Mean:

Arithmetic Mean is called average. It is used to measure of central tendency.

There are two types of A Mean

① Simple Mean ② Weighted Mean

Simple Mean:

It is equal to sum of variables divided by their number.

a) Direct Method

$$\text{Mean} = \bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{N}$$

$$\bar{X} = \frac{\Sigma X}{N}$$ ΣX → The Sum of variable; N → No. of Observation.

b) Shortcut Method:

S1 Assume any one value as assumed mean (A)

S2 Find out diffⁿ of each value from assumed mean i.e ($d = X - A$).

S3 Add all deviations (Σd)

S4 Apply formula

$$\boxed{\bar{X} = A \pm \frac{\Sigma d}{N}}$$

(2)

Page No.

Date.

Weighted Arithmetic Mean:

It is defined as an average whose component items are multiplied by certain value (weight) and the aggregate of the products are divided by total of weights.

$$\boxed{\bar{X} = \frac{N_1 X_1 + N_2 X_2}{N_1 + N_2}}$$ Doubt

Eg Calculate Mean

Profit per shop (Rs)	No. of shops
100 - 200	10
200 - 300	18
300 - 400	20
400 - 500	26
500 - 600	30
600 - 700	28
700 - 800	18

Sol: a) Direct Method:

Profit	Mid Point $\left(\frac{U+L}{2}\right)(x)$	No. of shop (f)	(f.x)
100 - 200	150	10	1500
200 - 300	250	18	4500
300 - 400	350	20	7000
400 - 500	450	26	11,700

500-600	550	30	16,500
600-700	650	28	18,200
700-800	750	18	13,500
		$N = \Sigma f = 150$	$\Sigma fx = 72,900$

$$\bar{x} = \frac{\Sigma fx}{\Sigma f}$$

$$= \frac{72,900}{150} = \underline{\underline{486}} \text{ Ans}$$

b) <u>Short Cut Method</u>:

(A = 450)

<u>Profit</u>	<u>x</u>	d = x - A	f	fd
100-200	150	-300	10	-3000
200-300	250	-200	18	-3600
300-400	350	-100	20	-2000
400-500	450	0	26	0
500-600	550	100	30	3000
600-700	650	200	28	5600
700-800	750	300	18	5400
			$\Sigma f = 150$	$\Sigma fd = 5400$

$$\bar{x} = A \pm \frac{\Sigma fd}{N \rightarrow \Sigma f}$$

$$= 450 + \frac{5400}{150}$$

$$= 450 + 36$$

$$= \underline{\underline{486}} \text{ Ans}$$

(4)

Page No.

Date

c) Step-Deviation Method :

S1 Find out mid value of each group (x)
S2 Assume any one of mid value as average (A)
S3 Find out deviation of midvalue from A (d)
S4 Deviation are divided by common factor (d')
S5 Multiply d' of each class by frequency (fd')
S6 Add up the product of S5 i.e ($\Sigma fd'$)
S7 Apply formula

$$\boxed{\bar{X} = A \pm \frac{\Sigma fd'}{N} \times C} \rightarrow \text{Step deviation}$$

In above eg :-

Profit	x	f	d ($A=450$)	d' ($\frac{x-450}{100}$)	fd'
100-200	150	10	-300	-3	-30
200-300	250	18	-200	-2	-36
300-400	350	20	-100	-1	-20
400-500	450	26	0	0	0
500-600	550	30	100	1	30
600-700	650	28	200	2	56
700-800	750	18	3000	3	54
		$\Sigma f = 150$			$\Sigma fd' = 54$

$$\bar{X} = \frac{\Sigma fd'}{N} \times C + A$$

$$= 450 + \frac{54}{150} \times 100$$

$$= 450 + 36 = \underline{\underline{486}} \quad \text{Ans}$$

MEDIAN :

It is defined as value of that item which divides the series into two equal parts, one half containing values greater then it and other half containing values less than it.

