD
## (Completely Randomized Design)

Introduction :-

It is the simplest design based on two basic principles of design of experiment viz. replication & randomization. Here, treatments are allocated in experimental unit in such a way that each treatment may occur in any no. of times. Here, we wish to study effect of treatments only we use one way ANOVA to analyze CRD.

Let us consider 't' treatments replicated r times then an experimental unit should be divided into

$t \cdot r = N$ subunits

Layout of CRD

let us consider 4 treatments A, B, C, D replicated 3 times then layout of CRD may be as

| | | | |
|---|---|---|---|
| A | A | B | C |
| B | C | A | D |
| C | B | D | D |

Scanned by CamScanner

# Statistical Analysis of CRD

$y_{ij} = \mu + \tau_i + \epsilon_{ij}$ $\forall$ $i = 1, 2, \ldots t$
$j = 1, 2, \ldots r$

$y_{ij}$ = Observation of $j^{th}$ repeatation from $i^{th}$ treatment

$\mu$ = overall mean
$\tau_i$ = $i^{th}$ treatment effect
$\epsilon_{ij}$ = Error term such that $\epsilon_{ij}$ are i.i.d v.v.

Here, we split total variation into two components as

$$\sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij}-\bar{y}_{..})^2 = \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij}-\bar{y}_{i.}+\bar{y}_{i.}-\bar{y}_{..})^2$$

$$= \sum_{i=1}^{t}\sum_{j=1}^{r}\{(\bar{y}_{i.}-\bar{y}_{..})+(y_{ij}-\bar{y}_{i.})\}^2$$

$$= \sum_{i=1}^{t}\sum_{j=1}^{r}(\bar{y}_{i.}-\bar{y}_{..})^2 + \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij}-\bar{y}_{i.})^2 + \text{cross product term}$$

[Since sum of product of deviation taken from A.M is zero]

$$\therefore \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij}-\bar{y}_{..})^2 = r\sum_{i=1}^{t}(\bar{y}_{i.}-\bar{y}_{..})^2 + \sum_{i=1}^{t}\sum_{j=1}^{t}(y_{ij}-\bar{y}_{i.})^2$$

$\therefore$ TSS = SST + SSE

Scanned by CamScanner

Now ANOVA table for CRD

| S.V | d.f | S.S | M.SS | F-ratio |
|---|---|---|---|---|
| Treat | $t-1$ | SST | $MSST = \frac{SST}{t-1}$ | $F_T = \frac{MSST}{MSSE}$ |
| Error | $t(r-1)$ | SSE | $MSSE = \frac{SSE}{t(r-1)}$ | |
| Total | $r.t-1$ | TSS | | |

Decision

* If $F_{cal} \leq F_{tab}$, we accept $H_0$ & reject $H_1$
* If $F_{cal} > F_{tab}$, we reject $H_0$ & accept $H_1$

* For Practical Purpose

$$SST = \sum_{i=1}^{t} \sum_{j=1}^{r} y_{ij}^2 - C.F$$

$$TSS = \sum_{i=1}^{t} \frac{y_{i.}^2}{r} - C.F$$

$$C.F = \frac{G^2}{N} = \frac{y_{..}^2}{r.t} \text{ (Correction factor)}$$

$$SS.E = SST - T.S.S$$

Scanned by CamScanner

Date
Page

Scanned by CamScanner

ES

We know

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad \forall \quad i = 1, 2, \ldots t$$
$$j = 1, 2 \ldots r$$

Here, we estimate $\mu$ & $\tau_i$ such that Error sum of square will be minimized

$$E = \sum\sum \varepsilon_{ij}^2$$

Then, $E = \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{ij} - \mu - \tau_i)^2$

For $\mu$

$$\frac{\partial E}{\partial \mu} = -2 \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{ij} - \mu - \tau_i)$$

$$\frac{\partial^2 E}{\partial \mu^2} = -2 \sum_{i=1}^{t} \sum_{j=1}^{r} (-1) = 2.t.r = 2N > 0$$

Here For estimation of $\mu$ :-

$$\frac{\partial E}{\partial \mu} = 0$$

$$\Rightarrow -2 \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{ij} - \mu - \tau_i) = 0$$

$$\Rightarrow \sum_{i=1}^{t} \sum_{j=1}^{r} y_{ij} - \sum_{i=1}^{t} \sum_{j=1}^{r} \mu - \sum_{i=1}^{t} \sum_{j=1}^{r} \tau_i = 0$$

$$\Rightarrow y_{..} - t.r.\mu - r \sum_{i=1}^{t} \tau_i = 0 \quad \left( \text{Imposing } \sum_{i=1}^{t} \tau_i = 0 \right)$$

$$\Rightarrow y_{..} - N\mu = 0$$

Scanned by CamScanner

$\therefore \quad N\mu = y_{..}$

$\therefore \quad \mu = \dfrac{y_{..}}{N}$

$\therefore \quad \hat{\mu} = \bar{y}_{..}$

For $\tau_i$

$$\frac{\partial E}{\partial \tau_i} = -2 \sum_{j=1}^{r} (y_{ij} - \mu - \tau_i)$$

$$\frac{\partial^2 E}{\partial \tau_i^2} = -2 \sum_{j=1}^{r} (-1) = 2r > 0$$

Now, for estimation of $\tau_i$

$$\frac{\partial E}{\partial \hat{\tau}_i} = 0$$

$$\therefore \quad -2 \sum_{j=1}^{r} (y_{ij} - \mu - \tau_i) = 0$$

$$\therefore \quad \sum_{j=1}^{r} y_{ij} - r\mu - r\tau_i = 0$$

$\therefore \quad y_{i.} - r\mu = r\tau_i$

$\therefore \quad \tau_i = \dfrac{y_{i.}}{r} - \mu$

$\therefore \quad \hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$

Hence $\hat{\mu} = \bar{y}_{..}$

$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$

Scanned by CamScanner

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 10 | B | 5 | A | 20 | C | 15 |
| B | 6 | A | 15 | C | 11 | B | 10 |
| C | 22 | B | 12 | C | 18 | A | 16 |

Calc$^n$ table for various sum of variation (for CRD)

| | | | | | $y_{i.}$ |
|---|---|---|---|---|---|
| A | 10 | 15 | 20 | 16 | 61 |
| B | 5 | 6 | 10 | 12 | 33 |
| C | 15 | 11 | 22 | 18 | 66 |
| | | | | | 160 |

Here $t = 3$, $r = 4$, $N = t \cdot r = 3 \times 4 = 12$

For C.F

$$C.F = \frac{G^2}{N} = \frac{160^2}{12} = 2133.33$$

For TSS

$$T.SS = \sum\sum y_{ij}^2 - C.F.$$

$$= 10^2 + 15^2 + \cdots + 18^2 - 2133.33$$

$$\therefore T.S.S = 306.67$$

For SST

$$S.S.T = \sum \frac{y_{i.}^2}{r} - C.F$$

$$= \frac{61^2 + 33^2 + 66^2}{4} - 2133.33$$

$$= 158.17$$

For SSE

$$SSE = TSS - SST = 306.67 - 158.17$$

$$\therefore SSE = 148.5$$

Scanned by CamScanner

Now, ANOVA Table

| S.V | d.f | S.S | M.SS | F. ratio |
|---|---|---|---|---|
| Treat | 3-1=2 | 158.17 | 79.085 | $F_T = 4.79$ |
| Error | 11-2=9 | 148.5 | 16.5 | |
| Total | 12-1=11 | 306.67 | | |

To test whether there is sig diff due to treatment or not we set up hypothesis as

$H_0 : \mu_A = \mu_B = \mu_C$ i.e. there is no sig diff due to treatment

Vs $H_1 : \mu_i \neq \mu_j$ (For atleast one pair) i.e there is sig diff due to at least one pair of treatment

Under $H_0$ Test statistic

$$F_T = \frac{MSST}{MSSE}$$

$F_{cal} = 4.79$

Here, level of sig $(\alpha) = 5\%$

d.f $= (2, 9)$

$F_{tab} = 4.26$

Decision

Since $F_{cal} > F_{tab}$, we reject $H_0$ & accept $H_1$ with the conclusion that there is sig diff due to at least one pair of observation.

