1

Statistics 15-9-14 ①

The branch of science which deals with, to collect the data, organize, analyze and interpret the data.

Population:-

"Population is a group of items, units or subjects which is under reference of study."

Population may consist of finite or infinite number of units.

Types of population

Population can be classified into four catagories.

(i) Finite (ii) Infinite (iii) Real Population (iv) Hypothetical population.

(i) Finite Population

If the number of units constiting the population is fixed and limmited.

eg:- Students in a college.

Workers in a factory.

(ii) Infinit Population

If the population consists of an infinite number of items.

e.g. Population of stars in the sky

(iii) Real Population

A population consisting of the items which are all present physically is termed as real population.

(IV) Hypothetical Population

The population consists of the result of repeated trails hypothetical population. head or Trail.

Sample:-

A representative part of population is called sample.

Statistic:-

A numerical quantity computed from sample is called statistic.

Parameter:-

A numerical quantity computed from population is called parameter.

Sampling Unit

Individual member of the population is called sampling unit.

e.g.- Avg income per family.

Avg yeild of wheat.

Sampling Method:-

It is manner or scheme through which the required number of units are selected in a sample from a population.

Objective of Sampling:-

The purpose of sampling is to gather maximum information about the population under consideration at minimum cost, time and human power.

Sampling frame

A list or a map identifying each sampling unit by a number.

* Difference b/w complete enumeration and sampling study:-

In complete enumeration each and every unit of population is studied and results are based on all units of the population whereas in a sampling study only a selected number of units are studied and results are based on the data of these units are supposed to yeild information about the whole population.

→ Any value or formula with the help which we can find a single value.

Estimator:

An estimator is a rule or a function of variates for estimating the population parameters.

Estimate:–

A particular value of an estimator from a fixed set of values of a random sample is known an estimate.

Types of Sampling Schemes:–

There are two types of sampling schemes.

i) Unrestricted random Sampling.

(II) Restricted random Sampling.

⇒ Unrestricted Random Sampling

In this type of sampling each and every unit of population has equal chance of being included in the sample.

Simple random sampling is an example of unrestricted.

⇒ Restricted Random Sampling.

If an investigator has any idea about hetrogenity of sampling units, the population is divided into homogenous groups and sample is drawn independently from each group

Stratified sampling, systematic, multistage Sampling. There are two types to select sample.

Sampling with Replacement

Sampling is said to be with replacement when from a finite population a sampling unite is drawn, observed and returned to the population before another unit is drawn. Formula: N^n

Sampling without Replacement:

Sampling unit is choosen and not returned to the population after it has been observed, the sampling is said to be without replacement.

$$^{N}C_{n} = \frac{N!}{(N-n)!\, n!}$$

Catagories of Sampling Method

There are two types of sampling method:

(i):- Probability Sampling.
ii):- Non-Probability Sampling.

Probability Sampling

The sampling procedure in which each unit (item) of the population gets a definite probability of being included in the sample is said to be probability Sampling.

(i):- Simple Random Sampling
(ii):- Statitied random Sampling
(iii):- Systematic Sampling
(iv):- Cluster Sampling.

Non-Probability Sampling

A non-probability sampling also called random sampling is a process in which the personal Judgement determines which units of the population are selected for a sample.

(i):- Purposive Sampling
(ii):- Quota Sampling

Errors:-

There are two types of error:-

(i):- Sampling error (Random Error)
(ii):- Non-Sampling error (Non-Random Error)

Sampling Error

The difference b/w value of sample statistic and the true value of the corresponding population parameter.

Sampling Error = $\bar{x} - \mu$.

As the sampling size increase the sampling error is reduced.

Non-Sampling Error.

The errors which occur at the stage of gathering & processing of data. Non-sampling errors include all kinds of human errors faulty sampling frame, biased method of selection of units, errors of observations and measurement.

Sampling Distribution 17-9-14 (2)

A sampling distribution is defined as a probability distribution of the values of a statistic such as a mean, a standard deviation, a proportion etc, computed from all possible samples of the same size, which might be selected with or without replacement from a population.

Probability distribution.

x	f	$f(x)$
0	1	1/12
1	4	4/12
2	3	3/12
3	2	2/12
4	1	1/12
5	1	1/12

$\Sigma f = 12$

Sampling Distribution for $\bar{x}$ (Mean Sample)

(i) Sampling with Replacement N^n

(ii) Sampling without Replacement ${}^{N}C_{n}$

Q: $N = 3, 4, 6, 8, 10$ 9.14.16

Size $n = 2$, $N = 5$

Sampling with Replacement $= N^n = 5^2 = 25$

$\frac{25}{5} = 5$, $\frac{5}{5} = 1$

S.No.	Sample	$\bar{x}$	S.No.	Sample	$\bar{x}$
1	3, 3	$\frac{3+3}{2} = 3$	14	6, 8	$14/2 = 7$
2	3, 4	$7/2 = 3.5$	15	6, 10	$16/2 = 8$
3	3, 6	$9/2 = 4.5$	16	8, 3	$11/2 = 5.5$
4	3, 8	$11/2 = 5.5$	17	8, 4	$12/2 = 6$
5	3, 10	$13/2 = 6.5$	18	8, 6	$14/2 = 7$
6	4, 3	$7/2 = 3.5$	19	8, 8	$16/2 = 8$
7	4, 4	$8/2 = 4$	20	8, 10	$18/2 = 9$
8	4, 6	$10/2 = 5$	21	10, 3	$13/2 = 6.5$
9	4, 8	$12/2 = 6$	22	10, 4	$14/2 = 7$
10	4, 10	$14/2 = 7$	23	10, 6	$16/2 = 8$
11	6, 3	$9/2 = 4.5$	24	10, 8	$18/2 = 9$
12	6, 4	$10/2 = 5$	25	10, 10	$20/2 = 10$
13	6, 6	$12/2 = 6$			

Sampling distribution of $\bar{x}$

$\bar{x}$	f	$f(\bar{x})$	$\bar{x} f(\bar{x})$	$\bar{x}^2 f(\bar{x})$
3	1	1/25	3/25	9/25
3.5	2	2/25	7/25	24.5/25
4	1	1/25	4/25	16/25
4.5	2	2/25	9/25	40.5/25
5	2	2/25	10/25	50/25
5.5	2	2/25	11/25	60.5/25
6	3	3/25	18/25	108/25
6.5	2	2/25	13/25	84.5/25
7	4	4/25	28/25	196/25
8	3	3/25	24/25	192/25
9	2	2/25	18/25	162/25
10	1	1/25	10/25	100/25
			$\sum \bar{x} f(\bar{x}) = 6.2$	$\sum \bar{x}^2 f(\bar{x}) = \frac{1043}{25} = 41.72$

$$\mu(\bar{x}) = \sum \bar{x} f(\bar{x}) = 6.2$$

$$\sigma^2(\bar{x}) = \sum \bar{x}^2 f(\bar{x}) - \left[\sum \bar{x} f(\bar{x})\right]^2$$

$$= 41.72 - (6.2)^2$$

$$= 3.28$$

$$\mu = \frac{\sum X}{N} = \frac{3+4+6+8+10}{5} = 31/5 = 6.2$$

$$\sigma^2 = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2 = \frac{9+16+36+64+100}{5} - (6.2)^2$$

$S^2 = 45 - 38.44 = 6.56$

To check:

$$S^2(\bar{x}) = \frac{S^2}{n}$$

$$3.28 = \frac{6.56}{2}$$

$$3.28 = 3.28$$

Sampling without Replacing.

Q:- $N = 2, 4, 6, 6, 8, 10$

$n = 3$

$${}^{N}C_{n} = {}^{6}C_{3} = 20$$

S.No.	Sample $\bar{X}$	$\bar{x}$	S.No.	Samples $\bar{X}$	$\bar{x}$
1	2, 4, 6	4	11	4, 6, 6	16/3
2	2, 4, 6	4	12	4, 6, 8	6
3	2, 4, 8	14/3	13	4, 6, 10	20/3
4	2, 4, 10	16/3	14	4, 6, 8	6
5	2, 6, 6	14/3	15	4, 6, 10	20/3
6	2, 6, 8	16/3	16	4, 8, 10	22/3
7	2, 6, 10	6	17	6, 6, 8	20/3
8	2, 6, 8	16/3	18	6, 6, 10	22/3
9	2, 6, 10	6	19	6, 8, 10	8
10	2, 8, 10	20/3	20	6, 8, 10	8

Sampling distribution of $\bar{x}$

$\bar{x}$	f	$f(\bar{x})$	$\bar{x}f(\bar{x})$	$\bar{x}^2 f(\bar{x})$
4	2	1/10	$\frac{2}{5}=\frac{6}{15}$	$\frac{8}{5}=1.6$
14/3	2	1/10	7/15	$\frac{98}{45}=2.18$
16/3	4	1/5	16/15	$\frac{256}{45}=5.69$
6	4	1/5	$\frac{6}{5}=\frac{18}{15}$	$\frac{36}{5}=7.2$
20/3	4	1/5	20/15	$\frac{80}{9}=8.89$
22/3	2	1/10	11/15	$\frac{242}{45}=5.37$
8	2	1/10	$\frac{4}{15}=\frac{12}{15}$	$\frac{32}{5}=6.4$

$\Sigma \bar{x}f(\bar{x}) = \frac{90}{15} = 6$ $\quad \Sigma \bar{x}^2 f(\bar{x}) = 37.33$

$$\mu(\bar{x}) = \Sigma\, \bar{x} f(\bar{x}) = 6$$

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

$$= 37.33 - (6)^2$$

$$= 1.33$$

$$\mu = \frac{\Sigma X}{N} = \frac{2+4+6+6+8+10}{6} = \frac{36}{6} = 6$$

$$\sigma^2 = \frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2$$

$$= \frac{4+16+36+36+64+100}{6} - 36$$

$$= 42.667 - 36 = 6.667$$

To check:- $\quad \sigma^2(\bar{x}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

$$1.33 = \frac{6.667}{3}\cdot\left(\frac{6-3}{6-1}\right) = \frac{6.667}{5} = 1.33$$

Q:-

$N = 6, 6, 10, 12, 15, 18$

$n = 4$

Without Replacement $= {}^{N}C_{n}$

$= {}^{6}C_{4} = 15$

S·N	Samples $\bar{X}$	$\bar{x}$	S·N	Samples $\bar{X}$	$\bar{x}$
1	6, 6, 10, 12	8·5	9	6, 10, 15, 18	12·25
2	6, 6, 10, 15	9·25	10	6, 12, 15, 18	12·75
3	6, 6, 10, 18	10	11	6, 10, 12, 15	10·75
4	6, 6, 12, 15	9·75	12	6, 10, 12, 18	11·5
5	6, 6, 12, 18	10·5	13	6, 10, 15, 18	12·25
6	6, 6, 15, 18.	11·25	14	6, 12, 15, 18	12·75
7	6, 10, 12, 15	10·75	15	10, 12, 15, 18	13·75
8	6, 10, 12, 18	11·5			

$\bar{x}$	f	$f(\bar{x})$	$\bar{x}f(\bar{x})$	$\bar{x}^2 f(\bar{x})$
8.5	1	$1/15$	$8.5/15 = 0.57$	$72.25/15 = 4.82$
9.25	1	$1/15$	$\frac{9.25}{15} = 0.62$	$85.56/15 = 5.70$
9.75	1	$1/15$	$\frac{9.75}{15} = 0.65$	$95.06/15 = 6.34$
10	1	$1/15$	$10/15 = 0.67$	$100/15 = 6.67$
10.5	1	$1/15$	$10.5/15 = 0.7$	$\frac{110.25}{15} = 7.35$
10.75	2	$2/15$	$21.5/15 = 1.43$	$\frac{231.12}{15} = 15.41$
11.25	1	$1/15$	$\frac{11.25}{15} = 0.75$	$\frac{126.56}{15} = 8.44$
11.5	2	$2/15$	$23/15 = 1.53$	$\frac{264.5}{15} = 17.63$
12.25	2	$2/15$	$24.5/15 = 1.63$	$\frac{300.13}{15} = 20.01$
12.75	2	$2/15$	$25.5/15 = 1.7$	$\frac{325.12}{15} = 21.67$
13.75	1	$1/15$	$13.75/15 = 0.92$	$\frac{189.06}{15} = 12.6$

$\Sigma \bar{x}f(\bar{x}) = 11.1667$ $\quad \Sigma \bar{x}^2 f(\bar{x}) = 126.64$

$$\mu(\bar{x}) = \Sigma \bar{x}f(\bar{x}) = 11.167$$

$$\delta^2_{(\bar{x})} = \Sigma \bar{x}^2 f(\bar{x}) - [\Sigma \bar{x} f(\bar{x})]^2$$

$$= 126.64 - (11.1667)^2$$

$$= 1.945$$

$$\mu = \frac{\Sigma X}{N} = \frac{6+6+10+12+15+18}{6} = 11.167$$

$$\delta^2 = \frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2$$

$$= \frac{36+36+100+144+225+324}{6} - 124.7$$

$$= 19.467$$

To check:-

$$S^2(\bar{x}) = \frac{S^2}{n}\left(\frac{N-n}{N-1}\right)$$

$$1.945 = \frac{19.467}{4}\left(\frac{6-4}{6-1}\right)$$

$$1.945 = 1.946$$

3

Sampling distribution for Proportion

22-9-14

Q:- $N = 1, 3, 5, 6, 8$

$n = 3$

W.O.R $^NC_n = \binom{5}{3} = 10$

P = Propotion of even number.

S.No.	Samples	$\hat{p}$	S.No.	Samples	$\hat{p}$
1	1, 3, 5	$0/3 = 0$	6	1, 6, 8	$2/3$
2	1, 3, 6	$1/3$	7	3, 5, 6	$1/3$
3	1, 3, 8	$1/3$	8	3, 5, 8	$1/3$
4	1, 5, 6	$1/3$	9	3, 6, 8	$2/3$
5	1, 5, 8	$1/3$	10	5, 6, 8	$2/3$

$\hat{p}$	f	$f(\hat{p})$	$\hat{p}f(\hat{p})$	$\hat{p}^2 f(\hat{p})$
0	1	$1/10$	0	0
$1/3$	6	$6/10$	$2/10$	$2/30$
$2/3$	3	$3/10$	$2/10$	$4/30$
			$2/5$	$1/5$

$$\Sigma \hat{p}^2 \, q(\hat{p}) = 0.2$$

$$\mu(\hat{p}) = \Sigma \hat{p} \, q(\hat{p}) = 0.4$$

$$\delta^2(\hat{p}) = \Sigma \hat{p}^2 q(\hat{p}) - \left[\Sigma \hat{p} \, q(\hat{p})\right]^2$$

$$= 0.2 - (0.4)^2$$

$$= 0.04$$

$$P = \frac{X}{N} = \frac{2}{5} = 0.4$$

So, $\mu(\hat{p}) = P$

$$0.4 = 0.4$$

Now

$$\delta^2(\hat{p}) = \frac{PQ}{n}\left(\frac{N-n}{N-1}\right)$$

$$Q = 1 - P$$
$$= 1 - \frac{2}{5} = \frac{3}{5}$$

$$0.04 = \frac{\frac{2}{5} \cdot \frac{3}{5}}{3}\left(\frac{5-3}{5-1}\right)$$

$$= \frac{2}{25} \cdot \frac{2}{4}$$

$$0.04 = \frac{1}{25} = 0.04$$

Q:- ~~$N = 1, 2, 3, 4, 5, 6$~~

~~$n = 2$~~

~~With Replacement.~~

~~$\hat{p}$ = Propotion of even number.~~

~~(i):- $\mu(\hat{p}) = P$ (ii):- $\delta^2(\hat{p}) = \frac{PQ}{n}$~~

~~Answer on next pages.~~

Q.

$N = BBGBBG$

$n = 3$

W.O.R

${}^{N}C_{n} = {}^{6}C_{3} = 20$

For B

S.N	Samples	$\hat{p}$	S.N	Samples	$\hat{p}$
1	BBG	2/3	11	BGB	2/3
2	BBB	3/3=1	12	BGB	2/3
3	BBB	3/3=1	13	BGG	1/3
4	BBG	2/3	14	BBB	3/3=1
5	BGB	2/3	15	BBG	2/3
6	BGB	2/3	16	BBG	2/3
7	BGG	1/3	17	GBB	2/3
8.	BBB	3/3=1	18	GBG	1/3
9	BBG	2/3	19	GBG	1/3
10	BBG	2/3	20	BBG	2/3

$\hat{p}$	f	$f(\hat{p})$	$\hat{p}f(\hat{p})$	$\hat{p}^2 f(\hat{p})$
1/3	4	4/20 = 1/5	1/15	1/45
2/3	12	12/20 = 3/5	6/15	12/45
1	4	4/20 = 1/5	3/15	9/45
			10/15 = 2/3	22/45 = 0.489

$\mu(\hat{p}) = \sum \hat{p}\, f(\hat{p}) = 0.667$

$\delta^2(\hat{p}) = \sum \hat{p}^2 f(\hat{p}) - (\sum \hat{p}\, f(\hat{p}))^2$

$= 0.489 - (0.667)^2$

$= 0.0444$

$\mu(\hat{p}) = p$

$0.667 = \frac{X}{N} = 4/6$

$0.669 = 0.667$

Now

$\delta^2(\hat{p}) = \frac{pq}{n}\left(\frac{N-n}{N-1}\right)$ $\quad q = 1-p = 1 - \frac{2}{3} = 1/3$

$0.0444 = \frac{\frac{2}{3}\cdot\frac{1}{3}}{3}\left(\frac{6-3}{6-1}\right)$

$= \frac{2}{9\times\cancel{3}}\cdot\frac{\cancel{3}}{5}$

$= \frac{2}{45}$

$= 0.0444$

Solution to perivious question.