Eg Find out Median of following item :

~~Size of item~~ 10, 15, 9, 25, 19

Sol: Arrange them in ascending Order

9, 10, 15, 19, 25

In case of odd items

$$\boxed{\text{Median} = \text{Size of } \left[\frac{(N+1)}{2}\right]^{th} \text{ item}}$$

$$\text{Median} = \text{Size of } \left(\frac{5+1}{2}\right)^{th} \text{ item}$$

$$= 3^{rd} \text{ item}$$

$$= \underline{\underline{15}} \quad \text{Ans}$$

6

Page No.......... Date..........

∴ In case of Even item,

$$\text{Median} = \text{Size of } \left(\frac{\left(\frac{n}{2}\right)^{th} \text{item} + \left(\frac{n}{2}+1\right)^{th} \text{item}}{2} \right)$$

Eg: Find Median of following

57, 58, 61, 42, 38, 65, 72, 66

Sol:

38, 42, 57, 58, 61, 65, 66, 72

$$\text{Median} = \frac{4^{th} + 5^{th}}{2} \Rightarrow \frac{58+61}{2}$$

$$= \underline{\underline{59.5}} \text{ Ans.}$$

∴ Median of Discrete Series:

S1 Arrange the data is ascending order.
S2 Find cummulative frequency.
S3 Apply formula

odd ← $\boxed{\text{Med} = \text{Size of } \left(\frac{N+1}{2}\right)^{th} \text{ item.}}$

even ← $\boxed{= \text{Size of } \left(\frac{N}{2}\right)^{th} \text{ item}}$

Eg **Locate Median :**

Size of Shoes	frequency
5	10
5.5	16
6	28
6.5	15
7	30
7.5	40
8	34

Sol:

Size of Shoes	f	cf
5	10	10
5.5	16	26
6	28	54
6.5	15	69
7	30	99
7.5	40	139
8	34	173

Median = Size of $\left(\frac{N+1}{2}\right)^{th}$ item

= Size of $\left(\frac{173+1}{2}\right)^{th}$ item

= 87^{th} item

= $\boxed{7}$ Ans

Page No.
Date.

Median for Continous Series:

S1 Find out median by using $\frac{N}{2}$.

S2 Find out the class in which median lies.

S3 Apply formula:

$$\boxed{\text{Median} = L_1 + \frac{\frac{N}{2} - C.F}{f} \times i}$$

where

L_1 = Lower limit of Median Class.
f = frequency of Median Class
$C.F$ = Cummulative frequency of the class preeceding median class.
i = Class interval of median class

Eg. Calculate Median:

Marks	Frequency
10-25	6
25-40	20
40-55	44
55-70	26
70-85	3
85-100	1

Sol:

Marks (x)	f	C.F
10-25	6	6
25-40	20	26
40-55	44	70
55-70	26	96
70-85	3	99
85-100	1	100

Median item $= \frac{N}{2} = \frac{100}{2} = 50$

Median lies in 40-55 mark group. ~~Median is~~

Median Class = 40-55

$$\text{Median} = L_1 + \frac{\frac{N}{2} - C.F}{f} \times i$$

$L_1 = 40$, $\frac{N}{2} = 50$, $C.F = 26$, $f = 44$

$i = 5$ [55-40]

$$\text{Median} = 40 + \left(\frac{50-26}{44}\right) \times 5$$

$= 40 + 8.18 = 48.18$ marks Ans

Alternative

taking mid point

$$\boxed{\text{Median} = L_1 + \frac{L_2 - L_1}{f}(\bar{x} - C.F)}$$

(10)

Page No.

Date

Mode:

Mode is the value which occurs the greatest no. of frequency in a series.

$$\boxed{Mode = L_1 + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i\right)}$$

☆ $\boxed{Mode = 3\,Median - 2\,Mean}$

Dispersion:

Dispersion is measure of the variation of the items. i.e the degree to which numerical data tend to spread about an avg. value.

Deviation = $X - \bar{X}$.

∴ The measure of central tendency - mean, median & mode are based on real values of items of a series and are called Averages of First Order. The measure of variation are called Average of Second Order.