Scanned by CamScanner

Date
Page

# RBD (Randomized Block Design)

It is the designed based on all three principles of design of experiment. Here we allocate treatment in completely random manner. Each treatment must occur once & only once in each row or column.

## Layout of RBD.

Let us consider 't' treatments ~~& b block~~ are allocated in b blocks then experimental unit should be divided into t.r sub units. Here we are intesested to study the effect of treatment and blocks so we used two way ANOVA to analyze RBD

For 4 treatments & 3 blocks layout of RBD may be as

| | Block I | Block II | Block III |
|---|---|---|---|
| | A | B | D |
| | B | A | A |
| | C | C | B |
| | D | D | C |

OR

| | | | | | |
|---|---|---|---|---|---|
| Block | I | B | C | D | A |
| | II | A | B | D | C |
| | III | A | C | B | D |

Scanned by CamScanner

Statistical Analysis of RBD

Statistical model for RBD is

$$y_{ij} = \mu + \tau_i + B_j + \varepsilon_{ij} \quad \forall \; i = 1, 2 \ldots t$$
$$j = 1, 2 \ldots r$$

where, $y_{ij}$ = Observation of $i^{th}$ treatment from $j^{th}$ block

$\mu$ = overall mean

$\tau_i$ = $i^{th}$ treatment effect

$B_j$ = $j^{th}$ Block

$\varepsilon_{ij}$ = Error term such that $\varepsilon_{ij}$ are i.i.d.r.v with $\varepsilon_{ij} \sim N(0, \sigma_e^2)$

Total sum of square = Sum of square due to treatment, sum of square due to block & sum of square due to error

$$\therefore \; T.S.S = \sum \sum (y_{ij} - \bar{y}_{..})^2$$

Here

$$\sum_{i=1}^{t} \sum_{j=1}^{r} (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{i.} - \bar{y}_{..})^2$$

$$= \sum_{i=1}^{t} \sum_{j=1}^{r} \{ (\bar{y}_{i.} - \bar{y}_{..}) + (y_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) \}^2$$

$$= \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j})^2$$

or, $$\sum_{i=1}^{t} \sum_{j=1}^{r} (y_{ij} - \bar{y}_{..})^2 = r \sum_{i=1}^{t} (\bar{y}_{i.} - \bar{y}_{..})^2 + t \sum_{j=1}^{r} (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

$$TSS = SST + SSB + SSE$$

Scanned by CamScanner

ANOVA table for RBD

| S.V | d.f | S.S | M.SS | F. ratio |
|---|---|---|---|---|
| treatment | $t-1$ | SST | $MSST = \frac{SST}{t-1}$ | $F_T = \frac{MSST}{MSSE}$ |
| block | $r-1$ | SSB | $MSSB = \frac{SSB}{r-1}$ | $F_B = \frac{MSSB}{MSSE}$ |
| Error | $(t-1)(r-1)$ | SSE | $MSSE = \frac{SSE}{(t-1)(r-1)}$ | |
| total | $rt-1$ | T.S.S | | |

Decision

*) If $F_{cal} \leq F_{tab}$, we accept $H_0$ & reject $H_1$

*) If $F_{cal} > F_{tab}$, we reject $H_0$ & accept $H_1$

For Practical Purpose

$$C.F = \frac{G^2}{N} = \frac{y_{..}^2}{N}$$

$$TSS = \sum_{i=1}^{t} \sum_{j=1}^{r} y_{ij}^2 - C.F$$

$$SST = \sum_{i=1}^{t} \frac{y_{i.}^2}{r} - C.F$$

$$SSB = \sum_{j=1}^{r} \frac{y_{.j}^2}{t} - C.F$$

$$SSE = TSS - SST - SSB$$

Scanned by CamScanner

# Estimation of Parameters of RBD

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad \forall \; i = 1, 2 \ldots t$$
$$j = 1, 2 \ldots r$$

Here, we estimate parameters of RBD in such a way that errors sum of square will be minimized.

$$E = \sum \varepsilon_{ij}^2$$

$$\therefore E = \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{ij} - \mu - \tau_i - \beta_j)^2 \quad \text{——— (i)}$$

We should impose $\sum_{i=1}^{t} \tau_i = \sum_{j=1}^{r} \beta_j = 0$

For $\mu$

$$\frac{\partial E}{\partial \mu} = -2 \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{ij} - \mu - \tau_i - \beta_j)$$

$$\frac{\partial^2 E}{\partial \mu^2} = -2 \sum_{i=1}^{t} \sum_{j=1}^{r} (-1) = 2 \cdot t \cdot r = 2N > 0$$

Here, $\frac{\partial E}{\partial \mu} = 0$

or, $-2 \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{ij} - \mu - \tau_i - \beta_j) = 0$

or, $\sum_{i=1}^{t} \sum_{j=1}^{r} y_{ij} - t \cdot r \cdot \mu - r \sum_{i=1}^{t} \tau_i - t \sum_{j=1}^{r} \beta_j = 0$

or, $y_{..} - N\mu - 0 - 0 = 0$

or $N\mu = y_{..}$

or $\mu = \frac{y_{..}}{N}$

$\therefore \hat{\mu} = \bar{y}_{..}$

Scanned by CamScanner

For $\tau_i$

$$\frac{\partial E}{\partial \tau_i} = -2 \sum_{j=1}^{r} (y_{ij} - \mu - \tau_i - \beta_j)$$

$$\frac{\partial^2 E}{\partial \tau_i^2} = -2 \sum_{j=1}^{r} (-1) = 2r > 0$$

Here $\frac{\partial E}{\partial \tau_i} = 0$

or, $-2 \left[ \sum_{j=1}^{r} y_{ij} - r.\mu - r\,\tau_i - \sum_{j=1}^{r} \beta_j \right] = 0$

or, $y_{i.} - r\mu - r\tau_i - 0 = 0$

or, $r\tau_i = y_{i.} - r\mu$

or, $\tau_i = \frac{y_{i.}}{r} - \mu$

or $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$

For $\beta_j$

$$\frac{\partial E}{\partial \beta_j} = -2 \sum_{j=1}^{t} (y_{ij} - \mu - \tau_i - \beta_j)$$

$$\frac{\partial^2 E}{\partial \beta_j^2} = -2 \sum_{j=1}^{t} (-1) = 2t > 0$$

Here, $\frac{\partial E}{\partial \beta_j} = 0$

or, $-2 \sum_{i=1}^{t} (y_{ij} - \mu - \tau_i - \beta_j) = 0$

or, $\sum_{i=1}^{t} y_{ij} - t\mu - \sum_{i=1}^{t} \tau_i - t\beta_j = 0$

or, $y_{.j} - t\mu = t\beta_j$

or, $\beta_j = \frac{y_{.j}}{t} - \mu$

$\therefore \hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$

Scanned by CamScanner

Hence, estimation of parameter for RBD

$\hat{\mu} = \bar{y}_{..}$

$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$

$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$

Note

Relative efficiency of RBD with CRD

$$E = \frac{r(t-1)\,MSE + (r-1)\,MSB}{(r+-1)\,MSE}$$

* If $E > 1$, RBD is more efficient than CRD
* If $E = 1$ " " equally " " "
* If $E < 1$ " " less " " "

Sol

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | 8 | C | 10 | A | 6 | B | 10 |
| C | 12 | B | 8 | B | 9 | A | 8 |
| B | 10 | A | 8 | C | 10 | C | 9 |

| Treatments | Blocks I | II | III | IV | $y_{i.}$ |
|---|---|---|---|---|---|
| A | 8 | 8 | 6 | 8 | 30 |
| B | 10 | 8 | 9 | 10 | 37 |
| C | 12 | 10 | 10 | 9 | 41 |
| $y_{.j}$ | 30 | 26 | 25 | 27 | $y_{..} = G = 108$ |