$N = 1, 2, 3, 4, 5, 6$

$n = 2$ W.R

$\hat{p}$ = propotion of even number

$N^n = 6^2 = 36$

S.N	Samples	$\hat{p}$	S.N	Samples	$\hat{p}$	S.N	Samples	$\hat{p}$
1	1,1	0/2	13	3,1	0/2	25	5,1	0/2
2	1,2	1/2	14	3,2	1/2	26	5,2	1/2
3	1,3	0/2	15	3,3	0/2	27	5,3	0/2
4	1,4	1/2	16	3,4	1/2	28	5,4	1/2
5	1,5	0/2	17	3,5	0/2	29	5,5	0/2
6	1,6	1/2	18	3,6	1/2	30	5,6	1/2
7	2,1	1/2	19	4,1	1/2	31	6,1	1/2
8	2,2	2/2	20	4,2	2/2	32	6,2	2/2
9	2,3	1/2	21	4,3	1/2	33	6,3	1/2
10	2,4	2/2	22	4,4	2/2	34	6,4	2/2
11	2,5	1/2	23	4,5	1/2	35	6,5	1/2
12	2,6	2/2	24	4,6	2/2	36	6,6	2/2

$\hat{p}$	f	$f(\hat{p})$	$\hat{p}\,f(\hat{p})$	$\hat{p}^2 f(\hat{p})$
0	9	$9/36 = 1/4$	0	0
1/2	18	$18/36 = 1/2$	1/4	1/8
1	9	$9/36 = 1/4$	1/4	1/4
			0.50	0.375

$$\mu(\hat{p}) = \sum \hat{p}\, f(\hat{p}) = 0.50$$

$$\sigma^2(\hat{p}) = \sum \hat{p}^2 f(\hat{p}) - \left[\sum \hat{p}\, f(\hat{p})\right]^2$$

$$= 0.375 - (0.5)^2$$

$$= 0.125$$

$$P = \frac{X}{N} = \frac{3}{6} = \frac{1}{2} = 0.5$$

So, $\mu(\hat{P}) = P$

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

$$\sigma^2(\hat{P}) = \frac{Pq}{n}$$

$$0.125 = \frac{\frac{1}{2}\cdot\frac{1}{2}}{2}$$

$$0.125 = \frac{1}{8}$$

$$0.125 = 0.125$$

Sampling Distribution of (4) 29-9-14

Difference Mean ($\bar{x}_1 - \bar{x}_2$ or $\bar{x}_2 - \bar{x}_1$)

Q:- $N_1 = 3, 4, 6$ $n_1 = 2$ W.R

$N_2 = 4, 6, 8$ $n_2 = 2$ W.R

S.No.	Samples	$\bar{x}_1$	S.N	Samples	$\bar{x}_1$
1	3,3	3	6	4,6	5
2	3,4	3.5	7	6,3	4.5
3	3,6	4.5	8	6,4	5
4	4,3	3.5	9	6,6	6
5	4,4	4			

S.N	Samples	$\bar{x}_2$	S.N	Samples	$\bar{x}_2$
1	4,4	4	6	6,8	7
2	4,6	5	7	8,4	6
3	4,8	6	8	8,6	7
4	6,4	5	9	8,8	8
5	6,6	6			

$\bar{x}_2 - \bar{x}_1$

$\bar{x}_1$ \ $\bar{x}_2$	4	5	6	5	6	7	6	7	8
3	1	2	3	2	3	4	3	4	5
3.5	0.5	1.5	2.5	1.5	2.5	3.5	2.5	3.5	4.5
4.5	-0.5	0.5	1.5	0.5	1.5	2.5	1.5	2.5	3.5
3.5	0.5	1.5	2.5	1.5	2.5	3.5	2.5	3.5	4.5
4	0	1	2	1	2	3	2	3	4
5	-1	0	1	0	1	2	1	2	3
4.5	-0.5	0.5	1.5	0.5	1.5	2.5	1.5	2.5	3.5
5	-1	0	1	0	1	2	1	2	3
6	-2	-1	0	-1	0	1	0	1	2

Let $\bar{x}_2 - \bar{x}_1 = d$

d	f	$f(d)$	$d\,f(d)$	$d^2 f(d)$
-2	1	1/81	-2/81	4/81
-1	4	4/81	-4/81	4/81
-0.5	2	2/81	-1/81	0.5/81
0	8	8/81	0	0
0.5	6	6/81	3/81	1.5/81
1	11	11/81	11/81	11/81
1.5	10	10/81	15/81	22.5/81
2	10	10/81	20/81	40/81
2.5	10	10/81	25/81	62.5/81
3	7	7/81	21/81	63/81
3.5	6	6/81	21/81	73.5/81
4	3	3/81	12/81	48/81
4.5	2	2/81	9/81	40.5/81
5	1	1/81	5/81	25/81
			$\frac{135}{81} = 1.667$	$\frac{396}{81} = 4.889$

$$\mu(d) = \sum d\,f(d) = 1.667$$

$$\bar{X}_2 = \frac{\sum \bar{X}_2}{N} = \frac{4+6+8}{3} = 6$$

$$\bar{X}_1 = \frac{\sum \bar{X}_1}{N} = \frac{3+4+6}{3} = 4.333$$

$$\bar{X}_2 - \bar{X}_1 = 6 - 4.333 = 1.667$$

$$\Rightarrow \mu(d) = \bar{X}_2 - \bar{X}_1$$

$$\delta^2(d) = \Sigma d^2 f(d) - [\Sigma d f(d)]^2$$
$$= 4.889 - (1.667)^2$$
$$= 2.110$$

$$\delta_1^2 = \frac{\Sigma X_1^2}{N} - \left(\frac{\Sigma X_1}{N}\right)^2$$
$$= \frac{9+16+36}{3} - \left(\frac{13}{3}\right)^2$$
$$= 20.333 - 18.777$$
$$= 1.556$$

$$\delta_2^2 = \frac{\Sigma X_2^2}{N} - \left(\frac{\Sigma X_2}{N}\right)^2$$
$$= \frac{16+36+64}{3} - \left(\frac{18}{3}\right)^2$$
$$= 38.667 - 36$$
$$= 2.667$$

Now

$$\delta^2(d) = \frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}$$
$$= \frac{1.556}{2} + \frac{2.667}{2}$$
$$= 0.778 + 1.3335$$
$$= 2.111$$

Q:-

$N_1 = 4, 6, 8, 10$ $\qquad n_1 = 3$

$N_2 = 8, 10, 12, 14$ $\qquad n_2 = 3$

W.O.R

$$^4C_3 = \frac{4!}{3!(4-3)!} = 4$$

S.N	Samples	$\bar{x}_1$	S.N	Samples	$\bar{x}_2$
1	4, 6, 8	6	1	8, 10, 12	10
2	4, 6, 10	6.67	2	8, 10, 14	10.67
3	4, 8, 10	7.33	3	8, 12, 14	11.33
4	6, 8, 10	8	4	10, 12, 14	12

$\bar{x}_2 - \bar{x}_1$

$\bar{x}_1 \backslash \bar{x}_2$	10	10.67	11.33	12
6	4	4.67	5.33	6
6.67	3.33	4	4.67	5.33
7.33	2.67	3.33	4	4.67
8	2	2.67	3.33	4

d	f	f(d)	d f(d)	d^2 f(d)
2	1	1/16	2/16	4/16
2.67	2	2/16	5.34/16	14.26/16
3.33	3	3/16	9.99/16	33.26/16
4	4	4/16	16/16	64/16
4.67	3	3/16	14.01/16	65.43/16
5.33	2	2/16	10.66/16	56.81/16
6	1	1/16	6/16	36/16
			$64/16 = 4$	$\frac{273.76}{16} = 17.11$

$$\mu d = \sum d f(d) = 4$$

$$\bar{X}_2 = \frac{\sum X_2}{N_2} = \frac{8+10+12+14}{4} = 44/4 = 11$$

$$\bar{X}_1 = \frac{\sum X_1}{N_1} = \frac{4+6+8+10}{4} = 28/4 = 7$$

$$\bar{X}_2 - \bar{X}_1 = 11 - 7 = 4$$

$$\sigma^2(d) = \sum d^2 f(d) - [\sum d f(d)]^2$$

$$= 17.11 - (4)^2$$

$$= 17.11 - 16$$

$$= 1.11$$

$$\sigma_1^2 = \frac{\sum X_1^2}{N_1} - \left(\frac{\sum X_1}{N_1}\right)^2 = \frac{16+36+64+100}{4} - (7)^2$$

$$= 5$$

$$\sigma_2^2 = \frac{64+100+144+196}{4} - (11)^2 = 504/4 - 121 = 5$$

$$S^2 d = (S_1^2 + S_2^2)\left[\frac{1}{n}\left(\frac{N-n}{N-1}\right)\right]$$

$$1.11 = \frac{5+5}{3}\cdot\left(\frac{4-3}{4-1}\right)$$

$$1.11 = \frac{10}{9} = 1.11$$

(5)

1/14/10 Sampling Dist. of diff. proportion $(\hat{P}_2 - \hat{P}_1)$

Q:-

$N_1 = 1, 2, 3, 4, 5$ $\quad n_1 = 3$ $\quad$ W.O.R

$N_2 = 1, 1, 2, 3, 4$ $\quad n_2 = 3$ $\quad$ "

Propotion is odd no.

$$^5C_3 = \frac{5!}{3!(5-3)!} = \frac{5\times4\times3!}{3!\times2} = 10$$

S.N	Samples	$\hat{P}_1$	S.N	Samples	$\hat{P}_1$
1	1,2,3	2/3	6	1,4,5	2/3
2	1,2,4	1/3	7	2,3,4	1/3
3	1,2,5	2/3	8	2,3,5	2/3
4	1,3,4	2/3	9	2,4,5	1/3
5	1,3,5	3/3	10	3,4,5	2/3

S.N	Samples	$\hat{P}_2$	S.N	Samples	$\hat{P}_2$
1	1,1,2	2/3	6	1,3,4	2/3
2	1,1,3	3/3	7	1,2,3	2/3
3	1,1,4	2/3	8	1,2,4	1/3
4	1,2,3	2/3	9	1,3,4	2/3
5	1,2,4	1/3	10	2,3,4	1/3

$\hat{P}_2 - \hat{P}_1$

$\hat{P}_1$ \ $\hat{P}_2$	2/3	3/3	2/3	2/3	1/3	2/3	2/3	1/3	2/3	1/3
2/3	0	1/3	0	0	-1/3	0	0	-1/3	0	-1/3
1/3	1/3	2/3	1/3	1/3	0	1/3	1/3	0	1/3	0
2/3	0	1/3	0	0	-1/3	0	0	-1/3	0	-1/3
2/3	0	1/3	0	0	-1/3	0	0	-1/3	0	-1/3
3/3	-1/3	0	-1/3	-1/3	-2/3	-1/3	-1/3	-2/3	-1/3	-2/3
2/3	0	1/3	0	0	-1/3	0	0	-1/3	0	-1/3
1/3	1/3	2/3	1/3	1/3	0	1/3	1/3	0	1/3	0
2/3	0	1/3	0	0	-1/3	0	0	-1/3	0	-1/3
1/3	1/3	2/3	1/3	1/3	0	1/3	1/3	0	1/3	0
2/3	0	1/3	0	0	-1/3	0	0	-1/3	0	-1/3

d	f	$f(d)$	$d f(d)$	$d^2 f(d)$
$-2/3$	3	$3/100$	$\frac{-1}{50} = -2/100$	$2/150 = \frac{1}{75}$
$-1/3$	24	$24/100$	$\frac{-4}{50} = -8/100$	$4/150 = \frac{2}{75}$
0	46	$46/100$	0	0
$1/3$	24	$24/100$	$\frac{4}{50} = 8/100$	$4/150 = \frac{2}{75}$
$2/3$	3	$3/100$	$\frac{1}{50} = 2/100$	$2/150 = \frac{1}{75}$
	100		0	$\frac{2}{25} = 12/150 = 0.08$

$\mu(d) = P_2 - P_1$

$P_2 = \frac{X_2}{N_2} = \frac{3}{5} = 0.6$

$P_1 = \frac{X_1}{N_1} = \frac{3}{5} = 0.6$

So,

$\mu(d) = P_2 - P_1$

$= 0.6 - 0.6$

$= 0$

$\delta^2(d) = \Sigma d^2 f(d) - [\Sigma d f(d)]^2$

$= 0.08 - (0)^2$

$= 0.08$

$q_1 = 1 - 0.6 = 0.4$

$q_2 = 1 - 0.6 = 0.4$

$\delta^2(d) = \frac{P_1 q_1}{n_1}\left(\frac{N_1 - n_1}{N_1 - 1}\right) + \frac{P_2 q_2}{n_2}\left(\frac{N_2 - n_2}{N_2 - 1}\right)$

$= \frac{0.6 \times 0.4}{3}\left(\frac{2}{4}\right) + \frac{0.6 \times 0.4}{3}\left(\frac{2}{4}\right)$

$$S^2(d) = \frac{0.24}{3} = 0.08$$

Q:- $N_1 = 1,2,3,4$ $n_1 = 2$ W.R

odd

$N_2 = 1,1,2,3$ $n_2 = 2$

S.N	Samples	$\hat{P}_1$	S.N	Samples	$\hat{P}_1$
1	1,1	2/2	9	3,1	2/2
2	1,2	1/2	10	3,2	1/2
3	1,3	2/2	11	3,3	2/2
4	1,4	1/2	12	3,4	1/2
5	2,1	1/2	13	4,1	1/2
6	2,2	0/2	14	4,2	0/2
7	2,3	1/2	15	4,3	1/2
8	2,4	0/2	16	4,4	0/2

S.N	Samples	$\hat{P}_2$	S.N.	Samples	$\hat{P}_2$
1	1,1	2/2	9	2,1	1/2
2	1,1	2/2	10	2,1	1/2
3	1,2	1/2	11	2,2	0/2
4	1,3	2/2	12	2,3	1/2
5	1,1	2/2	13	3,1	2/2
6	1,1	2/2	14	3,1	2/2
7	1,2	1/2	15	3,2	1/2
8	1,3	2/2	16	3,3	2/2

$\hat{P}_2 - \hat{P}_1$

$\hat{P}_1 \backslash \hat{P}_2$	2/2	2/2	1/2	2/2	2/2	2/2	1/2	2/2	1/2	1/2	0	1/2	2/2	2/2	1/2	2/2
2/2	0	0	-1/2	0	0	0	-1/2	0	-1/2	-1/2	-2/2	-1/2	0	0	-1/2	0
1/2	1/2	1/2	0	1/2	1/2	1/2	0	1/2	0	0	-1/2	0	1/2	1/2	0	1/2
2/2	0	0	-1/2	0	0	0	-1/2	0	-1/2	-1/2	-2/2	-1/2	0	0	-1/2	0
1/2	1/2	1/2	0	1/2	1/2	1/2	0	1/2	0	0	-1/2	0	1/2	1/2	0	1/2
1/2	1/2	1/2	0	1/2	1/2	1/2	0	1/2	0	0	-1/2	0	1/2	1/2	0	1/2
0	2/2	2/2	1/2	2/2	2/2	2/2	1/2	2/2	1/2	1/2	0	1/2	2/2	2/2	1/2	2/2
1/2	1/2	1/2	0	1/2	1/2	1/2	0	1/2	0	0	-1/2	0	1/2	1/2	0	1/2
0	2/2	2/2	1/2	2/2	2/2	2/2	1/2	2/2	1/2	1/2	0	1/2	2/2	2/2	1/2	2/2
1	0	0	-1/2	0	0	0	-1/2	0	-1/2	-1/2	-2/2	-1/2	0	0	-1/2	0
1/2	1/2	1/2	0	1/2	1/2	1/2	0	1/2	0	0	-1/2	0	1/2	1/2	0	1/2
1	0	0	-1/2	0	0	0	-1/2	0	-1/2	-1/2	-2/2	-1/2	0	0	-1/2	0
1/2	1/2	1/2	0	1/2	1/2	1/2	0	1/2	0	0	-1/2	0	1/2	1/2	0	1/2
1/2	1/2	1/2	0	1/2	1/2	1/2	0	1/2	0	0	-1/2	0	1/2	1/2	0	1/2
0	2/2	2/2	1/2	2/2	2/2	2/2	1/2	2/2	1/2	1/2	0	1/2	2/2	2/2	1/2	2/2
1/2	1/2	1/2	0	1/2	1/2	1/2	0	1/2	0	0	-1/2	0	1/2	1/2	0	1/2
0	2/2	2/2	1/2	2/2	2/2	2/2	1/2	2/2	1/2	1/2	0	1/2	2/2	2/2	1/2	2/2

d	f	$f(d)$	$d\,f(d)$	$d^2 f(d)$
-1	4	$1/64$	$-1/64$	$1/64$
$-1/2$	32	$8/64$	$-4/64$	$2/64$
0	88	$22/64$	0	0
$1/2$	96	$24/64$	$12/64$	$6/64$
1	36	$9/64$	$9/64$	$9/64$
	256	1	$\frac{1}{4} = \frac{16}{64} = 0.25$	$\frac{9}{32} = \frac{18}{64} = 0.28125$

$$\mu_d = \Sigma d\, f(d)$$
$$= 0.25$$

Sol

$$P_1 = \frac{X_1}{N_1} = \frac{2}{4} = 0.5$$

$$P_2 = \frac{X_2}{N_2} = \frac{3}{4} = 0.75$$

$$\mu d = P_2 - P_1$$
$$0.25 = 0.75 - 0.5$$
$$0.25 = 0.25$$

$$\sigma^2 d = \Sigma d^2 f(d) - [\Sigma d\, f(d)]^2$$
$$= 0.28125 - (0.25)^2$$
$$= 0.21875$$

$$q_1 = 1 - P_1 = 1 - 0.5 = 0.5$$

$$q_2 = 1 - P_2 = 1 - 0.75 = 0.25$$

$$S^2 d = \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}$$

$$= \frac{0.5 \times 0.5}{2} + \frac{0.75 \times 0.25}{2}$$

$$= \frac{0.4375}{2}$$

$$= 0.21875$$

Sampling dist. of Sample Variance 13-10-14

(6)