Purpose of Measuring Variation:

1. To test the reliability of an avg.
2. To serve as a basis of control of variability.
3. To compare two or more series with regard to their variability.
4. To facilitate as a basis for further statistical analysis.

Methods of Measuring Dispersion:

(1) Range (2) Inter - Quartile Range (3) Mean Deviation

(12)

Page No.

Date.

(4) Standard Deviation (5) Lorenz Curve.

Mean Deviation:

It is the arithmetic mean of the deviations of a series computed from any measure of central tendency; all deviation are taken as (+) ve.

$$M.D \text{ from } \bar{X} = \frac{\Sigma |D|}{N}$$

→ For Individual Series:

S1 : Calculate the avg. mean, median or mode of Series.

S2: Take deviation of items from average, ignoring (+) ve & (-) ve signs.

S3 : Compute the total of these deviation ΣD.

S4 : Divide this total by no. of items.

Eg: Calculate mean deviation from mean for following

100, 150, 200, 250, 360, 490, 500, 600, 671

Sol:

X	$\lvert D \rvert = X - \bar{X}$
100	269
150	~~169~~ 219
200	~~148~~ 169
250	~~8~~ 119

360	~~121~~ 9
490	~~131~~ 121
500	~~23~~ 131
600	231
671	302
$\Sigma x = 3321$	$\Sigma \lvert D \rvert = 1570$

$$\bar{x} = \frac{\Sigma x}{N} = \frac{3321}{9} = 369$$

$$M.D = \frac{\Sigma |D|}{N} = \frac{1570}{9} = 174.44 \text{ Ans}$$

<u>Coefficient</u> of <u>M.D</u>

$$= \frac{M.D}{Mean}$$

Hence, in above eg

$$\text{Coefficient of } M.D = \frac{174.44}{369}$$

$$= 0.47 \text{ Ans}$$

For Discrete Series:

S1: Find out average (mean, median, mode).

S2: Find out deviation. |D|.

S3: Multiply the deviation by its respective frequency ~~&~~ total them i.e $\Sigma f|D|$.

S4: Divide the total by ~~total~~ frequency

(14) Page No. Date.

$$M.D = \frac{\Sigma f|D|}{N}$$

Eg: Calculate mean deviation from mean:

Class Interval: 2-4 4-6 6-8 80-10

frequency : 3 4 2 1

Sol:

Class	x	f	fx	D ($X-\bar{X}$)	$f\|D\|$
2-4	3	3	9	2.2	6.6
4-6	5	4	20	0.2	0.8
6-8	7	2	14	1.8	3.6
8-10	9	1	9	3.8	3.8
		$N=10$	$\Sigma fx = 52$		$\Sigma f\|d\| = 14.8$

$\bar{x} = \frac{52}{10} = 5.2$

$$M.D = \frac{\Sigma f|D|}{N}$$

$$= \frac{14.8}{10} = 1.48 \text{ Ans.}$$

Standard Deviation: (σ)

It is defined as positive square root of the arithmetic mean of the square of deviation of given observation from their arithmetic mean.

It is denoted by Sigma (σ).

∴ There are two methods for calculating S.D
a) → Deviation from actual mean
b) → Deviation from assumed mean.

a) This method is taken when the mean is a whole number.

S1 Find out actual mean ($\bar{x}$)
S2 Find out deviation value from mean ($x-\bar{x}$)
S3 Square the deviation & take sum. Σx^2.
S4 Divide the total (Σx^2) by no. of observations. The square root of quotient is S.D

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$$

b) This method is adopted when the mean is ~~a~~ fractional value.

S1 Assume anyone of item as average (A).
S2 Find out deviation from assumed mean $X-A$.
S3 Find out total of deviation = Σd.
S4 Square the deviation i.e d^2 & add them i.e Σd^2
S5 Formual

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

Page No.........

Date.........