Here, $t = 3$, $r = 4$, $N = t.r = 12$

For C.F

$$C.F = \frac{G^2}{N} = \frac{108^2}{12} = 972$$

For TSS

$$TSS = \sum\sum y_{ij}^2 - CF = 8^2 + 8^2 + \dots + 9^2 - 972$$
$$= 26$$

Scanned by CamScanner

For SST

$$SST = \frac{\sum y_{i.}^2}{r} - C.F$$

$$= \frac{36^2 + 37^2 + 41^2}{4} - 972$$

$$= 15.5$$

For SSB

$$SSB = \frac{\sum y_{.j}^2}{t} - C.F = \frac{30^2 + 26^2 + 25^2 + 27^2}{3} - 972$$

$$= 4.67$$

For SSE

$$SSE = TSS - SST - SSB$$

$$= 26 - 15.5 - 4.67$$

$$= 5.83$$

Now

ANOVA table

| S.V | d.f | S.S | MSS | F-ratio |
|---|---|---|---|---|
| Trat | 3-1=2 | 15.5 | 7.75 | $F_T = 7.98$ |
| Block | 4-1=3 | 4.67 | 1.56 | $F_B = 1.61$ |
| Error | 6 | 5.83 | 0.971 | |
| Total | 12-1=11 | 26 | | |

Scanned by CamScanner

For Treatment

To test whether there is sig. diff due to treatment or not we set up hypothesis as

$H_0 : \mu_1 = \mu_2 = \mu_3$ i.e. there is no sig diff due to treatment

Vs $H_1 : \mu_i \neq \mu_j$ (For atleast one pair) i.e. there is sig diff due to at least one pair of treatments

2nd

Under $H_0$ test-statistics

$$F_T = \frac{MSST}{MSSE} \Rightarrow F_{cal} = 7.98$$

Here, level of sig $(\alpha) = 5\%$, d.f $= (2, 6)$

From table

$$F_{tab} = \cancel{10.9}\ 5.14$$

Decision

Since, $F_{cal} > F_{tab}$, we reject $H_0$ & accept $H_1$ with the conclusion that there is significant difference due to atleast one pair of treatments.

For Blocks

To test whether there is

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ ie

Vs $H_1 : \mu_i \neq \mu_j$ (For atleast one pair) i.e.

Under $H_0$, test - statistic

$$F_B = \frac{MSSB}{MSSE} \Rightarrow F_{cal} = 1.61$$

Here level of sig $(\alpha) = 5\%$ d.f $= (3, 6)$

From table

$$F_{tab} = 4.76$$

Decision:

Since $F_{cal} < F_{tab}$; we reject $H_1$ & accept $H_0$ with the conclusion that there is no sig diff. due to blocks.

part

For efficiency of RBD relative to CRD

$$E = \frac{r(t-1) \cdot MSE + (r-1) \cdot MSB}{(rt-1) MSE}$$

$$= \frac{4 \times 2 \times 0.971 + 3 \times 1.56}{11 \times 0.971}$$

$\therefore E = 1.1626 > 1$

$\Rightarrow$ RBD is more efficient that CRD

Scanned by CamScanner

# Estimation of Missing Value of RBD.

Let $y_{ij} = x$ be missing value in RBD

$y'_{..}$ = sum of all known observation

$y'_{i.}$ = Sum of all known obj of $i^{th}$ treatment

$y'_{.j}$ = Sum of all know obj of $j^{th}$ blade

Here, $G = y'_{..} + x$, $N = rt$

For $C.F = \dfrac{(y'_{..} + x)^2}{N}$

For T.S.S

$$TSS = \sum\sum y_{ij}^2 - C.F$$

$$\therefore TSS = x^2 + \text{constant} - \frac{(y'_{..} + x)^2}{N}$$

For SST

$$SST = \frac{\sum y_{i.}^2}{r} - C.F$$

$$= \frac{(y'_{i.} + x)^2}{r} + \text{constant} - \frac{(y'_{..} + x)^2}{N}$$

For SSB

$$SSB = \frac{\sum y_{.j}^2}{t} - C.F$$

$$= \frac{(y'_{.j} + x)^2}{t} + \text{constant} - \frac{(y'_{..} + x)^2}{N}$$

Here, we estimate missing value in such a way that sum of square due to error will be minimized.

Scanned by CamScanner

Here, $SSE = T.S.S - SST - SSB$

$= x^2 + \text{constant} - \frac{(y_{i.}' + x)^2}{r} - \frac{(y_{.j}' + x)^2}{t} + \frac{(y_{..}' + x)^2}{N}$

For Minima

$$\frac{\partial SSE}{\partial x} = 0$$

or, $2x + 0 - \frac{2(y_{i.}' + x)}{r} - \frac{2(y_{.j}' + x)}{t} + \frac{2(y_{..}' + x)}{N} = 0$

or, $r.t.x - t(y_{i.}' + x) - r(y_{.j}' + x) + (y_{..}' + x) = 0$

or, $(rt - t - r + 1)x = t\,y_{i.}' + r\,y_{.j}' - y_{..}'$

$\therefore \hat{x} = \frac{t\,y_{i.}' + r\,y_{.j}' - y_{..}'}{(r-1)(t-1)}$

Now, for missing case

d.f for total $= rt - 2$

Scanned by CamScanner

14.) Sol$^n$

let missing observation be $x$

Now, Cal$^n$ table for various sum

| Treat | Block I | II | III | IV | $y_{i.}$ |
|---|---|---|---|---|---|
| P | 19 | 21 | 23 | 26 | $x+68$ |
| Q | 26 | 28 | 27 | 33 | 114 |
| R | 21 | 29 | 22 | 26 | 98 |
| $y_{.j}$ | 66 | $x+57$ | 72 | 85 | $x+280$ |

Here $t=3$, $r=4$ $N=12$

$y'_{i.} = 68$, $y'_{.j} = 57$ $y'_{..} = 280$

We know, for RBD estimate of missing obs;

$$x = \frac{t.y'_{i.} + r.y'_{.j} - y'_{..}}{(r-1)(t-1)} = \frac{3\times68 + 4\times57 - 280}{3\times2} = 25.3$$

For C.F

$$C.F = \frac{G^2}{N} = \frac{305.3^2}{12} = 7767.34$$

For TSS

$$TSS = \sum\sum y_{ij}^2 - C.F$$
$$= 19^2 + 25.3^2 + \dots + 26^2 - 7767.34$$
$$= 158.75$$

For SST

$$SST = \sum \frac{y_{i.}^2}{r} - C.F = \frac{133^2 + 114^2 + 98^2}{4} - 7767.37$$
$$= 58.88$$

For SSB

$$SSB = \sum \frac{y_{.j}^2}{t} - C.F = \frac{66^2 + 82.3^2 + 72^2 + 85^2}{3} - 7767.34$$
$$= 78.27$$

Scanned by CamScanner

For SSE

$SSE = TSS - SST - SSB$

$=$

$SSE = 21.12$

Now ANOVA table for given problem

| S.V | df | S.S | M.S.S | F-ratio |
|---|---|---|---|---|
| Treat | $3-1=2$ | 58.88 | 29.44 | $F_T = 6.96$ |
| Block | $4-1=3$ | 78.75 | 26.25 | $F_B = 6.21$ |
| Error | 6 | 21.12 | 4.224 | |
| Total | $12-1=11$ | | | |

For treat

To test whether there is sig diff or not we setup hypothesis as

$H_0 : \mu_1 = \mu_2 = \mu_3$ i.e. there is no sig. diff due to treatment

$H_1 : \mu_i \neq \mu_j$ (for at least 1 pair) i.e. there is sig diff " " "

Under $H_0$ - test statistic

$F_T = \frac{MSST}{MSSE} \Rightarrow F_{cal} = 6.96$

Here level of sig $(\alpha) = 5\%$, d.f $= (2, 5)$

$F_{tab} = 5.79$

Decision

Since $F_{cal} > F_{tab}$ we reject $H_0$ & accept $H_1$ with the conclusion that there is significant difference due to at least one pair of treatments.

Scanned by CamScanner

# Latin Square Design

This design is based on all three principles of experimental design. ~~of experiment~~. Here each treatment must occur only once in each row and column. So this design is also known as row-wise, column-wise and treatment was designed. Here, no. of rows must be equal to no. of columns and no. of treatments. In this design we are interested to design study the variations due to row, column and treatment, So we are incomplete 3 way ANOVA to analyse L.S.D

Since latin alphabets are used to denote the treatments So, named latin Square Design.