Q:- $N = 2, 3, 6, 8, 11$ $n = 2$ W.R

$E(\bar{x}) = \sum \bar{x}\, P(\bar{x})$

$\mu = \frac{\sum X}{N}$, $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$E(S^2) = \sum S^2\, P(S^2)$

$\delta^2 = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2$

S.N	Samples	$\bar{x}$	$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$	
1	2, 2	2	$\frac{1}{2-1}[(2-2)^2 + (2-2)^2] = 0$	
2	2, 3	2.5	$\frac{1}{2-1}[(2-2.5)^2 + (3-2.5)^2] = 0.5$	
3	2, 6	4	$\frac{1}{2-1}[(2-4)^2 + (6-4)^2] = 8$	
4	2, 8	5	$\frac{1}{2-1}[(2-5)^2 + (8-5)^2] = 18$	
5	2, 11	6.5	$\frac{1}{2-1}[(2-6.5)^2 + (11-6.5)^2] = 40.5$	
6	3, 2	2.5	$\frac{1}{2-1}[(3-2.5)^2 + (2-2.5)^2] = 0.5$	
7	3, 3	3	$\frac{1}{2-1}[(3-3)^2 + (3-3)^2] = 0$	

S.N	Samples	$\bar{x}$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
8	3,6	4.5	$\frac{1}{2-1}[(3-4.5)^2+(6-4.5)^2] = 4.5$
9	3,8	5.5	$\frac{1}{2-1}[(3-5.5)^2+(8-5.5)^2] = 12.5$
10	3,11	7	$\frac{1}{2-1}[(3-7)^2+(11-7)^2] = 32$
11	6,2	4	$\frac{1}{2-1}[(6-4)^2+(2-4)^2] = 8$
12	6,3	4.5	$\frac{1}{2-1}[(6-4.5)^2+(3-4.5)^2] = 4.5$
13	6,6	6	$\frac{1}{2-1}[(6-6)^2+(6-6)^2] = 0$
14	6,8	7	$\frac{1}{2-1}[(6-7)^2+(8-7)^2] = 2$
15	6,11	8.5	$\frac{1}{2-1}[(6-8.5)^2+(11-8.5)^2] = 12.5$
16	8,2	5	$\frac{1}{2-1}[(8-5)^2+(2-5)^2] = 18$
17	8,3	5.5	$\frac{1}{2-1}[(8-5.5)^2+(3-5.5)^2] = 12.5$
18	8,6	7	$\frac{1}{2-1}[(8-7)^2+(6-7)^2] = 2$
19	8,8	8	$\frac{1}{2-1}[(8-8)^2+(8-8)^2] = 0$
20	8,11	9.5	$\frac{1}{2-1}[(8-9.5)^2+(11-9.5)^2] = 4.5$
21	11,2	6.5	$\frac{1}{2-1}[(11-6.5)^2+(2-6.5)^2] = 40.5$
22	11,3	7	$\frac{1}{2-1}[(11-7)^2+(3-7)^2] = 32$
23	11,6	8.5	$\frac{1}{2-1}[(11-8.5)^2+(6-8.5)^2] = 12.5$
24	11,8	9.5	$\frac{1}{2-1}[(11-9.5)^2+(8-9.5)^2] = 4.5$
25	11,11	11	$\frac{1}{2-1}[(11-11)^2+(11-11)^2] = 0$

Sampling dist of $\bar{x}$.

$\bar{x}$	f	$f(\bar{x})$	$\bar{x} f(\bar{x})$
2	1	$1/25$	$2/25$
2.5	2	$\frac{2}{25}$	$5/25$
3	1	$1/25$	$\frac{3}{25}$
4	2	$\frac{2}{25}$	$8/25$
4.5	2	$2/25$	$9/25$
5	2	$2/25$	$\frac{10}{25}$
5.5	2	$2/25$	$11/25$
6	1	$1/25$	$6/25$
6.5	2	$\frac{2}{25}$	$13/25$
7	4	$4/25$	$28/25$
8	1	$1/25$	$\frac{8}{25}$
8.5	2	$\frac{2}{25}$	$17/25$
9.5	2	$\frac{2}{25}$	$19/25$
11	1	$1/25$	$11/25$
	25		$\frac{150}{25} = 6$

$$E(\bar{x}) = \sum \bar{x} f(\bar{x}) = 6$$

$$\mu = \frac{\sum X}{N}$$

$$= \frac{2+3+6+8+11}{5} = \frac{30}{5}$$

$$= 6$$

So, $\mu = E(\bar{x})$

Sampling dist. of S^2.

S^2	f	$f(S^2)$	$S^2 f(S^2)$
0	5	$5/25$	0
0.5	2	$2/25$	$1/25$
2	2	$2/25$	$4/25$
4.5	4	$4/25$	$18/25$
8	2	$2/25$	$16/25$
12.5	4	$4/25$	$50/25$
18	2	$2/25$	$36/25$
32	2	$2/25$	$64/25$
40.5	2	$2/25$	$81/25$

$$\frac{54}{5} = \frac{270}{25} = 10.8$$

$$E(S^2) = \Sigma S^2 f(S^2)$$

$$= 10.8$$

$$\delta^2(x) = \frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2$$

$$= \frac{4+9+36+64+121}{5} - (6)^2$$

$$= 46.8 - 36$$

$$= 10.8$$

So,

$$E(S^2) = \delta^2(x)$$

Chap# Statistical Inference.

The process of drawing inferences about a population on the basis of information contained in a sample taken from the population is called statistical inference.

It has two types:-

(i):- Estimation of parameter

(ii):- Testing of Hypothesis

(i) Estimation of Parameter.

Estimation is a procedure by which we obtain an estimate of the true but unknown value of a population parameter by using the sample observations $X_1, X_2, \cdots, X_n$ from the population.

(ii):- Testing of Hypothesis

Testing of hypothesis is a procedure in which enables us to decide on the basis of information obtained by sampling whether to accept or reject any specified statement or hypothesis regarding the value of the parameter in a statistical problem.

An $\boxed{\text{Estimate}}$ is a numerical value of the unknown parameter obtained by applying a rule or a formula, called an Estimator.

"Types of Estimate"

(i) Point estimate.

(ii) Interval estimate

⇒ (i) Point Estimate

When an estimate for the unknown population parameter is expressed by a single value, it is called a point estimate.

$$\bar{X} = \frac{\Sigma X}{N}$$

(ii) Interval Estimate :-

An estimate expressed by a range of values within ~~which~~ the true value of the population parameter is believed to lie, is referred to as an interval estimate.

(B) If an estimate can be expressed as a sum of the weighted observations (i.e as a linear combination), it is said to be a linear estimate.

$\bar{x}$ is L.E of μ ⇒ $\bar{x} = \frac{1}{n}x_1 + \frac{1}{n}x_2 + \dots + \frac{1}{n}x_n$

Criteria For Good Point Estimators

(i):- Unbiasedness
(ii):- Consistency
(iii):- Efficiency
(iv):- Sufficiency.

Point Estimation

A procedure in which we calculate the point estimate is called point estimation.

Interval Estimation

A procedure in which we calculate the interval estimate is called interval estimation.

(I):- Unbiasedness

An estimator is defined to be unbiased if the statistic used as an estimator has it's expected value equal to the true value of the population parameter being estimated.

$$E(\bar{X}) = \mu$$

Proof:- $L.H.S = E(\bar{X})$

$$= E\left(\sum_{i=1}^{n}\left(\frac{X_i}{n}\right)\right)$$

$$= \frac{1}{n} E[X_1 + X_2 + \cdots + X_n]$$

$$= \frac{1}{n}[E(X_1) + E(X_2) + \cdots + E(X_n)]$$

$$= \frac{1}{n}[\underbrace{\mu + \mu + \cdots + \mu}_{n\text{-times}}]$$

$$= \frac{1}{n}[n\mu]$$

$$= \mu$$

$$= R.H.S$$

Now prove $E(S^2) = \sigma^2$

$$L.H.S = E(S^2)$$

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

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

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

$$= \frac{1}{n-1} E\{((X_1 - \mu) - (\bar{X} - \mu))^2 + ((X_2 - \mu) - (\bar{X} - \mu))^2 + \cdots + ((X_n - \mu) - (\bar{X} - \mu))^2\}$$

n-times

$$= \frac{1}{n-1} E[\{(X_1 - \mu)^2 - 2(X_1 - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2\}$$
$$+ \{(X_2 - \mu)^2 - 2(X_2 - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2\}$$
$$+ \cdots + \{(X_n - \mu)^2 - 2(X_n - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2\}]$$

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

We know that

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

$$\Rightarrow (\bar{X}-\mu) = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)$$

$$\Rightarrow n(\bar{X}-\mu) = \sum_{i=1}^{n}(X_i-\mu)$$

So,

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n}(X_i-\mu)^2 - 2(\bar{X}-\mu)\cdot n(\bar{X}-\mu) + n(\bar{X}-\mu)^2\right]$$

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

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

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i-\mu)^2 - n\,E(\bar{X}-\mu)^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} Var(X_i) - n\,Var(\bar{X})\right]$$

$$= \frac{1}{n-1}\left[n\sigma^2 - \cancel{n}\cdot\frac{\sigma^2}{\cancel{n}}\right] \qquad \because \sum_{i=1}^{n} Var(X_i) = n\sigma^2$$

$$\& \quad Var(\bar{X}) = \frac{\sigma^2}{n}$$

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

$$= \sigma^2$$

$$= R.H.S$$

(ii) Consistency:-

An estimator is said to be constant if the statistic to be used as estimator becomes closer & closer to the population parameter being estimated as the sample size n increases.

$\hat{\theta}$ is a consistent estimator of the parameter θ if, for any arbitrarily small positive quantity e,

$$\lim_{n\to\infty} P[|\hat{\theta}-\theta| \leq e] = 1$$

Probability Estimated Value Population parameter

A consistent estimator may or may not be unbaised.

A statistic whose standard error decreased with the increasing sample size, will be consistent.

(iii) Efficiency:-

An unbiased estimator is defined to be efficient if the variance of it's sampling distribution is smaller than that of the sampling distribution of any other unbiased estimator of the same parameter.

(B) Suppose there are two unbiased estimators T_1 & T_2 of the same parameter θ, then T_1 will be said to be more efficient estimator than T_2 if

$$Var(T_1) < Var(T_2).$$ if var $\downarrow$ efficiency $\uparrow$

Example: 15.7

$$T_1 = \frac{X_1 + X_2 + X_3 + X_4}{4}$$

$$T_2 = \frac{X_1 + 2X_2 + 3X_3 + X_4}{7}$$

if $\frac{Var(T_1)}{Var(T_2)} < 1 \Rightarrow T_1$ is more efficient.

if $\frac{Var(T_1)}{Var(T_2)} > 1 \Rightarrow T_2$ is more efficient.

Sol:- $E(T_1) = E\left[\frac{X_1 + X_2 + X_3 + X_4}{4}\right]$

$$= \frac{1}{4}\left[E(X_1) + E(X_2) + E(X_3) + E(X_4)\right]$$

$$= \frac{1}{4}\left[\mu + \mu + \mu + \mu\right]$$

$$= \frac{1}{4}(4\mu)$$

$$= \mu$$

$$Var(T_1) = Var\left[\frac{X_1 + X_2 + X_3 + X_4}{4}\right]$$

$$= \frac{1}{16}\left[Var(X_1) + Var(X_2) + Var(X_3) + Var(X_4)\right]$$

$$= \frac{1}{16}\left[\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2\right]$$

$$= \frac{1}{16}\left[4\sigma^2\right] = \frac{1}{4}\sigma^2$$

Now

$$E(T_2) = E\left[\frac{X_1 + 2X_2 + 3X_3 + X_4}{7}\right]$$

$$= \frac{1}{7}\left[E(X_1) + 2E(X_2) + 3E(X_3) + E(X_4)\right]$$

$$= \frac{1}{7}\left[\mu + 2\mu + 3\mu + \mu\right]$$

$$= \frac{1}{7}\left[7\mu\right]$$

$$= \mu$$

$$Var(T_2) = Var\left[\frac{X_1 + 2X_2 + 3X_3 + X_4}{7}\right]$$

$$= Var\left(\frac{X_1}{7}\right) + Var\left(\frac{2X_2}{7}\right) + Var\left(\frac{3X_3}{7}\right) + Var\left(\frac{X_4}{7}\right)$$

$$= \frac{1}{49}Var(X_1) + \frac{4}{49}Var(X_2) + \frac{9}{49}Var(X_3) + \frac{1}{49}Var(X_4)$$

$$= \frac{1}{49}\left[\sigma^2 + 4\sigma^2 + 9\sigma^2 + \sigma^2\right]$$

$$= \frac{15}{49}\sigma^2$$

$$\frac{Var(T_1)}{Var(T_2)} = \frac{\sigma^2/4}{15\sigma^2/49}$$

$$= \frac{49}{60} < 1$$

$$\Rightarrow Var(T_1) < Var(T_2)$$

So,

T_1 is more efficient than T_2.

④ Sufficiency (8) 20-10-14

An estimator is defined to be sufficient if the statistic used as estimator uses all the information that is contained in the sample.

The sample mean $\bar{X}$ is a sufficient estimator of μ.

Mean is better than median.

Pooled Estimator.

$$\bar{X}_c = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

pooling = combining

example: $n_1 = 20$, $\bar{X}_1 = 5.79$

$n_2 = 40$, $\bar{X}_2 = 10.84$

$$\bar{\bar{X}}_c = \frac{20(5.79)+40(10.84)}{20+40}$$

$$= \frac{549.4}{60} = 9.16$$

Variance of Pooled

$$S_p^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1+n_2-2}$$

example
Q:- 15.10

X =	18	19	20	21	22	23
f =	12	23	10	7	5	3

A second sample of 40 observations

$\sum X = 800$

$\sum X^2 = 16052$

$$S_1^2 = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2$$

$$= \frac{16052}{40} - \left(\frac{800}{40}\right)^2$$

$$= 401.3 - 400$$

$$= 1.3$$

$$S_2^2 = \frac{\sum f X^2}{\sum f} - \left(\frac{\sum f X}{\sum f}\right)^2$$

$$= \frac{3888+8363+4000+3087+2420+1587}{60}$$

$$- \left(\frac{216+437+200+147+110+69}{60}\right)^2$$

$$S_2^2 = \frac{23285}{60} - \left(\frac{1179}{60}\right)^2$$

$$= 388.08 - (19.65)^2$$

$$= 1.9575$$

$$S_p^2 = \frac{40(1.39) + 60(1.9575)}{40+60-2}$$

$$= 1.729$$

$$= 1.73$$

Need for Interval Estimation.

Any point estimate has a limitation that it does not provide information about decision of the estimate that is about the magnitude of error due to sampling of such information are essential for proper interpretation of sample result.

Confidence Interval

22-10-14

Confidence interval to express the decision and uncertainity associated with a particular sampling method.

It has three parts:

(i):- Confidence level

(ii):- Margin of error = Critical value × S.D

(iii):- Statistic

(i) Confidence Level

A probability part of confidence interval is called confidence level.

(ii) Margin of Error

Range of values above and below the samples statistic is called margin of error.

σ known

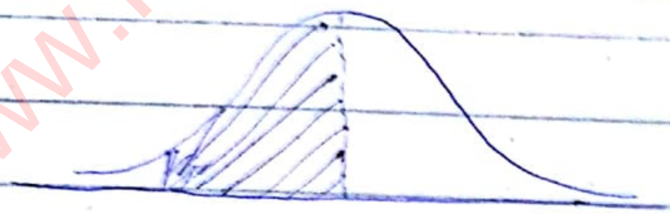
Z-test.

$\Rightarrow \bar{x} \pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$

σ unknown

t-test

$\Rightarrow \bar{x} \pm t_{v\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$



$\bar{x} - Z_{\alpha}\left(\frac{\sigma}{\sqrt{n}}\right) \longrightarrow$ one tail.

$\bar{x} - t_{v\alpha}\left(\frac{S}{\sqrt{n}}\right) \longrightarrow$ one tail.

Significance level (α)	Two tailed test	One Tailed Test
0.10	$\pm 1.645 = Z_{\alpha/2}$	$\pm 1.28 = Z_{\alpha}$
0.05	$\pm 1.96 = Z_{\alpha/2}$	$\pm 1.645 = Z_{\alpha}$
0.01	$\pm 2.58 = Z_{\alpha/2}$	$\pm 2.33 = Z_{\alpha}$

Steps to follow forming Confidence Interval.

(I):- Identify population of interest and state the conditions required for the validity of procedure being used to construct the confidence interval.

(II):- Give the procedure that will be used.

(III):- Construct the confidence interval.

(IV):- Interpret the results.