→ For Discrete Series:

a) Actual Mean Method:

S1 Calculate the mean of series

S2 Find deviation ; $d = X - \bar{X}$

S3 Square the deviation $(= d^2)$ & multiply by respective frequency i.e fd^2

S4 Total the product i.e Σfd^2. Then apply formula

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N}}$$

b) Assumed Mean Method:

S1 Assume any one of the item as average (A)

S2 Find out deviation from assumed mean i.e $X - A = d$

S3 Multiply these deviation by f i.e Σfd.

S4 Square the deviation d^2.

S5 Multiply the square deviation (d^2) by f & get Σfd^2.

S6 Apply formula

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

Eg: Calculate S.D from the following

Marks	No. of Students
10	8
20	12
30	20
40	10
50	7
60	3

Sol: a) Actual Mean Method:

X	f	fX	$d = X - \bar{X}$	d^2	fd^2
10	8	80	-20.8	432.64	3461.12
20	12	240	-10.8	116.64	1399.68
30	20	600	-0.8	0.64	12.80
40	10	400	9.2	84.64	846.40
50	7	350	19.2	368.64	2580.48
60	3	180	29.2	852.64	2557.92
$X=210$	$N=60$	$\Sigma fX = 1850$			$\Sigma fd^2 = 10,858.40$

$$\bar{X} = \frac{\Sigma fX}{N} = \frac{1850}{60} = 30.8$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N}}$$

$$= \sqrt{\frac{10858.40}{60}} = \underline{\underline{13.45}} \text{ Ans}$$

Page No.........
Date.........

b) Assumed Mean Method

Marks	f	d = X − 30	fd	fd²
10	8	−20	−160	3200
20	12	−10	−120	1200
30	20	0	0	0
40	10	10	100	1000
50	7	20	140	2800
60	3	30	90	2700
	N = 60		Σfd = 50	Σfd² = 10900

$$\sigma = \sqrt{\frac{fd^2}{N} - \left(\frac{fd}{N}\right)^2}$$

$$= \sqrt{\frac{10900}{60} - \left(\frac{50}{60}\right)^2}$$

$$= \sqrt{181.67 - 0.69} = \sqrt{180.98}$$

$$= 13.45 \quad \text{Ans.}$$

Coefficient of Variance:

Variance : Square of standard deviation

$$\text{Variance} = \sigma^2$$

$$\sigma = \sqrt{\text{Variance}}$$

$$\boxed{\text{Coefficient of Variance} = \frac{\sigma}{\bar{X}} \times 100}$$

Corelation :

The corelation refers to the relationship of two or more variables.

Type of Corelation:
→ Positive & Negative
→ Simple & Multiple
→ Partial & Total
→ Linear & Non-linear.

Methods for studying Correlation:

a) Graphic Method
→ Scatter Diagram
→ Simple Graph

b) Mathematical Method
→ Karl Pearson Coefficient of Corelation
→ Spearman's Rank Coefficient of Correlatio
→ Coefficient of Concurrent Deviation
→ Method of Least Squares.

Date:
P. No:

Coefficient of variation

Population — std deviation

$CV = \frac{\sigma}{\mu} \times 100\%$ — mean

Sample

$CV = \frac{S}{\bar{x}} \times 100\%$

Assumed mean
$dx = m - A$ ↳ middle value

$dx' = \frac{m-A}{c}$ → common term

~~Store wait time in minute~~

Step Deviation method

$$\bar{x}\,(\text{mean}) = \underset{\text{Assumed mean}}{A} + \frac{\Sigma f dx'}{N} \times c$$

Q. Calculate coefficient of variation,

marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of Students	2	4	5	9	10	5	15

marks	Number (f)	mid-value (m)	Deviation $dx = m-A$	$dx' = \frac{m-A}{c=10}$ (supposed)	fdx'	dx'^2	fdx'^2
0-10	2	5	-30	-3	-6	9	18
10-20	4	15	-20	-2	-8	4	16
20-30	5	25	-10	-1	-5	1	5
30-40	9	35 A Assumed	0	0	0	0	0
40-50	10	45	10	1	10	1	10
50-60	5	55	20	2	10	4	20
60-70	15	65	30	3	45	9	1
	N=50				$\Sigma fdx' = 46$		$\Sigma fdx'^2 = 204$