## Layout of L.S.D

Let us consider L.S.D with m treatments then an experimental unit is subdivided into m x m sub unit. i.e. there must be m rows & m-columns.

Let us consider 4x4 L.S.D then layout of L.S.D may be as

| | | | |
|---|---|---|---|
| B | A | D | C |
| C | D | A | B |
| A | B | C | D |
| D | C | B | A |

Standard LSD

| | | | |
|---|---|---|---|
| A | B | C | D |
| B | C | D | A |
| C | D | A | B |
| D | A | B | C |

Scanned by CamScanner

Date
Page

Analysis of L.S.D

Statistical model for L.S.D is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \varepsilon_{ijk} \quad \forall i,j,k = 1,2 \ldots m$$

where

$y_{ijk}$ = obs. of $k^{th}$ treatment from $i^{th}$ row & $j^{th}$ column

$\mu$ = overall effect

$\alpha_i$ = $i^{th}$ row effect

$\beta_j$ = $j^{th}$ column effect

$\tau_k$ = $k^{th}$ treat. effect

$\varepsilon_{ijk}$ = Error term s.t $\varepsilon_{ijk}$ are i.i.d.r.v &

$\varepsilon_{ijk} \sim N(0, \sigma_e^2)$

Here we split total variations into variation due to row, column, treatment & error variations as

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}(y_{ijk} - \bar{y}_{...})^2 = \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{.j.} + \bar{y}_{..k} - \bar{y}_{...})^2$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}\{(\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{..k} - \bar{y}_{...}) + (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...})\}^2$$

Since cross product term vanishes, so,

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}(y_{ijk} - \bar{y}_{...})^2 = \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}(\bar{y}_{i..} - \bar{y}_{...})^2 + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}(\bar{y}_{.j.} - \bar{y}_{...})^2$$

$$+ \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}(y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...})^2$$

Scanned by CamScanner

$$\therefore \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}(y_{ijk}-\bar{y}_{...})^2 = m\sum_{i=1}^{m}(\bar{y}_{i..}-\bar{y}_{...})^2 + m\sum_{j=1}^{m}(\bar{y}_{.j.}-\bar{y}_{...})^2$$

$$+ m\sum_{k=1}^{m}(\bar{y}_{..k}-\bar{y}_{...})^2 + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}(y_{ijk}-\bar{y}_{i..}-\bar{y}_{.j.}-\bar{y}_{..k}+\bar{y}_{...})^2$$

$$TSS = SSR + SSC + SST + SSE$$

ANOVA table for L.S.D

| S.V | d.f | S.S | M.S.S | F-ratio |
|---|---|---|---|---|
| Row | $m-1$ | SSR | $MSSR = \frac{SSR}{m-1}$ | $F_R = MSSR/MSSE$ |
| Column | $m-1$ | S.S.C | $MSSC = SSC/m-1$ | $F_C = MSSC/MSSE$ |
| Treat. | $m-1$ | S.S.T | $MSST = SSST/m-1$ | $F_T = MSST/MSSE$ |
| Error | $(m-1)(m-2)$ | S.S.E | $MSSE = \frac{SSE}{(m-1)(m-2)}$ | |
| Total | $m^2-1$ | T.SS | | |

Decision

If $F_{cal} \le F_{tab}$, we accept $H_0$ & reject $H_1$

If $F_{cal} > F_{tab}$ we reject $H_0$ & accept $H_1$

For Practical Purpose

$$C.F = \frac{G^2}{N} = \frac{y_{...}^2}{N}, \quad N = m^2$$

$$T.S.S = \sum\sum\sum y_{ijk}^2 - C.F$$

$$S.S.R = \frac{\sum y_{i..}^2}{m} - C.F$$

Scanned by CamScanner

Date
Page

$$S.S.C = \sum \frac{y_{.j.}^2}{m} - C.F$$

$$S.S.T = \sum \frac{y_{ijk}^2}{m} - C.F$$

$$SSE = T.S.S - S.S.R - SS.C - S.S.T$$

Scanned by CamScanner

# ESTIMATION OF MISSING VALUE IN LSD

let us consider $m \times m$ LSD whose one observation $y_{ijk}$ is missing.

Now, let $x = y_{ijk}$

$y'_{...}$ = Sum of all known observation

$y'_{i..}$ = Sum of " " " of $i^{th}$ row

$y'_{.j.}$ = Sum of " " " of $j^{th}$ column

$y'_{..k}$ = " " " " " $k^{th}$ column

Then

$$y_{...} = y'_{...} + x, \quad y_{i..} = y'_{i..} + x$$

$$y_{.j.} = y'_{.j.} + x, \quad y_{..k} = y'_{..k} + x$$

Here $N = m^2$

Now,

$$C.F = \frac{G^2}{N} = \frac{(y'_{...} + x)^2}{N}$$

For T.S.S

$$T.S.S = \sum\sum\sum y_{ijk}^2 - C.F$$

$$\therefore T.S.S = x^2 + \text{constant} - \frac{(y'_{...} + x)^2}{N}$$

For SSR

$$S.S.R = \sum \frac{y_{i..}^2}{m} - C.F$$

$$\therefore S.S.R = \frac{(y'_{i..} + x)^2}{m} + \text{constant} - \frac{(y'_{...} + x)^2}{N}$$

For S.S.C

$$SSC = \sum \frac{y_{.j.}^2}{m} - C.F$$

$$SSC = \frac{(y'_{.j.} + x)^2}{m} + \text{constant} - \frac{(y'_{...} + x)^2}{}$$

Scanned by CamScanner

For SST

$$SST = \sum \frac{y_{..k}^2}{m} - C.F$$

$$\therefore SST = \frac{(y'_{..k} + x)^2}{m} + \text{constant} - \frac{(y'_{...} + x)^2}{N}$$

Now

$$SSE = TSS - SSR - SSC - SST$$

$$\therefore SSE = x^2 + \text{constant} - \frac{(y'_{i..} + x)^2}{m} - \frac{(y'_{.j.} + x)^2}{m} - \frac{(y'_{..k} + x)^2}{m} + 2\frac{(y'_{...} + x)^2}{N}$$

Here we estimate $x$ in such a way that sum of squar due to error will be minimized

Now,

$$\frac{\partial SS}{\partial x} = 2x - \frac{2(y'_{i..} + x)}{m} - \frac{2(y'_{.j.} + x)}{m} - \frac{2(y'_{..k} + x)}{m} + \frac{4(y'_{...} + x)}{N}$$

$$\frac{\partial^2 SSE}{\partial x^2} = 2 - \frac{2}{m} - \frac{2}{m} - \frac{2}{m} + \frac{4}{N} = \frac{2}{m^2}(m^2 - 3m + 2)$$

$$= \frac{2}{m^2}(m-1)(m-2) > 0$$

For minima

$$\frac{\partial SSE}{\partial x} = 0$$

$$\text{or, } 2x - \frac{2(y'_{i..} + x)}{m} - \frac{2(y'_{.j.} + x)}{m} - \frac{2(y'_{..k} + x)}{m} + \frac{4(y'_{...} + x)}{n^2} = 0$$

Scanned by CamScanner

or, $\frac{2}{m^2}\left[m^2 x - m(y'_{i..} + x) - m(y'_{.j.} + x) - m(y'_{..k} + x) + 2(y'_{...} + x)\right] = 0$

or, $(m^2 - m - m - m + 2)x = m y'_{i..} + m y'_{.j.} + m y'_{..k} - 2y'_{...}$

$$\therefore \hat{x} = \frac{m y'_{i..} + m y'_{.j.} + m y'_{..k} - 2y'_{...}}{(m-1)(m-2)}$$

* EFFICIENCY OF LSD Relative to CRD

$$E = \frac{(m-1)MSE + MSR + MSC}{(m+1)MSE}$$

* Efficiency of LSD relative to R.B.D

i) Column is treated as Block

$$E = \frac{(m-1)MSE + MSR}{m.MSE}$$

ii) Row is treated as Block

$$E = \frac{(m-1)MSE + MSC}{m.MSE}$$

Scanned by CamScanner

Calⁿ table for varion sum

| | | | $y_{i..}$ |
|---|---|---|---|
| A 10 | B 15 | C 20 | 45 |
| B 25 | C 10 | A 15 | 50 |
| C 25 | A 20 | B 15 | 60 |
| 60 | 45 | 50 | G = |