Page # 96

Example:
Q:15.18 The standard deviation of the amounts poured into bottles by an automatatic filling machine is 1.8 ml. The amounts of fill in a random sample of bottles, in ml, were 481, 479, 482, 480, 477, 478, 481 & 482. Suppose the population of amounts of fill is normal. Construct a 90% confidence interval for the mean amount in all bottles filled by the machine.

Sol:- $\mu = \bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{3840}{8} = 480$$

$\sigma = 1.8$

$n = 8$

$1-\alpha = 0.90$

$\alpha = 1-0.90$

$\alpha = 0.10$

$\Rightarrow \quad \mu = 480 \pm 1.645\left(\frac{1.8}{\sqrt{8}}\right)$

$= 480 \pm 1.645\,(0.636)$

$= 480 \pm 1.05$

$\Rightarrow \quad 478.95 < \mu < 481.05$

Hence the 90% C.I for μ calculated from the given sample is $(478.95, 481.05)$

Example: 15.21

A sample of 100 observations from a population known to be non-normal yielded the sample values $\bar{x} = 182$ & $s^2 = 299$. Find an approximate 99% confidence interval for μ.

Solution:- $n = 100$

$1-\alpha = 0.99$

$\alpha = 0.01$

$\Rightarrow Z_{\alpha/2} = \pm 2.58$

$\bar{x} = 182$

$S^2 = 299$

$\Rightarrow S = 17.292$

$\mu = \bar{x} \pm Z_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$

$= 182 \pm (2.58)\left(\frac{17.292}{\sqrt{100}}\right)$

$= 182 \pm (2.58)(1.7292)$

$= 182 \pm 4.46$.

$\Rightarrow 177.54 < \mu < 186.46$

Hence an approximate 99% C.I for μ calculated from the given sample is $(177.54, 186.46)$

(10)

Confidence Interval for Proportion. 23-10-14

$$\hat{p} \pm Z_{\alpha/2}\left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

Q:- $n = 500$, $x = 40$, $\alpha = 1 - 0.99$
$\alpha = 0.01$

$\hat{p} = \frac{40}{500} = 0.08$

$\hat{q} = 1 - \hat{p}$
$= 1 - 0.08$
$= 0.92$

$$p = \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.08 \pm 2.58\left(\sqrt{\frac{(0.08)(0.92)}{500}}\right)$$

$$= 0.08 \pm 0.0313$$

$$\Rightarrow 0.048 < p < 0.111$$

Confidence Interval for difference Mean. $(\bar{x}_1 - \bar{x}_2)$

statement on p#104

Example: 15.24

$\bar{x}_1 = 345$, $\bar{x}_2 = 340$

$n_1 = 100 = n_2$

$\sigma_1^2 = 196$, $\sigma_2^2 = 204$

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

For 99%

$$\alpha = 1 - 0.99 = 0.01$$

$$\mu_1 - \mu_2 = 5 \pm 2.58\left(\sqrt{\frac{196}{100} + \frac{204}{100}}\right)$$

$$= 5 \pm 2.58\,(2)$$

$$= 5 \pm 5.16$$

$$\Rightarrow -0.16 < \mu_1 - \mu_2 < 10.16$$

for 95%

$$\alpha = 1 - 0.95 = 0.05$$

$$\Rightarrow \mu_1 - \mu_2 = 5 \pm 1.96 \sqrt{\frac{196}{100} + \frac{204}{100}}$$

$$= 5 \pm 1.96\,(2)$$

$$= 5 \pm 3.92$$

$$\Rightarrow 1.08 < \mu_1 - \mu_2 < 8.92$$

Example. 15.27

In a poll of college students in a large state university, 300 of 400 students living in dormitories approved a certain course of action, whereas 200 of 300 students not living in dormitories approved it. Estimate the difference in the proportions favouring the course of action and compute 90 % confidence interval for it.

Solution:- Let $\hat{p}_1$ & $\hat{p}_2$ be the observed proportions in the first & second sample respectively. Then

$\hat{p}_1 = \frac{300}{400} = 0.75$ $\quad\quad \hat{q}_1 = 1 - 0.75 = 0.25$

& $\hat{p}_2 = \frac{200}{300} = 0.667$ $\quad\quad \hat{q}_2 = 1 - 0.66 = 0.33$

$\alpha = 1 - 0.90$

$= 0.10$

$Z_{0.10/2} = \pm 1.645$

$$P_1 - P_2 = \hat{P}_1 - \hat{P}_2 \pm Z_{\alpha/2}\sqrt{\frac{\hat{P}_1\hat{q}_1}{n_1} + \frac{\hat{P}_2\hat{q}_2}{n_2}}$$

$$= (0.75 - 0.667) \pm 1.645\left[\sqrt{\frac{0.75\times0.25}{400} + \frac{(0.67)(0.33)}{300}}\right]$$

$$= 0.083 \pm (1.645)\left(\sqrt{0.0004687 + 0.000737}\right)$$

$$= 0.083 \pm (1.645)(0.03477)$$

$$= 0.083 \pm 0.0572$$

$$\Rightarrow \quad 0.023 < P_1 - P_2 < 0.137$$

Hence the 90% confidence interval for $P_1 - P_2$ is $\cancel{(0.023, 0.137)}$
$(0.02612, 0.14054)$

(1)

12-11-14

Chap# 16

"Statistical Hypothesis"

(Q) It is a procedure which enables us to decide on the basis of information obtained from sample data whether to accept or reject a statement (or an assumption) about the value of a population parameter. Such a statement (or assumption) which may or may not be true is called a statistical hypothesis.

"Null Hypothesis"

Any hypothesis which is to be tested for possible rejection under the assumption that it is true. It is denoted by H_0.

"Alternative Hypothesis"

Any hypothesis which we accept when the null hypothesis H_0 is rejected. It is denoted by H_1 (or H_A)

"Simple Hypothesis"

$\mu = 62$

A simple hypothesis is one in which all parameters of the distribution are specified. It is also called exact hypothesis. (but convers not necessary) (exact → simple)

"Composite Hypothesis"

A Hypothesis which is not simple (i.e in which not all of the parameters are specified) is called a composite hypothesis. It is also called inexact hypothesis. $\mu \neq 62$

"Test-Statistic"

A sample statistic which provides a basis for testing a null hypothesis, is called a test-statistic.

Rejection Region

Specified set of values of the test-statistic, the null hypothesis is rejected is called rejection region. It is also called critical region.

Two Tails Rejection Region

If the critical region is located equally in both tails of sampling distribution of test-statistic is called two-tails rejection region.

"One Tail Rejection Region"

If the critical region is located only one side may be right or may be left tail of the sampling distribution of test statistic is called one tail rejection region.

"Type I error"

The probability of rejecting H_0 when it is true is called type I error. It is denoted by α.

"Type II error"

The probability of accepting H_0 when it is false is called type II error. (or an error of the second kind.) It is denoted by β.

* The decision and the corresponding two types of error may be displayed in a tabular form as on next page:

True Situation	Decision: Accept H_0	Decision: Reject H_0 (or accept H_1)
H_0 is true	Correct decision (No error) $1-\alpha$	Wrong decision (Type-I error) α
H_0 is false (or H_1 is true)	Wrong decision (Type-II error) β	Correct decision (No error) $1-\beta$

(Book) Steps for testing statistic

It involves the following six steps:

(i) State your problem & formulate an appropriate null hypothesis H_0 with an alternative hypothesis H_1, which is to be accepted when H_0 is rejected.

(ii) Decide upon a significance level α of the test, which is the probability of rejecting the null hypothesis if it is true.

(iii) Choose an appropriate test-statistic, determine and sketch the sampling distribution of the test-statistic, assuming H_0 is true.

(iv):- Determine the rejection or critical region in such a way that the probability of rejecting the null hypothesis H_o, if it is true, is equal to the significance level α. The location of the critical region depends upon the form of H_1. The significance level will separate the acceptance region from the rejection region.

(V):- Compute the value of the test-statistic from the sample data in order to decide whether to accept or reject the null hypothesis H_o.

(Vi):- Formulate the decision rule as below:

a):- Reject the null hypothesis H_o, if the computed value of the test-statistic falls in the rejection region and conclude that H_1 is true.

b):- Otherwise accept the null hypothesis H_o.

(12)

"Testing Hypothesis about Mean" 19-11-14

(i):- $H_0: \mu = \mu_0$ | $H_0: \mu \leq \mu_0$ | $H_0: \mu \geq \mu_0$

$H_1: \mu \neq \mu_0$ | $H_1: \mu > \mu_0$ | $H_1: \mu < \mu_0$

(ii):- $\alpha = 0.05, 0.01, 0.10$

(iii):- Test Statistic

if σ is known

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

if σ is unknown

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

(iv):- Calculations

(v):- Rejection Region

Reject H_0 if

Rejection Region

two tail $H_1: \mu \neq \mu_0$ (two sided) then $Z < -Z_{\alpha/2}$ or $Z > Z_{\alpha/2}$

negative value   Positive value

one tail $H_1: \mu < \mu_0$ then $Z < -Z_{\alpha}$

one tail $H_1: \mu > \mu_0$ then $Z > Z_{\alpha}$

(vi) Result.

P# 140

Example 16.5

A random sample of $n=25$ values gives $\bar{x}=83$. Can this sample be regarded as drawn from a normal population with mean $\mu=80$ & $\sigma=7$?

$\alpha=0.05$

Solution:— i) $H_0: \mu=80$

& $H_1: \mu\neq 80$

ii) $\alpha=0.05$

iii) Test-statistic $Z=\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

By putting the values,

$$Z=\frac{83-80}{7/\sqrt{25}}=\frac{3\times 5}{7}$$

$$Z=2.14$$

~~iv~~ Critical Region for $\alpha=0.05$

$|Z|\geq 1.96$ where $Z_{\alpha/2}=1.96$

~~v~~ Rejection Region

$H_0: \mu=\mu_0 \longrightarrow$ two tailed.

Now $Z>Z_{\alpha/2}$

$2.14>1.96$

(vi): Result: Reject H_0.

Accept H_1.

Example: 16.6

Test the hypothesis that the mean of a normal population with known variance 70 is 31, if a sample of size 13 gave $\bar{x} = 34$. Let the alternative hypothesis be $H_1: \mu > 31$ & let $\alpha = 0.10$.

Solution:- $H_0: \mu \leq 31$ $\bar{x} = 34$

& $H_1: \mu > 31$ $n = 13$

$\sigma^2 = 70$

$\alpha = 0.10$

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

$$= \frac{34 - 31}{\sqrt{70}/\sqrt{13}}$$

$$= \frac{3 \times \sqrt{13}}{\sqrt{70}}$$

$$= 1.29$$

Rejection Region:-

$H_1: \mu > \mu_0 \Rightarrow Z > Z_\alpha$

$1.29 > 1.28$

Result:- Reject Ho.

Accept H1.

"Testing Hypothesis about Difference Mean:"

$H_0: \mu_1 - \mu_2 = \mu_0$

& $H_1: \mu_1 - \mu_2 \neq \mu_0$

$H_0: \mu_1 - \mu_2 \leq \mu_0$

& $H_1: \mu_1 - \mu_2 > \mu_0$

$H_0: \mu_1 - \mu_2 \geq \mu_0$

$H_1: \mu_1 - \mu_2 < \mu_0$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Example 16.10

A random sample of size 36 from a normal population with variance 24 gave $\bar{x}_1 = 15$. A second sample of size 28 from another normal population with variance 80 give $\bar{x}_2 = 13$. Test $H_0: \mu_1 - \mu_2 = 0$ against $H_1: \mu_1 - \mu_2 \neq 0$. Let $\alpha = 0.05$.

Sol:- $H_0 : \mu_1 - \mu_2 = 0$

& $H_1 : \mu_1 - \mu_2 \neq 0$

$\alpha = 0.05$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{(15-13)-(0)}{\sqrt{\frac{24}{36} + \frac{80}{28}}} = \frac{2}{\sqrt{0.66 + 2.85}}$$

$$= \frac{2}{1.877}$$

$$= 1.06$$

Rejection Region:

$H_1 : \mu_1 - \mu_2 \neq 0$

$\Rightarrow \quad Z > Z_{\alpha/2}$

$1.06 > 1.96$

Result: H_0 is accepted.

(13)

Testing about Mean when σ is unknown. & $n<30$: 2-11-14

Q:- Expensive test boring are made in an oil shall area to determine if the mean yeild of oil per ton of shell rock is greater than 4.5 barrels. 5 boring made at randomly selected points in the area indicate the following number of barrels per ton 4.8, 5.4, 3.9, 4.9 & 5.5. Suppose barrels per ton are normally distributive at $\alpha = 0.05$.

Solution:-

$$H_0 : \mu \leq 4.5$$

$$H_1 : \mu \geq 4.5$$

$$\alpha = 0.05$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \qquad \sigma \text{ is unknown}$$

$$S = \sqrt{\frac{1}{n-1}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)}$$

$$\bar{x} = 4.9 \quad , \quad n = 5$$

$$S = \sqrt{\frac{1}{5-1}(121.67 - 120.05)}$$

$$\boxed{S = 0.64}$$

$$t = \frac{4.9 - 4.5}{0.64/\sqrt{5}}$$

$$t = \frac{0.4}{0.286}$$

$$t = 1.4$$

Rejection Region:-

(i) $t > t_{\alpha(v)}$ $H_1: \mu > \mu_0$

d.f $= n-1$

$1.4 > 2.132$

Result:- Accept H_0

(ii) $H_1: \mu < \mu_0$ $\Rightarrow t < -t_{\alpha(v)}$

(iii) $H_1: \mu \neq \mu_0$ $\Rightarrow t < -t_{\frac{\alpha}{2}(v)}$ or $t > t_{\frac{\alpha}{2}(v)}$

Testing Hypothesis about Difference Mean.

$H_0: \mu_1 - \mu_2 = \mu_0$ $\quad H_0: \mu_1 - \mu_2 \leq \mu_0$ $\quad H_0: \mu_1 - \mu_2 \geq \mu_0$

$H_1: \mu_1 - \mu_2 \neq \mu_0$ $\quad H_1: \mu_1 - \mu_2 > \mu_0$ $\quad H_1: \mu_1 - \mu_2 < \mu_0$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Q:- Two random samples taken independently from normal population with identical variance (use formula of pooled var.) $n_1 = 12$, $n_2 = 18$, $\bar{X}_1 = 10$, $\bar{X}_2 = 25$, $S_1^2 = 1200$, $S_2^2 = 900$. Test the hypothesis that the true difference b/w the population mean is at most 10.

Sol:

$H_0 : \mu_2 - \mu_1 \leq 10$

$H_1 : \mu_2 - \mu_1 > 10$

$\alpha = 0.05$

$\bar{X}_2 - \bar{X}_1 = 15$

$\mu_2 - \mu_1 = 10$

$$S_p = \sqrt{\frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{(12-1)(1200) + (18-1)(900)}{12+18-2}}$$

$$= \sqrt{\frac{(11 \times 1200) + (17 \times 1900)}{28}}$$

$$S_p = 31.9$$

$$t = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{15 - 10}{31.9 \sqrt{\frac{1}{12} + \frac{1}{18}}} = \frac{5}{31.9 (0.3727)}$$

$$t = 0.42$$

R.R:- $t > t_{\alpha(v)}$ $\qquad$ $d.f = n_1 + n_2 - 2$

$0.42 > 1.701$ $\qquad$ $= 28$

Result:- Accept H_0.

Assignment #

Q:- What is the difference & relation b/w Z & t-test.

(14)

24/11/14

(Self) Chi-Square Goodness of Fit test

This test is based on the property that

"The χ^2-test is applicable when the cell probabilities depend upon unknown parameters, provided that the unknown parameters are replaced with their estimates and provided that one degree of freedom is deducted for each parameter estimated."

When there are k classes & the class probabilities are known, the number of degrees of freedom is $k-1$. When the probabilities depend upon m parameters, the degrees of freedom would be $k-1-m$.

The procedure for a goodness of fit test is as follows:

(i):- Formulate the null & alternative hypotheses as;

H_0: The population has a specified probability distribution, and

H_1: The population does not have the specified distribution.

(ii):- Choose the level of significance α - The commonly used value is at $\alpha = 0.05$

(iii):- The test-statistic to use is

$$\chi^2 = \sum_{i=1}^{K} \frac{(o_i - e_i)^2}{e_i}$$

which, if Ho is true, has an approximate chi-square distribution with d.f. = K-1- number of estimated parameters.

(iv):- Determine the critical region, which depends upon α and the degrees of freedom.

(V):- Compute the expected values and the value of χ^2.

(Vi):- Result.

Reject Ho, if the calculated value of χ^2 exeeds the χ^2_α values against the appropriate degrees of freedom from the χ^2-table.

Otherwise, Accept Ho.

(15)

26-11-14

Qualitative data: z, t-test

Quantitative data: χ^2 test

$$\chi^2 = \frac{\sum (O_{ij} - e_{ij})^2}{e_{ij}}$$

Rejection Region:-

Reject Ho if

$$\chi^2 > \chi^2_{\alpha(v)}$$

Q:- From the adult male population of 7 large cities random sample size indicate below were taken the number of married & single men were recorded.

City	A	B	C	D	E	F	G	Total
M	133	164	155	166	153	123	146	980
S	36	57	40	37	55	33	36	294
Total	169	221	195	143	208	156	182	1274

H_0 = the proportion man are same in 7 cities.