$\bar{x} = A + \frac{\Sigma fdx'}{N} \times c$

mean $(\bar{x}) = 35 + \frac{46}{50} \times 10$

$= \boxed{44.2}$

$\sigma\,(\text{Std Deviation}) = \sqrt{\frac{\Sigma fdx'^2}{N} - \left(\frac{\Sigma fdx'}{N}\right)^2} \times c$

$= \sqrt{\frac{204}{50} - \frac{46}{50}} \times 10 = \boxed{17}$

$CV = \frac{17}{44.2} \times 100 = \boxed{40.68}$

Date:
P. No:-

Correlation and Regression

* Correlation analysis deals with association b/w two or more variables
* The degree of relationship b/w the variables under consideration is measured through the correlation analysis.
* The measure of correlation called the "correlation coefficient "or" 'correlation index' summarizes in one figure the direction and degree of correlation.

X	y
10	15
12	20
15	22
18	25
20	37

x	y
20	40
30	30
40	22
60	15
80	10

Types of correlation

(direct correlation)
① Positive correlation
(If both the variables varies in same direction)
e.g. Height and weight
rainfall and yeild

$\boxed{x\uparrow\ y\uparrow}$ or $\boxed{x\downarrow\ y\downarrow}$

$r \rightarrow 0$ to 1

(inverse correlation)
② Negative correlation
(If both the variables varies in opposite direction)
e.g. price and Demand

$\boxed{x\uparrow\ y\downarrow}$ $\boxed{x\downarrow\ y\uparrow}$

$r \rightarrow -1$ to 0

Correlation coeff- (r)

$r \rightarrow -1$ to 1

$r = 1 \rightarrow$ perfect positive correlation
$r = -1 \rightarrow$ perfect negative correlation
$r = 0 \rightarrow$ no correlation

Methods of finding correlation coeff-

$-1 \quad\quad 0 \quad\quad +1$

very week correction
0·1, −0·1
0·2, −0·15

***
① Karl pearson's product moment method.
② spearman's Rank correlation coeff-
③ Scatter diagram method
④ coeff. of concurrent deviation

$$r = \frac{Cov(x, y)}{\sigma_x \, \sigma_y}$$

If direct value given we this

or $Cov(x, y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}$

$$\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$$\sigma_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}}$$

x	y
-	-
-	-
-	-
-	-

for this type of que mostly we use

no. of pair of observation $\rightarrow n = 4$

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\,\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

PTO

Q. Find coefficient of correlation of the following data

Date:
P. No:

Direct method

x	y	x^2	y^2	xy
9	15	81	225	135
8	16	64	256	128
7	14	49	196	98
6	13	36	169	78
5	11	25	121	55
4	12	16	144	48
3	10	9	100	30
2	8	4	64	16
1	9	1	81	9
45	108	$\Sigma x^2 = 285$	$\Sigma y^2 = 1356$	597

$$\Rightarrow r = \frac{n\Sigma xy - \Sigma x\Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{9(597) - (45)(108)}{\sqrt{9(285) - (45)^2}\sqrt{9(1356) - (108)^2}}$$

$$= \frac{5373 - 4860}{\sqrt{540}\sqrt{540}}$$

$$= \frac{513}{540} = 0.95 \text{ Ans}$$

nearest to 1
strong +ve correlation

when obser vary बड़ी हो large no. वाली

Shortcut method

when diff same न हो पाए

when diff

replaced

$x - A$ $x \rightarrow dx \rightarrow$ deviated x $\frac{x-A}{h_1}$

$x - B$ $y \rightarrow dy \rightarrow$ deviated y $\frac{y-B}{h_2}$

If diff b/w observation are same 10, 20, 30

$$r = \frac{n\Sigma dxdy - \Sigma dx\Sigma dy}{\sqrt{n\Sigma dx^2 - (\Sigma dx)^2}\sqrt{n\Sigma dy^2 - (\Sigma dy)^2}}$$