For treatment

| | | $y_{..k}$ |
|---|---|---|
| A : | 10 + 15 + 20 | 45 |
| B : | 15 + 25 + 15 | 55 |
| C : | 20 + 10 + 25 | 55 |
| | | G = 155 |

For SSC

$$SSC = \frac{\sum y_{.j.}^2}{m} - C.F$$

$$= \frac{60^2 + 45^2 + 50^2}{3} - 2669.44 = 38.89$$

For SST

$$SST = \frac{\sum y_{..k}^2}{m} - CF = \frac{45^2 + 55^2 + 55^2}{3} - 2669.44 = 23.22$$

For SSE

$$SSE = TSS - SSR - SSC - SST$$

$$= 255.56 - 38.89 - 38.89 - 23.22 = 154.56$$

Here $m = 3$, $N = m \times m = 9$

For C.F

$$C.F = \frac{G^2}{N} = \frac{155^2}{9} = 2669.44$$

For TSS

$$T.SS = \sum\sum\sum y_{ijk}^2 - C.F$$

$$= 10^2 + 15^2 + \dots + 15^2 - 2669.44$$

$$= 255.56$$

For SSR

$$SSR = \frac{\sum y_{i..}^2}{m} - C.F$$

$$= \frac{45^2 + 50^2 + 60^2}{3} - 2669.44$$

$$= 38.89$$

ANOVA TABLE

Scanned by CamScanner

| S.V | d.f | S.S | M.SS | F ratio |
|---|---|---|---|---|
| Row | 2 | 38.80 | 19.445 | $F_R = 0.252$ |
| Column | 2 | 38.80 | 19.445 | $F_C = 0.252$ |
| Treat | 2 | 23.22 | 11.61 | $F_T = 0.15$ |
| Error | 2 | 154.58 | 77.28 | |
| Total | 9-1=8 | 255.56 | | |

For Row

To test whether there is sig. diff due to row or not we set hypothesis as

$H_0 : \mu_1 = \mu_2 = \mu_3$ ie there is no sig diff due to row

Vs $H_1 : \mu_i \neq \mu_j$ ie there is sig diff due to 1 pair of row

Under $H_0$ : statistic

$$F_R = \frac{MSSR}{MSSE} \Rightarrow F_{cal} = 0.252$$

Here level of sig ($\alpha$) = 5%, d.f = (2,2)

From table $F_{tab} = 19.01$

Decision

Since $F_{cal} < F_{tab}$, we accept $H_0$ & reject $H_1$ with the conclusion that there is no significant difference due to row.

For Column

To test whether there is sig diff due to column or not we set up hypothesis as

$H_0 : \mu_1 = \mu_2 = \mu_3$ ie column

Vs $H_1 : \mu_i \neq \mu_j$ ie column

Under $H_0$ - statistic

$$F_C = \frac{MSSC}{MSSE} \Rightarrow F_C = 0.252$$

$F_{cal} < F_{tab}$

= 5%, d.f = (2,2) $F_{tab} = 19.01$ Accept $H_0$

For treatment

To ..... treatment or not ...

$H_0 : \mu_1 = \mu_2 = \mu_3$ ie ..... treatment

Vs $H_1 : \mu_i \neq \mu_j$ ..... treatment

Under $H_0$, test-statistic

$$F_T = \frac{MSST}{MSSE} \Rightarrow F_{cal} = 0.15$$

$\alpha = 5\%$, d.f $= (2,2)$ $F_{tab} = 19.01$

$F_{cal} < F_{tab}$, accept $H_0$ & reject $H_1$

Efficiency relative to CRD

$$E = \frac{(m-1)\,MSE + MSC + MSR}{(m+1)\,MSE}$$

$$= \frac{2 \times 77.28 + 19.445 + 19.445}{4 \times 77.28}$$

$$= 0.63$$

Efficiency relative to RBD

i) When Row is taken as Block

$$E = \frac{(m-1)\,MSE + MSC}{m\,MSE} = \frac{2 \times 77.28 + 19.445}{3 \times 77.28} = 0.75$$

ii) When Column is taken as Block

$$E = \frac{(m-1)\,MSE + MSR}{m \cdot MSE} = \frac{2 \times 77.48 + 19.445}{3 \times 77.28} = 0.75$$

Scanned by CamScanner

15) Do^n

| | | | |
|---|---|---|---|
| A 12 | C 19 | B 10 | D 8 |
| C 18 | B 12 | D 6 | A ? |
| B 22 | D 10 | A 5 | C 21 |
| D 12 | A 7 | C 27 | B 17 |

let missing value be $x$

Here $y'_{i..} = 18+12+6 = 36$

$y'_{.j.} = 8+21+17 = 46$ $\quad y'_{..k} = 12+5+7 = 24$

$y'_{...} = 12+19+\dots+17 = 206$

Here $m = 4$, $N = m^2 = 16$

~~$y'_{...} = 12+$~~

We know

$$\hat{x} = \frac{m y'_{i..} + m y'_{.j.} + m y'_{..k} - 2y'_{...}}{(m-1)(m-2)}$$

$$= \frac{4\times 36 + 4\times 46 + 4\times 24 - 2\times 206}{3\times 2}$$

$\hat{x} = 2$

Now, Cal^n table for various sum

| | | | | | $y'_{i..}$ |
|---|---|---|---|---|---|
| | A 12 | C 19 | B 10 | D 8 | 49 |
| | C 18 | B 12 | D 6 | A 2 | 38 |
| | B 22 | D 10 | A 5 | C 21 | 36 |
| | D 12 | A 7 | C 27 | B 17 | 63 |
| $y_{.j.}$ | 64 | 48 | 48 | 48 | 208 |

For treatment

A : 12 + 2 + 5 + 7 $\quad y_{..k}$ 26

B : 10 + 12 + 22 + 17 $\quad$ 61

C : 19 + 18 + 21 + 27 $\quad$ 85

D : 8 + 6 + 10 + 12 $\quad$ 36

For C.F

$$C.F = \frac{G^2}{N} = \frac{208^2}{16}$$

$C.F = 2704$

Colu.

Treatme

Error

Total

Scanned by CamScanner

For TSS

$TSS = \sum\sum\sum y_{ijk}^2 - C.F = 12^2 + 19^2 + \ldots + 17^2 - 2704$

$= 734$

For SSR

$SSR = \frac{\sum y_{i..}^2}{m} - C.F = \frac{49^2 + 38^2 + 58^2 + 63^2}{4} - 2704$

$= 90.5$

For SSC

$SSC = \frac{\sum y_{.j.}^2}{m} - C.F = \frac{64^2 + 48^2 + 48^2 + 48^2}{4} - 2704 = 48$

For SST

$SST = \frac{\sum y_{..k}^2}{m} - C.F = \frac{26^2 + 61^2 + 85^2 + 36^2}{4} - 2704$

$= 525.5$

For SSE

$SSE = TSS - SSR - SSC - SST = 70$

ANOVA table

| S.V | d.f | S.S | M.S.S | F-ratio |
|---|---|---|---|---|
| Row | 3 | 90.5 | 30.17 | $F_R = 2.16$ |
| Lum | 3 | 48 | 16 | $F_C = 1.14$ |
| ment | 3 | 525.5 | 175.17 | $F_T = 12.51$ |
| r | 5 | 70 | 14 | |
| | $m^2-2 = 14$ | 734 | | |

Scanned by CamScanner

Here, we have

| S.V | d.f | S.S | M.SS | F-ratio |
|---|---|---|---|---|
| Blocks | 4 | 26.8 | | |
| Treat | 3 | | | |
| Error | 12 | | 2.5 | |
| Total | 19 | 85.3 | | |