H_1 = The " " " not " " "

Expected Frequency -

City	A	B	C	D	E	F	G	Total
M	$\frac{980\times169}{1274}=130$	170	150	110	160	120	140	980
S	$\frac{294\times169}{1274}=39$	51	45	33	48	36	42	294
Total	169	221	195	143	208	156	182	1274

O_{ij}	e_{ij}	$O_{ij}-e_{ij}$	$(O_{ij}-e_{ij})^2$	$\frac{(O_{ij}-e_{ij})^2}{e_{ij}}$
133	130	3	9	0.069
164	170	-6	36	0.2117
155	150	5	25	0.167
106	110	-4	16	0.14546
153	160	-7	49	0.30625
123	120	3	9	0.075
146	140	6	36	0.2571
36	39	-3	9	0.23076
57	51	6	36	0.70589
40	45	5	25	0.556
37	33	4	16	0.4849
55	48	7	49	1.0209
33	36	-3	9	0.25
36	42	-6	36	0.8571
				5.34

$\chi^2 > \chi^2_{\alpha(v)}$

$5.34 > 12.59$

$d.f = n-1$
$= 7-1 = 6$

$\Rightarrow$ Accept H_0

5 Q.17.57 (a).

In a study to determine whether or not the proportions of defectives produced by workers was the same for the day, evening or night shift worked the following data were collected.

	Shift		
	Day	Evening	Night
Defectives	45	455	70
Non-Defectives	905	890	870

Test the hypothesis, at the 0.025 level of significance, that the proportion of defectives is the same for all three shifts.

Solution:-

H_0: proportion of defectives is the same for all three shifts.

H_1: Proportion of defectives is not same for all three shifts.

$\alpha = 0.02$

	Shift			
	Day	Evening	Night	Total
Defectives	45	55	70	170
Non-Defectives	905	890	870	2665
Total	950	945	940	2835

Expected Frequency

	Shift			
	Day	Evening	Night	Total
Defectives	$\frac{950 \times 170}{2835} = 56.96$	56.66	56.37	170
Non-Defectives	$\frac{2665 \times 170}{2835} = 893.03$	888.34	883.63	2665
Total	950	945	940	2835

O_{ij}	e_{ij}	$O_{ij} - e_{ij}$	$(O_{ij} - e_{ij})^2$	$(O_{ij} - e_{ij})^2 / e_{ij}$
45	56.96	-11.96	143.04	2.51
55	56.66	-1.66	2.7556	0.048
70	56.37	13.634	185.88	3.297
905	893.03	11.97	143.28	0.1604
890	888.34	1.66	2.7556	0.003
870	883.63	-13.63	185.88	0.2104
				6.228

$\chi^2 > \chi^2_{\alpha(v)}$

$6.228 > 7.38$

$d.f = 3 - 1 = 2$

Result:- Accept Ho

$\Rightarrow$ Proportion of defectives is same for all three shifts.

(16)

Example 17.10 1-12-14

Five pennies were tossed 1000 times and the number of heads were observed as given below:

No. of Heads	0	1	2	3	4	5
Frequencies	38	144	342	287	164	25

Test whether a binomial distribution gives a satisfactory fit to these data.

H_0: The population distribution is a binomial with $n=5$.

H_1: The population distribution is not a binomial with $n=5$.

$\alpha = 0.05$

$$\chi^2 = \frac{\Sigma (o_{ij} - e_{ij})^2}{e_{ij}}$$

Solution:- Binomial

$$\binom{n}{x} p^x q^{n-x} \qquad x = 0, 1, 2, \cdots n$$

$$x = 0, 1, 2, \cdots, 5$$

Binomial distribution $\Rightarrow n = 5$

has two parameters p & n.

$$\bar{X} = np$$

$$\sigma^2 = npq$$

$$\bar{X} = \frac{\sum fx}{\sum f}$$

$$= \frac{0+144+684+861+656+125}{38+144+342+287+164+25}$$

$$= \frac{2470}{1000} = 2.47$$

$$\bar{X} = np$$

$$5p = 2.47$$

$$p = 0.494$$

$$q = 1 - 0.494$$

$$= 0.506$$

Expected frequency.

$$E = N \binom{n}{x} p^{x} q^{n-x}$$

$x = 0$

$$= 1000 \binom{5}{0} (0.494)^{0} (0.506)^{5-0}$$

$$= 33.2$$

$x = 1$

$$= 1000 \binom{5}{1} (0.494)^{1} (0.506)^{5-1}$$

$$= 161.9$$

$x = 2$

$$= 1000 \binom{5}{2} (0.494)^{2} (0.506)^{5-2}$$

$$= 316.2$$

$x=3$

$$= 1000 \binom{5}{3} (0.494)^3 (0.506)^2$$

$$= 308.7$$

$x=4$

$$= 1000 \binom{5}{4} (0.494)^4 (0.506)^1$$

$$= 150.7$$

$x=5$

$$= 1000 \binom{5}{5} (0.494)^5 (0.506)^0$$

$$= 29.4$$

O_{ij}	e_{ij}	$O_{ij}-e_{ij}$	$(O_{ij}-e_{ij})^2$	$\frac{(O_{ij}-e_{ij})^2}{e_{ij}}$
38	33.2	4.8	23.04	0.69
144	161.9	-17.9	320.41	1.98
342	316.2	25.8	665.64	2.15
287	308.7	-21.7	470.89	1.53
164	150.7	13.3	176.89	1.17
25	29.4	-4.4	19.36	0.66

$\chi^2 = 8.18$

$$d.f = K-1-m$$
$$= 6-1-1$$
$$= 4$$

SHAKOOR AHMAD
Mob # 0331-SHAKOOR

R.R.

$$\chi^2 > \chi^2_{\alpha(v)}$$

$$8.18 > 9.49$$

Result:- Accept H_0

So, The population distribution is a binomial distribution.

---

⑤ Poisson Distribution

Uses:- We use poisson approximation when p is 0.05 or less and n is 20 or more.

If we assume that n goes to infinity and p approaches zero in such a way that

$\mu = np$ remains constant

then the limiting form of the binomial probability distribution is

$$\lim_{\substack{n \to \infty \\ p \to 0}} b(x; \underbrace{n, p}_{\mu}) = \frac{\mu^x e^{-\mu}}{x!}$$

$x = 0, 1, 2, \ldots, \infty$

The poisson distribution has only one parameter $\mu > 0$

(17)

3-12-14

Example: 17·11

A skilled typist, on routine work, kept a record of mistakes made per day during 300 working day;

Mistakes per day (x)	0	1	2	3	4	5	6
Number of days (f)	143	90	42	12	9	3	1

Test the hypothesis that X has a Poisson distribution by applying the χ^2 goodness-of-fit-test. $\alpha = 0.05$

Sol:-

H_0: The population has a Poisson distribution.

H_1: The population does not have a Poisson distribution.

$$\mu = \frac{\Sigma(f x)}{\Sigma f}$$

$$= \frac{0+90+84+36+36+15+6}{143+90+42+12+9+3+1}$$

$$= \frac{267}{300} = 0.89$$

Expected Frequency

$$P(x; 0.89) = \frac{e^{-0.89}(0.89)^x}{x!}$$

$$x = 0, 1, 2, \ldots, 6$$

$$e_{ij} = N\left[\frac{e^{-0.89}(0.89)^x}{x!}\right]$$

Mistakes per day (x)	Observed f, O_{ij}	Expected f, e_{ij}	$O_{ij}-e_{ij}$	$\frac{(O_{ij}-e_{ij})^2}{e_{ij}}$
0	143	$300\left[\frac{e^{-0.89}(0.89)^0}{0!}\right] = 123.2$	19.8	3.18
1	90	109.6	−19.6	3.51
2	42	48.8	−6.8	0.95
3	12 } 25	14.5 } 18.475	6.6	2.37
4	9	3.2		
5	3	0.6		
6	1	0.1		

$\chi^2 = 10.01$

$$d.f = K-1-m$$
$$= 4-1-1$$
$$= 2$$

R.R.

$$\chi^2 > \chi^2_{\alpha(2)}$$
$$10.01 > 5.99$$

Result:- Reject H_0

$\Rightarrow$ The population does not have a poisson distribution.

(18)

Q= 17.46 4-12-14

A 1000 households are taken at random and divided into three groups A, B and C according to the total monthly income. The following table shows the numbers in each group having a colour television receiver, a black and white receiver, or no television at all.

	A	B	C
Colour TV	56	51	93
Black & white	118	207	375
None	26	42	32

Test the hypothesis that there is no association between total income and television ownership.

Solution:-

H_0:

H_1:

	A	B	C	Total
Colour TV	56	51	93	200
Black & white	118	207	375	700
None	26	42	32	100
Total	200	300	500	1000

Expected frequency.

	A	B	C	Total
Colour TV	$\frac{200\times 200}{1000}=40$	60	100	200
Black & white	140	210	350	700
None	20	30	50	100
Total	200	300	500	1000

O_{ij}	e_{ij}	$O_{ij}-e_{ij}$	$(O_{ij}-e_{ij})^2$	$(O_{ij}-e_{ij})^2/e_{ij}$
56	40	16	256	6.4
51	60	-9	81	1.35
93	100	-7	49	0.49
118	140	-22	484	3.45
207	210	-3	9	0.042
375	350	25	625	1.785
26	20	6	36	1.8
42	30	12	144	4.8
32	50	-18	324	6.48
				26.59

$$d.f. = (\overset{\text{Rows}}{r}-1)(\overset{\text{columns}}{c}-1)$$
$$= (3-1)(3-1)$$
$$= 4$$

R.R.

$$\chi^2 \geq \chi^2_{\alpha(v)}$$
$$26.56 > 9.49$$

Result:

Rejected H_0

$\Rightarrow$

Use Book S.M. Ch Vol. 1

Regression:- 8-12-14

Sir Francis Galton (1822-1911)

Def:- Regression is a process by which we estimate the value of dependent variable on the basis of 1 or more independent variables.

Linear Regression:-

Regression analysis involving one independent variable and one dependent variable in which the relationship b/w the variables is approximated by a straight line is called Linear regression.

Non-Linear Regression

Regression analysis involving one independent variable and one dependent variable in which the relationship b/w the variables is approximated by not a straight line is called non-linear regression.

Simple Regression Equation

An equation that includes one independent and one dependent variable is called simple regression eq..

Multiple Regression Equation

An equation that include one dependent variable and more than one independent variable is called multiple regression equation.

Residual:-

The difference between observed value of the dependent variable and the value predicted by using the

$-1 < r < 1$

estimated regression equation is called residual.

$$Y_i - \hat{Y} = e_i \leftarrow \text{residual}$$

observed estimate

Correlation:-

Correlation is measure of degree of linear association b/w two variables.

OR

Strength of linear association b/w two variables is called correlation.

Positive Correlation

When the values of two variables move in same direction. so that increases or decreases in the value of one variable is associated with an increases and decreases the value of other variable.

Negative Correlation

When the values of two variables move in different directions so that with an increase in the value of one variable the value of other variable is decreases and with a decreasing in the value of one variable the value of other variable is increases.

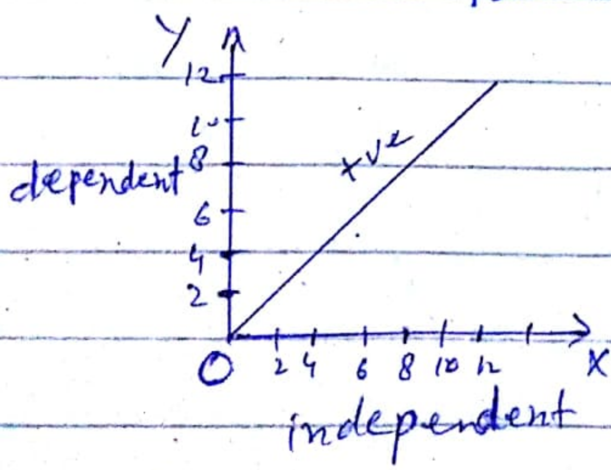
No correlation

When the variables have zero correlation is said to be no correlation.

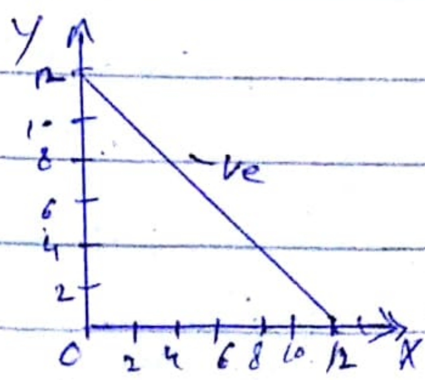
* Curvilinear Correlation:-

When correlation b/w two variables represent a curve that is not a straight line then the correlation is said to be curvilinear correlation.

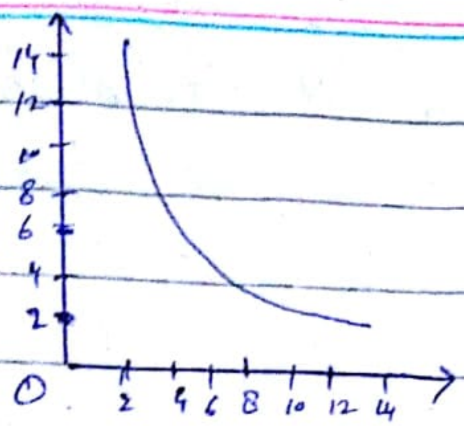
Scatter Plot:-

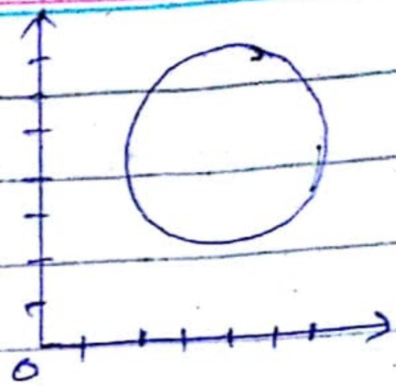


+ive



-ive





Curvilinear

No-correlation.

(20) 10-12-14

Methods of Estimation

(1):- Method of Least Square.

(2):- Method of Moments.

(3):- Method of Likelyhood.

Method of Least Square

The principle of least square consist of determining the values of unknown parameters that will minimize the sum of squares of residue,

$$e_i = (y_i - \hat{y})$$

$$\Sigma e_i^2 = \Sigma(y_i - \hat{y})^2$$

$$\Sigma(y_i - \hat{y}) = 0$$

Regression line by using least Square.

$$\hat{Y} = a + b X_i$$

Dependent variable

Intercept

Slop

Independent variable

Slope:→ Unit change in dependent variable w.r.t independent variable.

Y on X.

$$\hat{Y} = a + b X_i$$

$$b_{yx} = \frac{n \sum (X_i Y_i) - (\sum X_i)(\sum Y_i)}{n \sum X_i^2 - (\sum X_i)^2}$$

$$a = \bar{Y} - b\bar{X}$$

X on Y.

$$\hat{X} = a + b Y_i$$

$$b_{xy} = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n \sum Y_i^2 - (\sum Y_i)^2}$$

$$a = \bar{X} - b\bar{Y}$$

Example: 10.2

Load (lb) X_i	Length (inches) Y_i	X_i^2	Y_i^2	X_iY_i
3	10	9	100	30
5	12	25	144	60
6	15	36	225	90
9	18	81	324	162
10	20	100	400	200
12	22	144	484	264
15	27	225	729	405
20	30	400	900	600
22	32	484	1024	704
28	34	784	1156	982
130	220	2288	5486	3467

X on Y.

$$\hat{X} = a + bY_i$$

$$a = \bar{X} - b\bar{Y}$$

$$b_{xy} = \frac{n\sum X_iY_i - (\sum X_i)(\sum Y_i)}{n\sum Y_i^2 - (\sum Y_i)^2}$$

$$= \frac{10(3467) - (130)(220)}{10(5486) - (220)^2}$$

$$= \frac{34670 - 28600}{54860 - 48400} = \frac{6070}{6460} = 0.94$$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{130}{10} = 13$$

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{220}{10} = 22$$

$$a = \bar{X} - b\bar{Y}$$
$$= 13 - (0.94)22$$
$$= 13 - 20.64$$
$$= -7.68$$

$$\hat{X} = a + bY_i$$
$$\hat{X} = -7.68 + 0.94\,Y_i$$

To check: at $Y = 29$

$$\hat{X} = -7.68 + 0.94(29)$$
$$\hat{X} = 19.58$$

which lies b/w 15 and 20.

Another Formula for b.

$$b_{xy} = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(Y_i - \bar{Y})^2}$$

$$b_{yx} = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2}$$

Deterministic Model

$$Y_i = a + bX_i \qquad (i = 1, 2, \ldots, n)$$

Example: $F = 32 + \frac{9}{5}C$

Fahrenheit Celsius

If the relationship b/w the variables is exactly linear is called deterministic model.

Probabilistic Model (Non-deterministic Model)

If the relationship b/w the variables is not exactly (inexact) linear is called probabilistic model or non-deterministic model.

$$Y_i = a + bX_i + e_i \quad (i = 1, 2, \ldots, n)$$

Where e_i's are the unknown random errors.

⑤ 10.4.5 Properties of the Least-Squares Regression line

The least-squares linear regression line has the following properties:

(i):– The least squares regression line always goes through the point $(\bar{X}, \bar{Y})$, the means of the data.

(ii) The sum of the deviations of the observed values Y_i from the least squares regression line is always equal to zero.

i.e $\sum (Y_i - \hat{Y}) = 0$

(iii):- The sum of the squares of the deviations of the observed values from the least squares regression line is minimum i.e

$$\sum (Y_i - \hat{Y})^2 = \text{minimum}$$

(iv):- The least squares regression line obtained from a random sample is the line of best fit because "a" & "b" are the unbaised estimates of the parameters α and β.