This is RBD with 5 blocks and 4 treatment

We know,

d.f for Blocks $= r-1$

or $4 = r-1$

$\therefore r = 5$

& d.f for treatment $= t-1$

or, $3 = t-1$

$\therefore t = 4$

Now

d.f for Error $(r-1)(t-1) = 4 \times 3 = 12$

d.f for total $(rt-1) = 5 \times 4 - 1 = 19$

For MSSE, $MSSE = \frac{SSE}{d.f}$

or, $2.5 = \frac{SSE}{12}$

$\therefore SSE = 30$

We know

$TSS = SSB + SST + SSE$

or, $85.3 = 26.8 + SST + 30$

$\Rightarrow SST = 28.5$

For MSSB

$MSSB = \frac{SSB}{d.f} = \frac{26.8}{4} = 6.7$

For MSST

$MSST = \frac{SST}{d.f} = \frac{28.5}{3} = 9.5$

Now

$F_B = \frac{MSSB}{MSSE} = \frac{6.7}{2.5} =$

$F_T = \frac{MSST}{MSSE} = \frac{9.5}{2.5} =$

Scanned by CamScanner

Sol/

| S.V | d.f | S.S | M.S.S | F |
|---|---|---|---|---|
| Rows | ? | 72 | ? | 2 |
| Columns | ? | ? | 36 | ? |
| Treatment | ? | 180 | ? | ? |
| Error | 6 | ? | 12 | |
| Total | ? | ? | | |

d.f of error $= 6$

$(m-1)(m-2) = 3 \times 2$

$\therefore m-1 = 3$

$m = 4$

d.f of Rows $= 4-1 = 3$

" " Colum $= 4-1 = 3$

Treat $= 4-1 = 3$

total $= m^2 - 1 = 15$

~~SSR~~ Here,

$SSR = 72$

$C.F = \frac{G^2}{N} =$

$MSSR = \frac{SSR}{m-1} = \frac{72}{4-1} = 24$

$MSST = \frac{SST}{m-1}$

$= \frac{180}{3} = 60$

$MSSC = \frac{SSC}{m-1}$

or, $36 = \frac{SSC}{3}$

$\therefore SSC = 108$

$MSSE = \frac{SSE}{(m-1)(m-2)}$

or, $12 = \frac{MSSE}{6}$

$\therefore MSSE = 72$

$F_R = \frac{MSSR}{MSSE} = \frac{24}{72} = 0.333$

$F_C = \frac{MSSC}{MSSE} = \frac{36}{72} = 0.5$

$F_T = \frac{MSST}{MSSE} = \frac{60}{72} = 0.833$

For TSS

$SSE = TSS - SSR - SSC - SST$

$72 = TSS - 72 - 108 - 180$

$\therefore TSS = 288$

Scanned by CamScanner

# STOCHASTIC PROCESS

A family of random variables indexed by time as parameter is called stochastic process. It is also called chance or random process. For eg:- Markov process, Binomial process, Quotient process, Branching process, Random walk, Queue in system.

## TRANSITION PROBABILITY

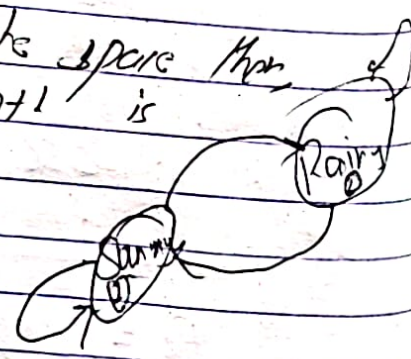
The probability of changing from one state to another or remain in the same state in a unit period of time is called transition probability. Since, probability of moving from one state to another depends upon the probability of preceding state so it is a conditional probability.

Transition probability matrix

It is matrix obtain by using various transition probability. It is always a square matrix.

Let $I = \{0, 1, 2, \ldots n\}$ be state space then transition prob matrix of order $n+1$ is

$$P = \begin{pmatrix} P_{00} & P_{01} & P_{02} & \cdots & P_{0n} \\ P_{10} & P_{11} & P_{12} & \cdots & P_{1n} \\ \vdots & & & & \vdots \\ P_{n0} & P_{n1} & P_{n2} & \cdots & P_{nn} \end{pmatrix}$$




Here, $P_{ij}$ = Transition probability of changing from state 'i' to state 'j'.

$$\sum^{n} P_{ij} = 1, \quad P_{ij} \geq 0$$

Scanned by CamScanner

$P_{ij} = P\{X_{t+1} = j / X_t = i\}$ = First step Transition Prob

$P_{ij}(n) = P\{X_{t+1} = j / X_t = i\}$ = n-step

Steady State Prob

A limiting probability by $x = \lim_{n \to \infty} P_n(x)$ is called steady state prob.

For steady state prob :-

$$xP = x, \quad \sum_x x_x = 1$$

It gives prob of a state in a long-run.

Binomial Process

It is descrete time/space counting process

Notation :-

$\lambda$ = Arrival rate, $\Delta$ = frame size

$P$ = Prob. of success [Prob. of arrival during one frame]

$X(t/\Delta)$ = Number of arrival in time $t$

$T$ = Inter arrival time or Time between two successive success

Computational Formulae

$P = \lambda \cdot \Delta$

$n = t/\Delta$

$Y$ = No. of frames [which follows geometric dist

$T = Y \cdot \Delta$

$E(T) = E(Y\Delta) = \Delta E(Y) = \Delta \cdot \frac{1}{P} = 1/\lambda \quad \left[\lambda = \frac{P}{\Delta}\right]$

$V(T) = V(Y\Delta) = \Delta^2 \cdot V(Y) = \Delta^2 \cdot \frac{1-P}{P^2} = \frac{1-P}{\lambda^2}$

Scanned by CamScanner

11) Sol$^n$ By ques

Here transition prob. matrix

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

For steady state,

$$\pi P = \pi$$

let $\pi = (\pi_1, \pi_2 \ \pi_3)$

Now

$$(\pi_1 \ \pi_2 \ \pi_3) \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1/2 & 1/2 & 0 \end{pmatrix} = (\pi_1 \ \pi_2 \ \pi_3)$$

$$\therefore \left( \pi_2 + \frac{\pi_3}{2} \quad \frac{\pi_1 + \pi_3}{2} \quad \frac{\pi_1}{2} \right) = (\pi_1 \ \pi_2 \ \pi_3)$$

By def

$$\frac{\pi_1}{2} = \pi_3 \quad \text{— (1)} \qquad \frac{\pi_1 + \pi_3}{2} = \pi_2 \Rightarrow \pi_2 = \frac{3\pi_3}{2}$$

Again, $\pi_1 + \pi_2 + \pi_3 = 1$

$\therefore \pi_3 = \frac{2}{9}$

From 1

$$\frac{\pi_1}{2} = \frac{2}{9} \Rightarrow \pi_1 = \frac{4}{9}$$

From 3 $\pi_2 = \frac{3}{2} \times \frac{2}{9} = \frac{1}{3}$

Hence steady state prob.

For state A $\pi_1 = 4/9$

B $\pi_2 = 1/3$

C $\pi_3 = 2/9$

Ans

Here we have, Arrival rate $(\lambda) = 2$ takes per minute

Frame size $(\Delta) = 4$ sec $= \frac{4}{60} = \frac{1}{15}$ min

i) For 8 min

$n = \frac{t}{\Delta} = \frac{8}{1/15} = 120$

We know

$p = \lambda\Delta = 2 \times \frac{1}{15} = \frac{2}{15}$

$q = 1 - \frac{2}{15} = \frac{13}{15}$

Now, reqd prob $= P(X_{(n)} > 3)$

$= 1 - P(X_{(n)} \le 3)$

$= 1 - [P(3) + P(2) + P(1) + P(0)]$ $\because X_{(n)} \sim B(n, p)$

$= 1 - [{}^{120}C_3 \times \left(\frac{2}{15}\right)^3 \times \left(\frac{13}{15}\right)^{117} + {}^{120}C_2 \left(\frac{2}{15}\right)^2 \left(\frac{13}{15}\right)^{118} +$

${}^{120}C_1 \times \frac{2}{15} \times \left(\frac{13}{15}\right)^{119} + \left(\frac{13}{15}\right)^{120}$

$= 0.9995$

ii) For 150 min

Sol/ Here we have

$\lambda = 3$ jobs per min

For Binomial counting process

i) $p = 0.2$, $\Delta = ?$

We have, $p = \lambda \cdot \Delta \Rightarrow \Delta = \frac{p}{\lambda} = \frac{0.2}{3} = \frac{1}{15}$ min

ii) For one minute

$E(T) = \frac{1}{\lambda} \Rightarrow E(T) = \frac{1}{3}$

$V(T) = \frac{1-p}{\lambda^2} = \frac{1-0.2}{3^2} = \frac{0.8}{3^2} = \frac{8}{90}$

$\Rightarrow S.D(T) = 0.298$

For one hr

$n = \frac{t}{\Delta} = \frac{60}{1/15} = 900$

Mean $= np = 900 \times 0.2 = 180$

variance $= npq = 900 \times 0.2 \times 0.8 = 144$

$S.D = \sqrt{144} = 12$

Scanned by CamScanner

15) Soln. Here we have

For binomial process

Frame size $(\Delta) = 2$ secs $= \frac{2}{60} = \frac{1}{30}$ min

Arrival rate $(\lambda) = 10$ per hour $= \frac{10}{60} = \frac{1}{6}$ per min

We know

$$p = \lambda\Delta = \frac{1}{30} \times \frac{1}{6} = \frac{1}{180}, \quad q = 1 - \frac{1}{180} = \frac{179}{180}$$

For 15 min $\quad n = \frac{6}{\Delta} = \frac{15}{1/30} = 450$

Now, Reqd Prob $= P(X(n) \geq 3)$

$$= 1 - P(X(n) < 3)$$

$$= P - [P(2) + P(1) + P(0)]$$

$$= 1 - \left[{}^{n}C_2\, p^2 q^{n-2} + {}^{n}C_1\, p^1 q^{n-1} + {}^{n}C_0\, p^0 q^{n-0}\right]$$

$$= 1 - \left[{}^{450}C_2 \cdot \left(\frac{1}{180}\right)^2 \times \left(\frac{179}{180}\right)^{448} + {}^{450}C_1 \times \frac{1}{180} \times \left(\frac{179}{180}\right)^{449} + \left(\frac{179}{180}\right)^{450}\right]$$

$\therefore$ Reqd prob =

Scanned by CamScanner

# POISSON PROCESS

It is a limiting case of binomial process. In binomial counting process if frame size $\Delta$ decreases towards zero and arrival rate $\lambda$, remain constant then we use Poisson process.

Let $X(t)$ = Number of arrivals occurring until time '$t$'.

$T$ = Inter arrival time

Here $X_t \sim P(\lambda t)$

& $T \sim$ Exponential $(\lambda)$

Here $E(X(t)) = np = p \cdot \frac{t}{\Delta} = \frac{p}{\Delta} \cdot t = \lambda t = V(X(t))$

20) Here we have

$\lambda = 5$ per 30 days $= \frac{5}{30}$ per day

a) For 15 days

Avg Arr rate $(\lambda t) = 15 \times \frac{5}{30} = 2.5$

Reqd prob $= P(X_t > 5)$

$= 1 - P(X_t \leq 5)$

$= 1 - (P(5) + \cdots + P(0))$

$= 1 - e^{-\lambda}$

$= 0.042$

For seven days

$\lambda t = 7 \times \frac{5}{30} = \frac{7}{6}$

Reqd prob $=$

$P(X_t = 0)$

$= e^{-\lambda} \frac{\lambda^0}{0!} = e^{-7/6} = 0.311$

Scanned by CamScanner

# Queuing System

It is a system design to perform certain task or process certain jobs by one or several servers. In a queue jobs are waiting to be processed.

Features of queue

1.) Arrival process :-
Arrival rate follows poisson distribution with parameter lamda ($\lambda$).

2.) Service process :-
Service rate follows exponential distribution with parameter ($\mu$).

3.) Queueing configuration
Queue is single waiting line with unlimited space (until stated)

4.) Queue discipline
It is based on 1st come 1st serve (FCFS)

## BERNOULLE SINGLE SERVER QUEUEING PROCESS

It is descrete time in process in one serve and unlimited capacity.

Let, $P_D$ = Prob of new arrival in the process
$P_S$ = Prob of departure from the services

Then



Scanned by CamScanner

$P_A = \lambda \cdot \Delta$

$P_S = \mu \Delta$

where $\Delta$ = frame size

Here,

$P_{00} = P(\text{no arrivals}) = 1 - P_A$

$P_{01} = P(\text{one arrival}) = P_A$

$P_{10} = P(\text{no arrival} \cap \text{one departure}) = (1-P_A).P_S$

$P_{11} = P(\text{one arrival} \cap \text{one departure})$ or $P(\text{no arrival} \cap \text{no departure}) = (1-P_A)(1-P_S) + P_A.P_S$

$P_{12} = P(\text{One arrival} \cap \text{no departure}) = P_A.(1-P_S)$

Now, transition prob matrix

$$P = \begin{pmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ P_{20} & P_{21} & P_{22} & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

$$\therefore P = \begin{pmatrix} 1-P_A & P_A & 0 & 0 & \cdots \\ (1-P_A)P_S & (1-P_A)(1-P_S) + P_A P_S & P_A(1-P_S) & 0 & \cdots \\ 0 & (1-P_A).P_S & (1-P_A)(1-P_S) + P_A P_S & P_A(1-P_S) & \cdots \end{pmatrix}$$

Now, transition prob. matrix

$$P = \begin{pmatrix} 1-P_A & P_A & 0 & 0 \\ (1-P_A).P_S & (1-P_A)(1-P_S) + P_A P_S & P_A(1-P_S) & 0 \\ & & (1-P_A)(1-P_S) + P_A P_S & P_A(1-P_S) \end{pmatrix}$$

$$P = \begin{pmatrix} 0.9825 & 0.0375 & 0 & 0 & \cdots \\ 0.08 & 0.8854 & 0.0346 & 0 & \cdots \\ 0 & 0.08 & 0.8854 & 0.036 & \cdots \\ 0 & 0 & 0.08 & 0.8854 & 0.036 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Scanned by CamScanner

24) Sol$^n$

Here, we have

$\lambda = 9$ per hour

$\mu = \frac{1}{3} \times 60 = 20$ per hour

$\Delta = 15$ secs $= \frac{15}{60 \times 60} = \frac{1}{240}$ hr

We know

$P_A = \lambda \cdot \Delta = 9 \times \frac{1}{240} = \frac{3}{80}$, $1 - P_A = \frac{77}{80}$

$P_S = \mu \cdot \Delta = 20 \times \frac{1}{240} = \frac{1}{12}$, $1 - P_S = \frac{11}{12}$

Now, transition prob. matrix

$$P = \begin{pmatrix} 0.9625 & 0.0375 & 0 & 0 & \cdots \\ 0.08 & 0.8854 & 0.0346 & 0 & \cdots \\ 0 & 0.08 & 0.8854 & 0.0346 & \cdots \\ 0 & 0 & 0.08 & 0.8854 & 0.0346 \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

25.) Sol$^n$

Here we have

$\lambda = 10$ per hour

$\mu = \frac{1}{4} \times 60 = 15$ per hour

$\Delta = 1$ minute $= \frac{1}{60}$ hr

We know,

$P_A = \lambda \cdot \Delta = 10 \times 1/60 = 1/6$, $1 - P_A = 5/6$.