Q:

X_i	Y_i	X_iY_i	X_i^2	Y_i^2
3	25	75	9	625
4	24	96	16	576
5	20	100	25	400
6	20	120	36	400
7	19	133	49	361
8	17	136	64	289
9	16	140	81	256
10	13	130	100	169
11	10	110	121	100
12	6	72	144	36

$\sum X_i = 75$

$\sum Y_i = 170$

$\sum X_iY_i = 1116$

$\sum X_i^2 = 645$

$\sum Y_i^2 = 3212$

$\bar{x} = \frac{\sum x}{n} = \frac{75}{10} = 7.5$

$\bar{y} = \frac{\sum y}{n} = \frac{170}{10} = 17$

(i):- y on x

$\hat{y} = a + bx_i$

$b = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$

$= \frac{10(1116) - (75)(170)}{10(645) - (75)^2}$

$= \frac{11160 - 12750}{6450 - 5625} = \frac{-1590}{825}$

$\boxed{b = -1.927}$

$a = \bar{y} - b\bar{x}$

$= 17 - (-1.927)(7.5)$

$= 17 - 14.45$

$\boxed{a = 31.45}$

So, $\hat{y} = a + bx_i$

$\Rightarrow \hat{y} = 31.45 - 1.927 x_i$

(ii). X on Y.

$$\hat{X} = a + bY_i$$

$$b = \frac{n\sum X_iY_i - (\sum X_i)(\sum Y_i)}{n\sum Y_i^2 - (\sum Y_i)^2}$$

$$= \frac{10(1116) - (75)(170)}{10(3212) - (170)^2}$$

$$= \frac{-1590}{3220}$$

$$\Rightarrow \boxed{b = -0.4938}$$

$$\boxed{b = -0.5}$$

$$a = \bar{X} - b\bar{Y}$$

$$a = 7.5 - (-0.5)(17)$$

$$a = 7.5 + 8.5$$

$$\boxed{a = 16}$$

So, $\hat{X} = 16 - 0.5Y_i$

Correlation coefficient for the Group data:- 11-12-14 (21)

$$r = \frac{n\sum X_iY_i - (\sum X_i)(\sum Y_i)}{\sqrt{(n\sum X_i^2 - (\sum X_i)^2)(n\sum Y_i^2 - (\sum Y_i)^2)}}$$

OR

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

and

$$r = \pm\sqrt{b_{xy} \cdot b_{yx}}$$

Example: 10.5 Calculate the product moment co-efficient of correlation (r) b/w X and Y from the following data.

X	1	2	3	4	5
Y	2	5	3	8	7

Solution:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{15}{5} = 3$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{25}{5} = 5$$

X_i	Y_i	$(X_i-\bar{X})$	$(X_i-\bar{X})^2$	$(Y_i-\bar{Y})$	$(Y_i-\bar{Y})^2$	$(X_i-\bar{X})(Y_i-\bar{Y})$	X_i^2	Y_i^2	X_iY_i
1	2	-2	4	-3	9	6	1	4	2
2	5	-1	1	0	0	0	4	25	10
3	3	0	0	-2	4	0	9	9	9
4	8	1	1	3	9	3	16	64	32
5	7	2	4	2	4	4	25	49	35
15	25	0	10	0	26	13	55	151	88

$$r = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2\sum(Y_i-\bar{Y})^2}} = \frac{13}{\sqrt{10 \times 26}}$$

$$\boxed{r \simeq 0.8}$$

Also

By
$$r = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{(n\sum x_i^2 - (\sum x_i)^2)(n\sum y_i^2 - (\sum y_i)^2)}}$$

$$r = \frac{5(88) - (15)(25)}{\sqrt{(5(55) - (15)^2)(5(151) - (25)^2)}}$$

$$= \frac{440 - 375}{\sqrt{(50)(130)}}$$

$$= \frac{65}{80.62}$$

$$= 0.806$$

$$\boxed{r = 0.8}$$

Also $\quad b_{yx} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$

$$= \frac{13}{10} = 1.3$$

& $\quad b_{xy} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (y_i - \bar{y})^2}$

$$= \frac{13}{26} = 0.5$$

$$r = \pm\sqrt{b_{yx} \cdot b_{xy}}$$

$$r = \sqrt{1.3 \times 0.5}$$

$$\boxed{r = 0.8}$$

Q:- 10.33

X	5	2	6	8	1	7	4	9	3	10
Y	1	7	6	10	4	5	3	8	2	9
Z	6	4	9	8	1	2	3	10	5	7

In a painting competition, various entries are ranked by three Judges. Discuss which pair of Judges has the nearest approach to common tastes.

Solution:-

X_i	Y_i	Z_i	X_i^2	Y_i^2	Z_i^2	X_iY_i	X_iZ_i	Z_iY_i
5	1	6	25	1	36	5	30	6
2	7	4	4	49	16	14	8	28
6	6	9	36	36	81	36	54	54
8	10	8	64	100	64	80	64	80
1	4	1	1	16	1	4	1	4
7	5	2	49	25	4	35	14	10
4	3	3	16	9	9	12	12	9
9	8	10	81	64	100	72	90	80
3	2	5	9	4	25	6	15	10
10	9	7	100	81	49	90	70	63
55	55	55	385	385	385	354	358	344

$$r_{xy} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{(n\sum x_i^2 - (\sum x_i)^2)(n\sum y_i^2 - (\sum y_i)^2)}}$$

$$= \frac{10(354) - (55)(55)}{\sqrt{((10)(385) - (55)^2)(10(385) - (55)^2)}}$$

$$'' = \frac{3540 - 3025}{\sqrt{(825)(825)}}$$

$$'' = \frac{515}{825}$$

$$\boxed{r_{xy} = 0.624}$$

$$r_{xz} = \frac{n\sum x_i z_i - (\sum x_i)(\sum z_i)}{\sqrt{(n\sum x_i^2 - (\sum x_i)^2)(n\sum z_i^2 - (\sum z_i)^2)}}$$

$$= \frac{10(358) - (55)(55)}{\sqrt{((10)(385) - (55)^2)(10(385) - 55^2)}}$$

$$= \frac{3580 - 3025}{825} = \frac{555}{825}$$

$$\boxed{r_{xz} = 0.673}$$

$$r_{zy} = \frac{n\sum y_i z_i - (\sum y_i)(\sum z_i)}{\sqrt{(n\sum z_i^2 - (\sum z_i)^2)(n\sum y_i^2 - (\sum y_i)^2)}}$$

$$= \frac{10(344) - (55)(55)}{\sqrt{((10)(385) - (55)^2)(10(385) - 55^2)}}$$

$$= \frac{415}{825}$$

$$\boxed{r_{zy} = 0.503}$$

So, (X, Z) pair is best.

Q. 10.10

Y_i	6.5	5.3	8.6	1.2	4.2	2.9	1.1	3.0
X_i	3.2	2.7	4.5	1.0	2.0	1.7	0.6	1.9

Find Regression lines and the product moment co-efficient of correlation.

Sol:-

Y_i	X_i	Y_i^2	X_i^2	X_iY_i
6.5	3.2	42.25	10.24	20.8
5.3	2.7	28.09	7.29	14.31
8.6	4.5	73.96	20.25	38.7
1.2	1.0	1.44	1	1.2
4.2	2.0	17.64	4	8.4
2.9	1.7	8.41	2.89	4.93
1.1	0.6	1.21	0.36	0.66
3.0	1.9	9	3.61	5.7
32.8	17.6	182	49.64	94.7

$\bar{X} = \frac{\sum X_i}{n} = \frac{17.6}{8}$

$\bar{X} = 2.2$

$\bar{Y} = \frac{\sum Y_i}{n} = \frac{32.8}{8}$

$\bar{Y} = 4.1$

i) X on Y

$\hat{X} = a + bY_i$

$$b_{xy} = \frac{n\sum X_iY_i - (\sum X_i)\sum Y_i}{n\sum Y_i^2 - (\sum Y_i)^2}$$

$$= \frac{8(94.7) - (32.8)(17.6)}{8(182) - (32.8)^2}$$

$$= \frac{180.32}{380.16} = 0.47$$

$$a = \bar{X} - b\bar{Y}$$
$$= 2.2 - (0.47)(4.1)$$
$$\boxed{a = 0.27}$$

So, $\hat{X} = 0.27 + 0.47 Y_i$

(ii):- Y on X

$$\hat{Y} = a + bX_i$$

$$b_{yx} = \frac{n\sum X_iY_i - (\sum X_i)(\sum Y_i)}{n\sum X_i^2 - (\sum X_i)^2}$$

$$= \frac{8(94.7) - (32.8)(17.6)}{8(49.64) - (17.6)^2}$$

$$= \frac{180.32}{87.36}$$

$$\boxed{b_{yx} = 2.06}$$

$$a = \bar{Y} - b\bar{X}$$
$$= 4.1 - 2.06(2.2)$$
$$= -0.43$$

So, $\hat{Y} = -0.43 + 2.06 X_i$

(iii) $r = \pm\sqrt{b_{xy}\, b_{yx}}$

$$r = \sqrt{(0.47)(2.06)}$$

$$\boxed{r = 0.984}$$

22

15-12-14

(i) Y on X.

$$Y_i = a + bX_i + e_i$$

$$\hat{Y} = a + bX_i$$

$$e_i = Y_i - \hat{Y}$$

Least squares

a, b estimate $\quad \sum e_i^2 = \sum (Y_i - \hat{Y})^2$

$$= \sum (Y_i - a - bX_i)^2$$

Now taking partial derivatives w.r.t "a" and "b" equal to zero.

$$2\sum(Y - a - bX)(-1) = 0 \xrightarrow{\text{w.r.t } a} (i)$$

w.r.t b

$$2\sum(Y_i - a - bX_i)(-X_i) = 0 \longrightarrow (ii)$$

Now from eq. (i)

$$-2\sum(Y_i - a - bX_i) = 0$$

$$\sum(Y_i - a - bX_i) = 0$$

$$\sum Y_i - na - b\sum X_i = 0$$

$$\sum Y_i = na + b\sum X_i \longrightarrow (iii)$$

Multiply eq. (iii) with $\sum X_i$

$$\sum X_i \sum Y_i = na\sum X_i + b(\sum X_i)^2 \longrightarrow (iv)$$

Now from eq. (ii)

$$-2\sum(Y_i - a - bX_i)X_i = 0$$

$$\sum(X_i Y_i - X_i a - bX_i^2) = 0$$

$$\sum X_i Y_i - a\sum X_i - b\sum X_i^2 = 0$$

$$\sum X_i Y_i = a\sum X_i + b\sum X_i^2 \longrightarrow (V)$$

Multiplying eq. (V) with n.

$$n\sum X_i Y_i = na\sum X_i + nb\sum X_i^2 \longrightarrow (Vi)$$

By (Vi) − (iV).

$$n\sum X_i Y_i - \sum X_i \sum Y_i = nb\sum X_i^2 - b(\sum X_i)^2$$

$$n\sum X_i Y_i - \sum X_i \sum Y_i = b(n\sum X_i^2 - (\sum X_i)^2)$$

$$\boxed{\frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} = b}$$

eq (iii) Divided by n.

$$\frac{\sum Y_i}{n} = \frac{\cancel{n}a}{\cancel{n}} + b\frac{\sum X_i}{n}$$

$$\bar{Y} = a + b\bar{X}$$

$$\Rightarrow \boxed{a = \bar{Y} - b\bar{X}}$$

Q:- Derive "a" and "b" from least square method.

(ii): X on Y.

$$X_i = a + bY_i + e_i$$

$$\hat{X} = a + bY_i$$

$$\sum e_i^2 = \sum (X_i - \hat{X})^2$$

$$= \sum (X_i - a - bY_i)^2$$

Now taking partial derivatives w.r.t "a" & "b" equal to zero.

w.r.t a: $\rightarrow 2\sum (X_i - a - bY_i)(-1) = 0 \longrightarrow (i)$

w.r.t b. $\rightarrow 2\sum (X_i - a - bY_i)(-Y_i) = 0 \longrightarrow (ii)$

Now from (i).

$$-2\sum (X_i - a - bY_i) = 0$$

$$\sum (X_i - a - bY_i) = 0$$

$$\sum X_i - na - b\sum Y_i = 0$$

$$\sum X_i = na + b\sum Y_i \longrightarrow (iii)$$

xing by $\sum Y_i$.

$$\sum X_i \sum Y_i = na\sum Y_i + b(\sum Y_i)^2 \longrightarrow (iv)$$

Now from (ii).

$$-2\sum (X_i - a - bY_i) Y_i = 0$$

$$\sum (X_i Y_i - aY_i - bY_i^2) = 0$$

$$\sum X_i Y_i - a\sum Y_i - b\sum Y_i^2 = 0$$

$$\sum X_i Y_i = a\sum Y_i + b\sum Y_i^2$$

xing by n.

$$n\sum X_i Y_i = na\sum Y_i + nb\sum Y_i^2 \longrightarrow (v)$$

By (v) − (iv).

$$n\sum X_i Y_i - \sum X_i \sum Y_i = nb\sum Y_i^2 - b\left(\sum Y_i\right)^2$$

$$n\sum X_i Y_i - \sum X_i \sum Y_i = \left(n\sum Y_i^2 - \left(\sum Y_i\right)^2\right) b$$

$$\boxed{\frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum Y_i^2 - \left(\sum Y_i\right)^2} = b}$$

Dividing eq. (iii) by n.

$$\frac{\sum X_i}{n} = \frac{na}{n} + b\frac{\sum Y_i}{n}$$

$$\bar{X} = a + b\bar{Y}$$

$$\Rightarrow \boxed{a = \bar{X} - b\bar{Y}}$$

also try to Prove:– for Y on X.

$$b = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \quad \& \quad a = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2}$$

Example : 10.3

$\hat{Y} = 1.47 + 2.831X_i$ & Show that $\Sigma(Y_i - \hat{Y}) = 0$

X_i	5	6	8	10	12	13	15	16	17
Y_i	16	19	23	28	36	41	44	45	50

Sol:-

X_i	Y_i	$\hat{Y}$	$Y_i - \hat{Y}$
5	16	15.625	0.375
6	19	18.456	0.544
8	23	24.118	-1.118
10	28	29.78	-1.78
12	36	35.442	0.558
13	41	38.273	2.727
15	44	43.935	0.065
16	45	46.766	-1.766
17	50	49.597	0.403
			0.008

Q:-

X_i	78	89	97	69	59	79	68	61
Y_i	125	137	156	112	107	136	123	108

Show that $\Sigma(Y_i - \hat{Y}) = 0$

Sol.

X_i	Y_i	X_iY_i	X_i^2
78	125	9750	6084
89	137	12193	7921
97	156	15132	9409
69	112	7728	4761
59	107	6313	3481
79	136	10744	6241
68	123	8364	4624
61	108	6588	3721
600	1004	76812	46242

$\bar{x} = \frac{\sum X_i}{n} = \frac{600}{8} = 75$

$\bar{y} = \frac{\sum Y_i}{n} = \frac{1004}{8} = 125.5$

$$b = \frac{n \sum X_iY_i - (\sum X_i)(\sum Y_i)}{n \sum X_i^2 - (\sum X_i)^2}$$

$$= \frac{8(76812) - (600)(1004)}{8(46242) - (600)^2}$$

$$= \frac{12096}{9936} = 1.218$$

$$a = \bar{y} - b\bar{x}$$

$$= 125.5 - 1.218(75)$$

$$= 34.15$$

$$\Rightarrow \hat{y} = 34.15 + 1.218\, X_i$$

X_i	Y_i	$\hat{Y}$	$Y_i - \hat{Y}$
78	125	129.154	-4.154
89	137	142.552	-5.552
97	156	152.296	3.704
69	112	118.192	-6.192
59	107	106.012	0.988
79	136	130.372	5.628
68	123	116.974	6.026
61	108	108.448	-0.448

$\Sigma(Y_i - \hat{Y}) = 0$

(23)

Quality Control 18-12-2014

Quality control is a process by which entities review the quality of all factors envolves in production. ISO-9000 (international standradized organization) define quality control "A part of quality management focus on fullfilling quality requirment".

Three Basic Aspacts:

1:- Elements (Such as

control job management, define well manage process)

2:- Competence (Such as knowledge, skills, experience & qualifications).

3:- Soft Element (Such as personal organizational culture, motivation, team spirit and quality relationship.)

Control includes product inspaction where every product is examined visually and often using a stereo microscope for fine detail before the product is sold into external market.

Inspector will be provided with list and discription of unacceptable product defects such as crack on surface blemishes.

The quality of output is at risk if any of these three aspect is deficient in anyway.

24

Statistical Process Control (SPC) 12-1-15

It is a method of quality control which uses statistical method. S.P.C. applied in order to monitor and control the process. Monitring and controling the process ensure that it operates at it's full potension. It full potension the process can make as much confirming product as possible with minimum of waste.

SP.C can be applied to any process where the confirming product output can be measured. S.P.C is a statistical method of separating variation, resulting from special causes to natural variation and to establish and maintain consistancy in the process, enabling process improvement (Davis 2003)

Common Cause Variation

The variation which inharent in the process itself.

Special Cause Variation

The variation in process output that might be trace to a specific cause. The process is said to be out of control when special cause variation exists.

(5)

SP.C Toolbelt 15-1-15

(i):- Histogram

(ii):- Boxplot (or Scatter diagram)

(iii):- Check sheet

(iv):- Defect construction diagram or Flow chart

(v):- Pareto chart

(vi):- Cause and effect

(vii):- Control chart

"Classification of control chart"

(1):- Variable Control chart

2):- Attribute control chart

3):- Univariant Control chart

4):- Multivariant Control chart

5):- Parametric Control chart

6):- Non-Parametric Control chart

7):- Classical C.C 8):- Baysian C.C

9):- Memory Control Chart

10):- Memoryless Control chart.

56

Control Chart 19-1-15

A control chart contains time or sample number on the horizontal axis and some quality characteristics on vertical axis. Three decision lines are added to the structure which allow us to decide whether the process is working under incontrol or outcontrol situation.

These three lines are named as:

(i):- Upper Control limit

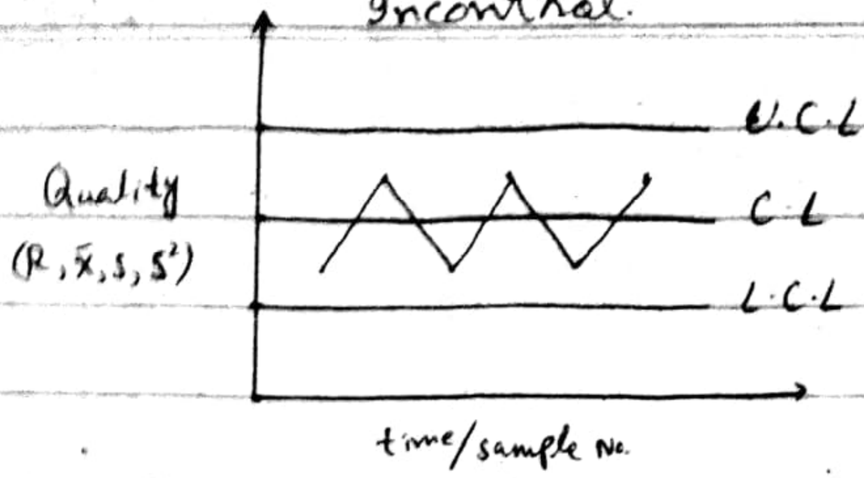
(ii):- Central Limit (Intermediat Limit)

(iii):- Lower Control Limit

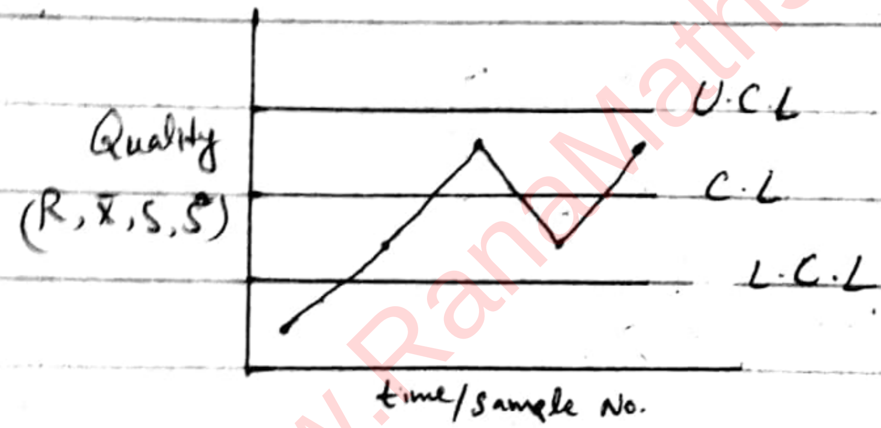
As long as the working within the control limit the process is declared to be **incontrol** but when the statistics $(\bar{X}, S, S^2$ & $R)$ falls outside these lines/limits then the process is said to be outcontrol.

These control limits are generally reffered as 3-sigma limits.

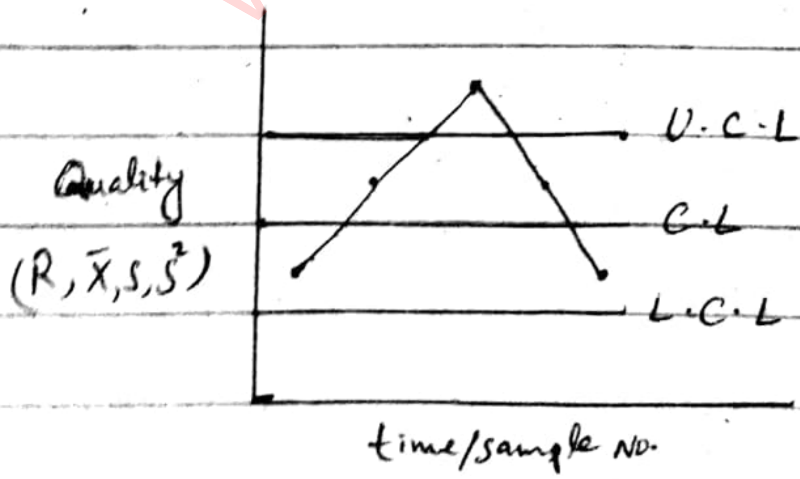
Incontrol.



(Outcontrol)



(Outcontrol)



Solⁿ Measure of dispersion & measure of central tendency.

(27)

21-01-15

R-chart, S-chart, S^2-chart $\longrightarrow$ despension

$\bar{X}$-chart $\longrightarrow$ location.

R-chart : when σ known

$U.C.L = (d_2 + 3d_3)\sigma_0$

$C.L = d_2 \sigma_0$

d_2, d_3 are constant.

$L.C.L = (d_2 - 3d_3)\sigma_0$

Q:-

Sample No.	Sub-Samples.	Range
1	1.3, 1.6, 1.3, 1.4, 1.5	0.3
2	1.5, 1.4, 1.4, 1.7, 1.6	0.3
3	1.6, 1.5, 1.7, 1.2, 1.4	0.5
4	1.3, 1.3, 1.4, 1.6, 1.3	0.3
5	1.6, 1.2, 1.5, 1.2, 1.6	0.4
6	1.7, 1.4, 1.6, 1.5, 1.7	0.3
7	1.3, 1.7, 1.6, 1.3, 1.5	0.4
8	1.2, 1.4, 1.5, 1.3, 1.5	0.3
9	1.4, 1.4, 1.7, 1.5, 1.3	0.4
10	1.5, 1.6, 1.5, 1.4, 1.7	0.3

$d_2 = 2.236$

$d_3 = 0.74$

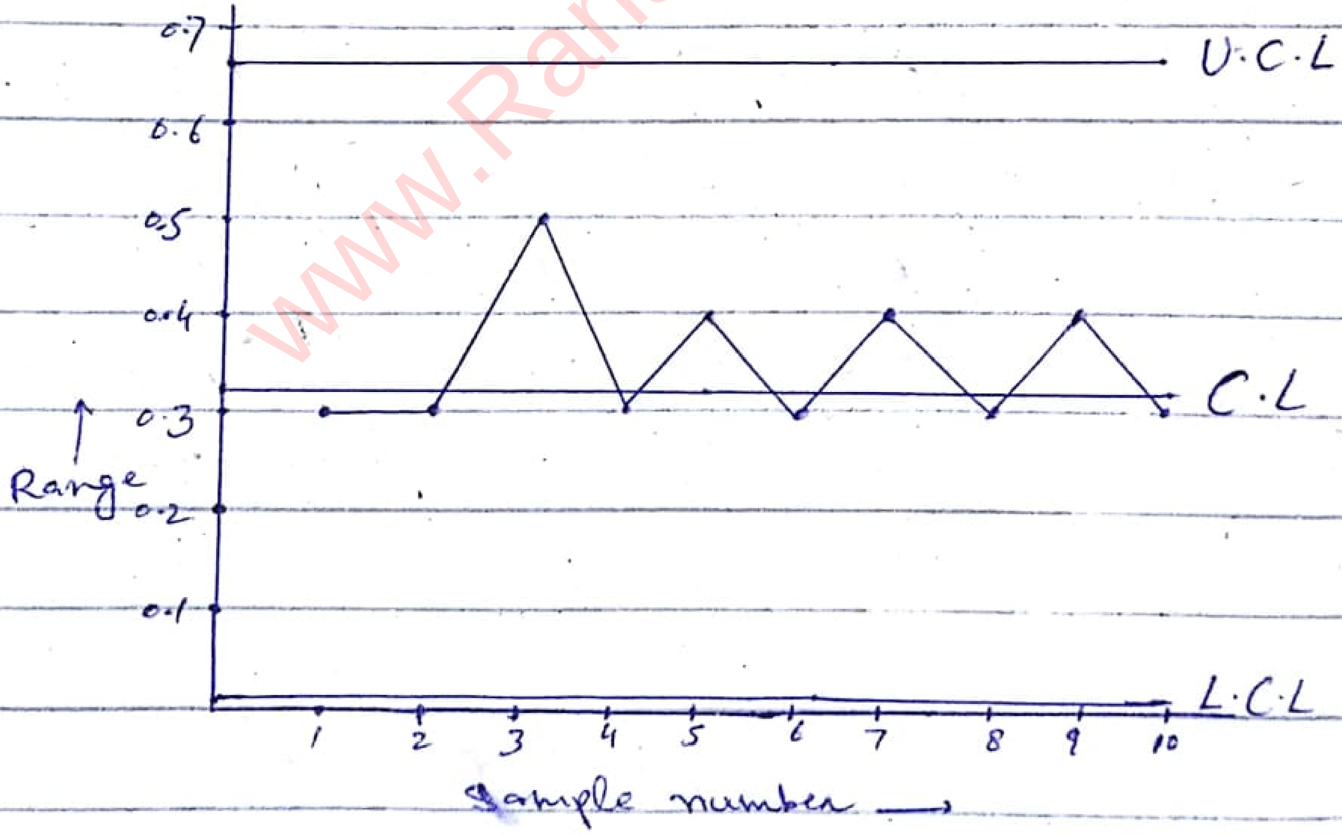
$\sigma = 0.15$

Solution:-

$$U.C.L = (d_2 + 3d_3)\sigma_0$$
$$= (2.236 + 3(0.74))\,0.15$$
$$= 0.6684$$

$$C.L = d_2\,\sigma_0$$
$$= (2.236)(0.15)$$
$$= 0.3354$$

$$L.C.L = (d_2 - 3d_3)\,\sigma_0$$
$$= (2.236 - 3(0.74))(0.15)$$
$$= 0.0024$$



Result:- Process is incontrol.

When σ is unknown

$$U.C.L = (d_2 + 3d_3)\hat{\sigma}$$

$$C.L = d_2\hat{\sigma}$$

$$L.C.L = (d_2 - 3d_3)\hat{\sigma}$$

$$\hat{\sigma} = \frac{\bar{R}}{d_2}$$

$$\bar{R} = \frac{\Sigma R}{m}$$

Q1- $d_2 = 2.236$

$d_3 = 0.74$

Sample No.	X_i	Range
1	1.3, 1.6, 1.3, 1.4, 1.5	0.3
2	1.5, 1.4, 1.4, 1.7, 1.6	0.3
3	1.6, 1.5, 1.7, 1.2, 1.4	0.5
4	1.3, 1.3, 1.4, 1.6, 1.3	0.3
5	1.6, 1.2, 1.5, 1.2, 1.6	0.4
6	1.7, 1.4, 1.6, 1.5, 1.7	0.3
7	1.3, 1.7, 1.6, 1.3, 1.5	0.4
8	1.2, 1.4, 1.5, 1.3, 1.5	0.3
9	1.4, 1.4, 1.7, 1.5, 1.3	0.4
10	1.5, 1.6, 1.5, 1.4, 1.7	0.3
		3.5

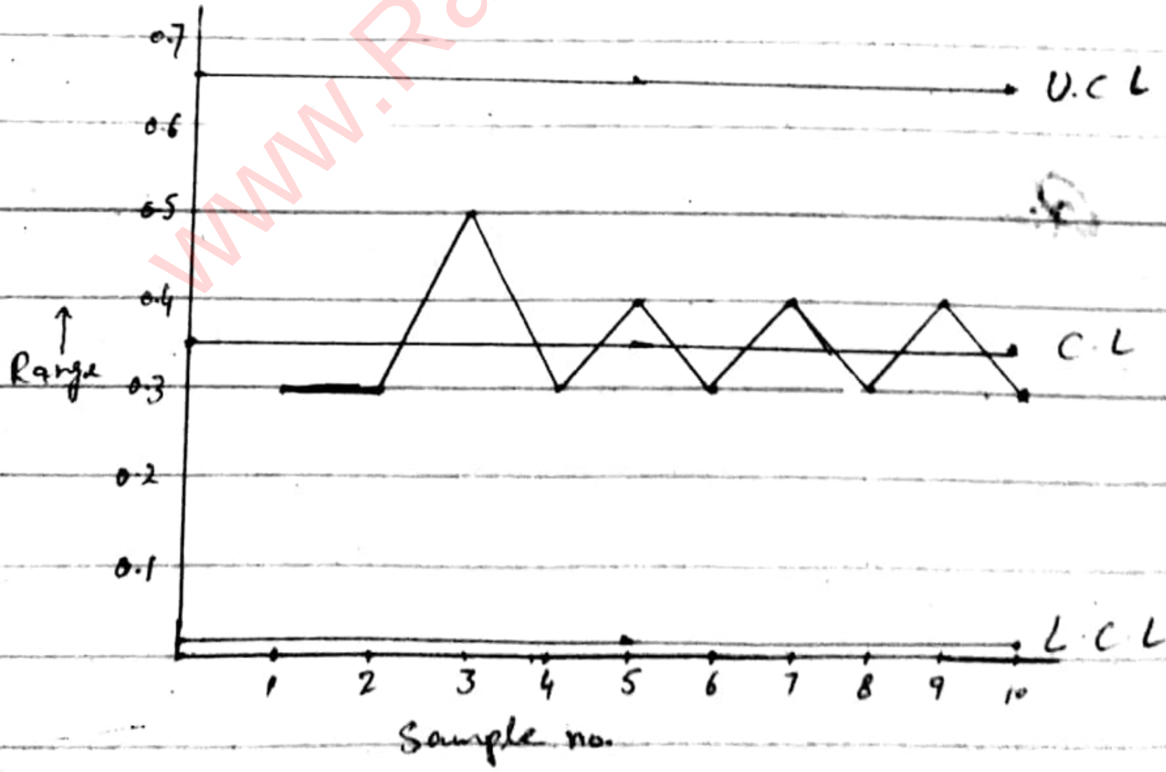
$$\bar{R} = \frac{\Sigma R}{m} = \frac{3.5}{10} = 0.35$$

$$\hat{\sigma} = \frac{\bar{R}}{d_2} = \frac{0.35}{2.236} = 0.156$$

$$U.C.L = (d_2 + 3d_3)\hat{\sigma}$$
$$= (2.236 + 3(0.74))(0.156)$$
$$= 0.695$$

$$C.L = d_2 \hat{\sigma}$$
$$= (2.236 \times 0.156)$$
$$= 0.349$$

$$L.C.L = (d_2 - 3d_3)\hat{\sigma}$$
$$= (2.236 - 3(0.74))(0.156)$$
$$= 0.0025$$



Result:- Process in incontrol.

$\bar{X}$-Chart $\longrightarrow$ For location

1):- When σ & μ are known.

$$U.C.L = \mu_0 + 3\frac{\sigma_0}{\sqrt{n}}$$

$$C.L = \mu_0$$

$$L.C.L = \mu_0 - 3\frac{\sigma_0}{\sqrt{n}}$$

2):- When σ & μ are unknown.

$$U.C.L = \bar{\bar{X}} + 3\frac{\hat{\sigma}}{\sqrt{n}}$$

$$C.L = \bar{\bar{X}}$$

$$L.C.L = \bar{\bar{X}} - 3\frac{\hat{\sigma}}{\sqrt{n}}$$

$$\bar{\bar{X}} = \frac{\Sigma \bar{X}}{m}$$

$$\hat{\sigma} = \frac{\bar{R}}{d_2}$$

$$\bar{R} = \frac{\Sigma R}{m}$$

22-01-15

Q:- $\mu = 1.57$, $\sigma = 0.24$ } known μ & σ

$n = 5$

$$U.C.L = \mu_0 + 3\frac{\sigma_0}{\sqrt{n}}$$

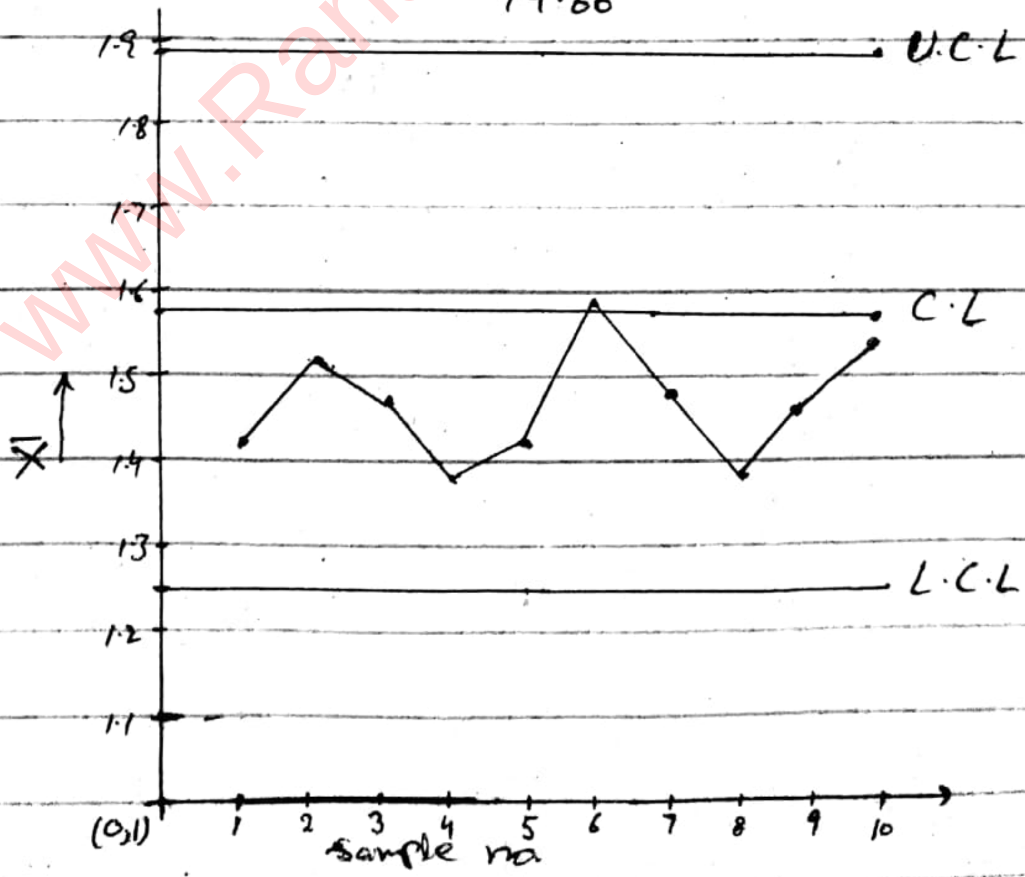
$$= 1.57 + 3\left(\frac{0.24}{\sqrt{5}}\right) = 1.892$$

$$C.L = \mu_0 = 1.57$$

$$L.C.L = \mu_0 - 3\frac{\sigma_0}{\sqrt{n}} = 1.57 - 3\left(\frac{0.24}{\sqrt{5}}\right)$$

$$= 1.248$$

Sample No.	x_i	$\bar{X}$
1	1.3, 1.6, 1.3, 1.4, 1.5	1.42
2	1.5, 1.4, 1.4, 1.7, 1.6	1.52
3	1.6, 1.5, 1.7, 1.2, 1.4	1.48
4	1.3, 1.3, 1.4, 1.6, 1.3	1.38
5	1.6, 1.2, 1.5, 1.2, 1.6	1.42
6	1.7, 1.4, 1.6, 1.5, 1.7	1.58
7	1.3, 1.7, 1.6, 1.3, 1.5	1.48
8	1.2, 1.4, 1.5, 1.3, 1.5	1.38
9	1.4, 1.4, 1.7, 1.5, 1.3	1.46
10	1.5, 1.6, 1.5, 1.4, 1.7	1.54
		14.66



Result:- Process is in control.