$P_S = \mu \cdot \Delta = 15 \times 1/60 = 1/4$, $1 - P_S = 3/4$

Scanned by CamScanner

Now, transition prob. matrix

$$P = \begin{pmatrix} 1-P_A & P_A & 0 \\ (1-P_A)P_S & (1-P_A)(1-P_S) + P_A P_S & P_A(1-P_S) \\ 0 & (1-P_A)P_S & (1-P_S) + P_A P_S \end{pmatrix}$$

Here, $P = \begin{pmatrix} 5/6 & 1/6 & 0 \\ 5/24 & 16/24 & 3/24 \\ 0 & 5/24 & 19/24 \end{pmatrix}$

For steady state prob

$$\lambda P = \lambda$$

or $(\lambda_1 \ \lambda_2 \ \lambda_3) \begin{pmatrix} 5/6 & 1/6 & 0 \\ 5/24 & 16/24 & 3/24 \\ 0 & 5/24 & 19/24 \end{pmatrix} = (\lambda_1, \lambda_2 \ \lambda_3)$

or, $\left(\frac{5}{6}\lambda_1 + \frac{5}{24}\lambda_2 \quad \frac{1}{6}\lambda_1 + \frac{16\lambda_2}{24} + \frac{5\lambda_3}{24} \quad \frac{3\lambda_2}{24} + \frac{19\lambda_3}{24}\right)$

$= (\lambda_1 \ \lambda_2 \ \lambda_3)$

$\Rightarrow \frac{5}{6}\lambda_1 + \frac{5}{24}\lambda_2 = \lambda_1$

or, $\frac{1}{6}\lambda_1 = \frac{5}{24}\lambda_2$

$\therefore \lambda_1 = \frac{5}{4}\lambda_2$ —— (i)

$\Rightarrow \frac{3}{24}\lambda_2 + \frac{15\lambda_3}{24} = \lambda_3$

or, $\frac{9}{24}\lambda_3 = \frac{3}{24}\lambda_2$

$\therefore \lambda_3 = \frac{3}{9}\lambda_2$ —— (2)

Now $\lambda_1 + \lambda_2 + \lambda_3 = 1$

or, $\frac{5}{4}\lambda_2 + \lambda_2 + \frac{3}{5}\lambda_2 = 1$

$\therefore \lambda_2 = \frac{20}{57} = 0.3509$

$\lambda_1 = \frac{5}{4} \times \frac{20}{57} = 0.4386$

$\lambda_3 = \frac{3}{4} \times \frac{20}{57} = 0.2105$

Scanned by CamScanner

Hence, reqd steady state prob
For state 1, $\lambda_1 = 0.4386$
2, $\lambda_2 = 0.3509$
3, $\lambda_3 = 0.2105$

It is limiting case of Bernoullie queue process for negligible prob.
It is continuous time of Markov process with one server, unlimited capacity, Poisson interarrival rate ($\lambda$), exponential service rate ($\mu$).

Notations
$\lambda$ = Arrival rate
$\mu$ = service rate
$\rho$ = Utilization ratio $= \frac{\lambda}{\mu}$

$L_s$ = Mean no. of customer in the system
$L_q$ = " " " " " " queue
$W_s$ = Mean waiting time in the system
$W_q$ = " " " " " queue

Computational Formula
Utilization rate $= \frac{\lambda}{\mu} = \rho =$ prob of server is busy

Idle rate $= 1 - \rho$
Prob of no customer in the queue $(P_0) = 1 - \rho$
" if one " " " " $= (P_1) = \rho(1-\rho)$

Scanned by CamScanner

" " two " " " " $(P_2) = \rho^2 (1-\rho)$

" " n-customer " " " " $(P_n) = \rho^n (1-\rho)$

Expected or Av no of customes in the system $(L_s) = \frac{\rho}{1-\rho}$

Exp or Av no of customer in the queue $(L_q) = \frac{\rho^2}{1-\rho}$

Exp waiting time of a customer in queue $(W_q) = \frac{L_q}{\lambda} = \frac{\rho^2}{\lambda(1-\rho)}$

Exp waiting time of a customer in system $(W_s) = \frac{L_s}{\lambda} = \frac{\rho}{\lambda(1-\rho)}$

Prob of at least k-customer in the system,

$P(n \geq k) = 1-\rho^k$

~~Exp~~ Prob of less than k-customer in the system

$P(n < k) = \rho^k$

Scanned by CamScanner

26. Sol/

Here, we have

$\lambda = \frac{1}{6}$ per minute

$\mu = \frac{1}{4}$ per minute

i) Here, we have to compute ($W_s$)

We know, $W_s = \frac{L_s}{\lambda} = \frac{\rho}{\lambda(1-\rho)} = \frac{2/3}{\frac{1}{6}(1-2/3)} = \frac{2}{3} \times 18$

$\therefore W_s = 12$ minutes

ii) Here reqd fraction of time $= \rho = \frac{2}{3}$

28) Sol/ Here we have

$\lambda = \frac{1}{10}$ per minute

$\mu = \frac{1}{3}$ per minute, $\rho = \frac{\lambda}{\mu} = \frac{3}{10}$

(b) Here, we should compute ($P_0$).

$P_0 = 1-\rho = 1 - \frac{3}{10} = 0.7$

(a) Here we should compute ($L_s$)

$L_s = \frac{\rho}{1-\rho} = \frac{3/10}{1-3/10} = \frac{9}{10}$

(c) Here we

Here we have

$\mu = \frac{1}{3}$ per minute

$W_q = 5$ mins, $\lambda = ?$

We know,

$$W_s = \frac{\rho}{\lambda(1-\rho)} = 1$$

$$\text{or, } 5 = \frac{\lambda/\mu}{\lambda(1-\lambda/\mu)}$$

$$\text{or, } 5 = \frac{\lambda}{\lambda/\mu - \lambda)}$$

$$\text{or } 5 = \frac{1}{\mu - \lambda}$$

$$\text{or } 5 = \frac{1}{1/3 - \lambda}$$

$$\text{or, } \frac{1}{3} - \lambda = \frac{1}{5}$$

$$\text{or } \lambda = \frac{1}{3} - \frac{1}{5} = \frac{2}{15}$$

$$W_q = \frac{\rho^2}{\lambda(1-\rho)}$$

$$= \frac{\lambda^2/\mu^2}{\lambda(1-\lambda/\mu)}$$

$$\text{or, } 5 = \frac{\lambda^2}{\mu^2} \times \frac{\mu}{\lambda(1-\lambda)}$$

$$\text{or, } 5 = \frac{\lambda}{\mu(\mu-\lambda)}$$

$$\text{or, } 5 \times \mu(\mu-\lambda) = \lambda$$

$$\text{or } 5 \times \frac{1}{3}\left(\frac{1}{3} - \lambda\right) = \lambda$$

$$\text{or } \frac{5}{9} - \frac{5}{3}\lambda = \lambda$$

$$\text{or } \lambda = \frac{5}{24} \text{ per minute}$$

$$\text{or } \lambda = \frac{5}{24} \times 60 \text{ per hour}$$

$$\therefore \lambda = 12.5 \text{ per hr.}$$

Scanned by CamScanner

32.) Sol$^n$ Here, we have

$\lambda = \frac{1}{10}$ per ~~ta~~ min

$\mu = \frac{1}{6}$ per min

$\Rightarrow \rho = \frac{\lambda}{\mu} = \frac{6}{10} = 0.6$

i) We know

$L_q = \frac{\rho^2}{1-\rho} = \frac{0.6^2}{1-0.6}$

$\therefore L_q = 0.9$

ii) We know

$W_s = \frac{L_s}{\lambda} = \frac{\rho}{\lambda(1-\rho)}$

$= \frac{0.6}{\frac{1}{10}(1-0.6)} = \frac{6}{0.4}$

$\therefore W_s = 15$ mins

iii) We know

$P_2 = \rho^2(1-\rho) = 0.6^2(1-0.6)$

$\therefore P_2 = 0.144$

Scanned by CamScanner

$$\frac{30}{60} = \frac{1}{2}$$

$$\rho = \frac{1/2}{3/2} = \frac{1}{3}$$

$$\frac{\rho}{\lambda(1-\rho)} = 4.5$$

$$1 + \frac{1}{2}$$

$$\frac{2\rho}{(1-\rho)} = 4.5$$

$$\frac{3}{2}$$

$$\frac{\rho}{(1-\rho)} = 2.25$$

$$\rho = 2.25 - 2.25\rho$$

$$3.25\rho = 2.25$$

$$\rho = \frac{2.25}{3.25} = 0.69$$

$$\frac{9}{13}$$

$$\rho = \frac{\lambda}{\mu}$$

$$\frac{9}{13} = \frac{1/2}{\mu}$$

$$\mu = \frac{1}{2} \times \frac{13}{9}$$

$$= \frac{13}{18}$$

Scanned by CamScanner