Q:- When σ & μ are unknown.

Sample No.	Subsamples	Range	$\bar{X}$
1	1.3, 1.6, 1.3, 1.4, 1.5	0.3	1.42
2	1.5, 1.4, 1.4, 1.7, 1.6	0.3	1.52
3	1.6, 1.5, 1.7, 1.2, 1.4	0.5	1.48
4	1.3, 1.3, 1.4, 1.6, 1.3	0.3	1.38
5	1.6, 1.2, 1.5, 1.2, 1.6	0.4	1.42
6	1.7, 1.4, 1.6, 1.5, 1.7	0.3	1.58
7	1.3, 1.7, 1.6, 1.3, 1.5	0.4	1.48
8	1.2, 1.4, 1.5, 1.3, 1.5	0.3	1.38
9	1.4, 1.4, 1.7, 1.5, 1.3	0.4	1.46
10	1.5, 1.6, 1.5, 1.4, 1.7	0.3	1.54
		3.5	14.66

$d_2 = 2.236 \qquad \bar{R} = \frac{\Sigma R}{m} = \frac{3.5}{10} = 0.35$

$$\hat{\sigma} = \frac{\bar{R}}{d_2} = \frac{0.35}{2.236} = 0.156$$

$$\bar{\bar{X}} = \frac{\Sigma \bar{X}}{m} = \frac{14.66}{10} = 1.466$$

$$U.C.L = \bar{\bar{X}} + 3\frac{\hat{\sigma}}{\sqrt{m}}$$

$$= 1.466 + 3\left(\frac{0.156}{\sqrt{5}}\right)$$

$$= 1.675$$

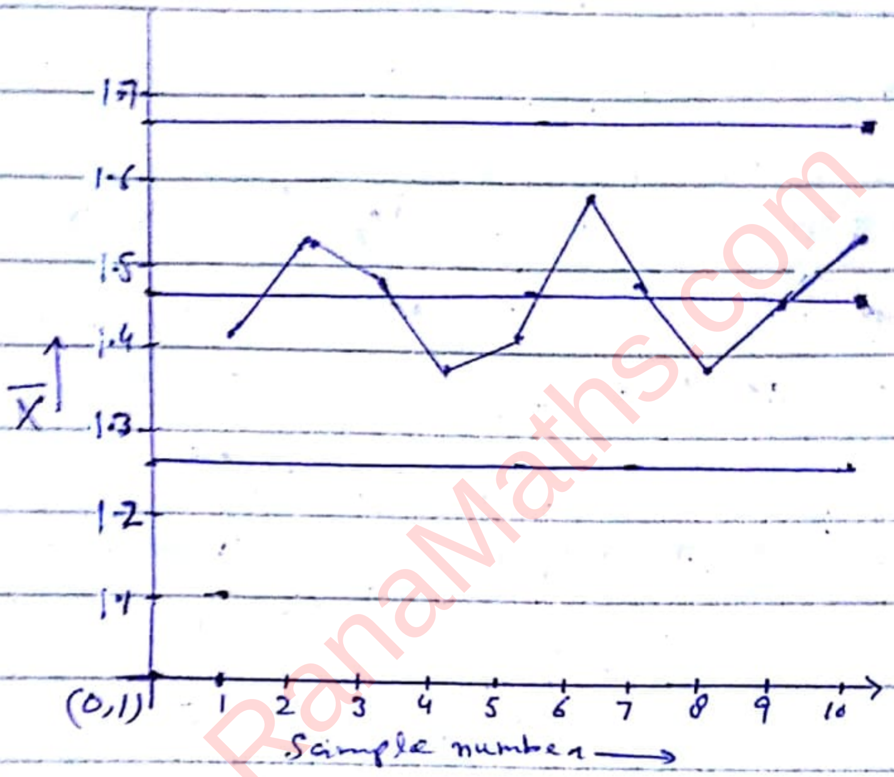
$$C.L = \bar{\bar{X}}$$

$$= 1.466$$

$$L.C.L = \bar{\bar{X}} - 3\left(\frac{\hat{\sigma}}{\sqrt{n}}\right)$$

$$= 1.466 - 3\left(\frac{0.156}{\sqrt{5}}\right)$$

$$= 1.257$$



Result:- Process is incontrol.

29

S-chart

26-1-15

① When σ is known.

$$U.C.L = \left(C_4 + 3\sqrt{1-C_4^2}\right)\sigma_0$$

$$C.L = C_4\sigma_0$$

$$L.C.L = \left(C_4 - 3\sqrt{1-C_4^2}\right)\sigma_0$$

② When σ is unknown.

$$U.C.L = \left(C_4 + 3\sqrt{1-C_4^2}\right)\hat{\sigma}$$

$$C.L = C_4\hat{\sigma}$$

$$L.C.L = \left(C_4 - 3\sqrt{1-C_4^2}\right)\hat{\sigma}$$

$$\bar{S} = \frac{\sum S}{m}$$

$$\hat{\sigma} = \frac{\bar{S}}{c_4}$$

Q:-

Sample No.	Sub sample	$S_j = \sqrt{\frac{1}{n-1}\left(\sum (x_i - \bar{x})^2\right)}$	$\bar{X}$
1	1.3, 1.6, 1.3, 1.4, 1.5	0.13038 4048	1.42
2	1.5, 1.4, 1.4, 1.7, 1.6	0.13038 4048	1.52
3	1.6, 1.5, 1.7, 1.2, 1.4	0.19235384	1.48
4	1.3, 1.3, 1.4, 1.6, 1.3	0.130384048	1.38
5	1.6, 1.2, 1.5, 1.2, 1.6	0.20493 9015	1.42
6	1.7, 1.4, 1.6, 1.5, 1.7	0.130384048	1.58
7	1.3, 1.7, 1.6, 1.3, 1.5	0.178885438	1.48
8	1.2, 1.4, 1.5, 1.3, 1.5	0.130384048	1.38
9	1.4, 1.4, 1.7, 1.5, 1.3	0.151657508	1.46
10	1.5, 1.6, 1.5, 1.4, 1.7	0.114017542	1.54

$$S_j = \sqrt{\frac{1}{n-1}\sum (X_i - \bar{X})^2} \qquad i = 1, \ldots, n. \quad j = 1, \ldots, m.$$

$$\Rightarrow S_1 = \sqrt{\tfrac{1}{4}\left[(1.3-1.42)^2 + (1.6-1.42)^2 + (1.3-1.42)^2 + (1.4-1.42)^2 + (1.5-1.42)^2\right]}$$

$$S_1 = \sqrt{\tfrac{1}{4}(0.068)} = \sqrt{0.017}$$

$$\boxed{S_1 = 0.13038}$$

$$\Rightarrow S_2 = \sqrt{\frac{1}{4}\left[(1.5-1.52)^2+(1.4-1.52)^2+(1.4-1.52)^2+(1.7-1.52)^2+(1.6-1.52)^2\right]}$$

$$S_2 = \sqrt{\frac{1}{4}(0.068)} = \sqrt{0.017}$$

$$\boxed{S_2 = 0.13038}$$

$$\Rightarrow S_3 = \sqrt{\frac{1}{4}\left[(1.6-1.48)^2+(1.5-1.48)^2+(1.7-1.48)^2+(1.2-1.48)^2+(1.4-1.48)^2\right]}$$

$$S_3 = \sqrt{\frac{1}{4}(0.148)} = \sqrt{0.037}$$

$$\boxed{S_3 = 0.19235}$$

$$\Rightarrow S_4 = \sqrt{\frac{1}{4}\left[(1.3-1.38)^2+(1.3-1.38)^2+(1.4-1.38)^2+(1.6-1.38)^2+(1.3-1.38)^2\right]}$$

$$S_4 = \sqrt{\frac{1}{4}(0.068)} = \sqrt{0.017}$$

$$\boxed{S_4 = 0.13038}$$

$$\Rightarrow S_5 = \sqrt{\frac{1}{4}\left[(1.6-1.42)^2+(1.2-1.42)^2+(1.5-1.42)^2+(1.2-1.42)^2+(1.6-1.42)^2\right]}$$

$$S_5 = \sqrt{\frac{1}{4}(0.168)} = \sqrt{0.042}$$

$$\boxed{S_5 = 0.2049}$$

$$\Rightarrow S_6 = \sqrt{\frac{1}{4}\left[(1.7-1.58)^2+(1.4-1.58)^2+(1.6-1.58)^2+(1.5-1.58)^2+(1.7-1.58)^2\right]}$$

$$S_6 = \sqrt{\frac{1}{4}(0.068)} = \sqrt{0.017}$$

$$\boxed{S_6 = 0.13038}$$

$$\Rightarrow S_7 = \sqrt{\frac{1}{4}\left[(1.3-1.48)^2+(1.7-1.48)^2+(1.6-1.48)^2+(1.3-1.48)^2+(1.5-1.48)^2\right]}$$

$$\boxed{S_7 = 0.1789}$$

$\Rightarrow S_8 = \sqrt{\frac{1}{4}\left[(1.2-1.38)^2+(1.4-1.38)^2+(1.5-1.38)^2+(1.3-1.38)^2+(1.5-1.38)^2\right]}$

$S_8 = \sqrt{\frac{1}{4}(0.068)} = \sqrt{0.017}$

$\boxed{S_8 = 0.13038}$

$\Rightarrow S_9 = \sqrt{\frac{1}{4}\left[(1.4-1.46)^2+(1.4-1.46)^2+(1.7-1.46)^2+(1.5-1.46)^2+(1.3-1.46)^2\right]}$

$S_9 = \sqrt{\frac{1}{4}(0.092)} = \sqrt{0.023}$

$\boxed{S_9 = 0.1516}$

$\Rightarrow S_{10} = \sqrt{\frac{1}{4}\left[(1.5-1.54)^2+(1.6-1.54)^2+(1.5-1.54)^2+(1.4-1.54)^2+(1.7-1.54)^2\right]}$

$S_{10} = \sqrt{\frac{1}{4}(0.052)} = \sqrt{0.013}$

$\boxed{S_{10} = 0.114}$

(i):— When σ is known

$C_4 = 0.9937$

$\sigma = 0.24$

$U.C.L = \left(C_4 + 3\sqrt{1-C_4^2}\right)\sigma_0$

$= \left(0.9937 + 3\sqrt{1-(0.9937)^2}\right)(0.24)$

$= 0.32$

$C.L = C_4\sigma_0$

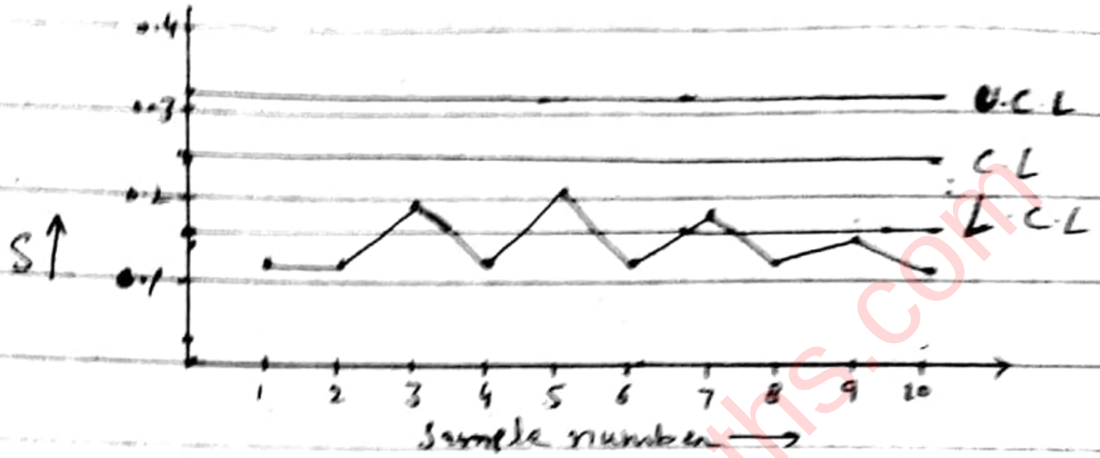
$= (0.9937)(0.24)$

$= 0.24$

$L.C.L = \left(C_4 - 3\sqrt{1-C_4^2}\right)\sigma_0$

$$L.C.L = \left(0.9937 - 3\sqrt{1-(0.9937)^2}\right)(0.24)$$

$$= 0.16$$



Result:- Process is outcontrol.

(ii):- σ is unknown.

$$\bar{S} = \frac{\Sigma S}{m} = \frac{1.494}{10} = 0.1494$$

$$C = 0.9937$$

$$\hat{\sigma} = \frac{\bar{S}}{C_4} = \frac{0.1494}{0.9937} = 0.15035$$

$$U.C.L = \left(C_4 + 3\sqrt{1-C_4^2}\right)\hat{\sigma}$$

$$= \left(0.9937 + 3\sqrt{1-0.9937^2}\right)(0.15035)$$

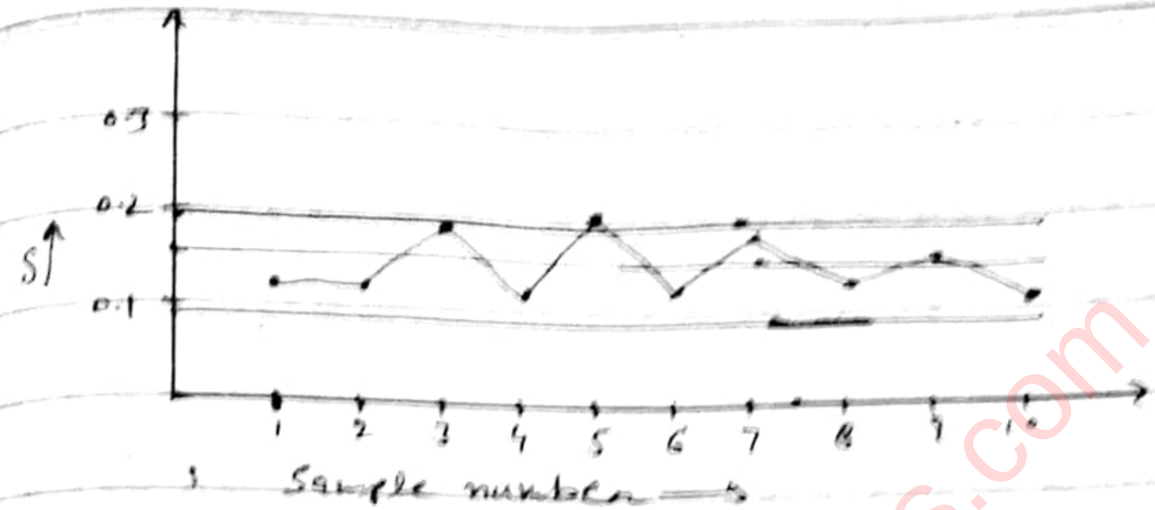
$$= 0.19995$$

$$C.L = C_4\hat{\sigma} = 0.149$$

$$L.C.L = \left(C_4 - 3\sqrt{1-C_4^2}\right)\hat{\sigma}$$

$$= \left(0.9937 - 3\sqrt{1-0.9937^2}\right)(0.15035)$$

$$= 0.0988$$



Result:- Process is outcontrol