x	y	$dx = x - A$	$dy = x - B$	dx^2	dy^2	$dxdy$
9	15	4	4	16	16	16
8	16	3	5	9	25	15
7	14	2	3	4	9	6
6	13	1	2	1	4	2
A ← 5	11 → B	0	0	0	0	0
4	12	−1	1	1	1	−1
3	10	−2	−1	4	1	2
2	8	−3	−3	9	9	9
1	9	−4	−2	16	4	8
		0	9	60	69	57

$$r = \frac{9(57) - 0(9)}{\sqrt{9(60) - (0)^2}\sqrt{9(69) - (9)^2}}$$

$$r = \frac{513}{\sqrt{540}\sqrt{540}} = 0.95 \text{ Ans}$$

Q. Find coefficient of correlation b/w size group and defect in quality from the following data.

Date:
P. No:.

$\frac{150}{200} \times 100 = 75$

Size-Group	15-16	16-17	17-18	18-19	19-20	20-21
No. of items	200	270	340	360	400	300
No. of defective items	150	162	170	180	180	114
% defective items	75	60	50	50	45	38

x	y	x(middle)	$dx = x - 17.5$	$dy = x - 50$	dx^2	dy^2	$dxdy$
15-16	75	$\frac{15+16}{2} = 15.5$	-2	25	4	625	-50
16-17	60	16.5	-1	10	1	100	-10
17-18	50	17.5	0	0	0	0	0
18-19	50	18.5	1	0	1	0	0
19-20	45	19.5	2	-5	4	25	-10
20-21	38	20.5	3	-12	9	144	-36
			3	18	19	894	-106

$$r = \frac{6(-106) - (3)(18)}{\sqrt{6(19) - (3)^2}\sqrt{6(894) - (18)^2}} \Rightarrow -0.949$$

Properties of correlation coefficient

(1) The correlation coefficient always lies b/w -1 to $+1$

i.e. $-1 \leq r \leq +1$

(2) Correlation coefficient is symmetric about x and y

i.e. $r_{xy} = r_{yx}$

(3) Cor. coe. is Independent of origin and scale

i.e. $r_{xy} = r_{uv}$

(4) C.C. is the product of two regression coefficients

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

(5) If $r = 0$ It means variable are independent

Date:
P. No:

Rank correlation :-

Q. A biologist wish to see the relationship b/w height of tall trees and their diameter. The following table showing height and Diameter of different trees.
Calculate the rank correlation coefficient and interprit the result.

ranking

Dia (in)	Height (ft)	R(x)	R(y)	d = R(x) − R(y)	d^2
1024	261	10	8	2	4
950	321	9	10	−1	1
451	219	2	6	−4	16
505	281	3	9	−7	49
761	159	8	4	4	16
644	83	6	1	5	25
707	191	7	5	2	4
586	141	5	3	2	4
442	232	1	7	−6	36
546	108	4	2	2	4

$\Sigma d^2 = 159$

$$\text{ranking} = 1 - \left(\frac{6 \times \Sigma d^2}{n(n^2-1)}\right)$$

$$= 1 - \frac{6(159)}{10(10^2-1)} = 1 - \left(\frac{954}{990}\right) = 0.036$$

There is low relationship b/w dia and height of trees

Date:
P. No:

Difference b/w regression and correlation

(1) Regression is the dependent of one variable upon the other while correlation is the interdependency of both variables

(2) In correlation $r_{xy} = r_{yx}$ but in regression $b_{xy} \neq b_{yx}$

(3) Regression provides the idea of slope while correlation do not.

Regression Analysis.

* It is the technique for measuring or estimating the relationsh among variables
* Regression Analysis provides estimates of values of the dependent variable from the values of the independent variable
* The device used to accomplish this estimation procedure is the regression line.
* The regression line describes the average relationship existing b/w x and y.

(4) In regression it is possible to study cause and effect relationship while in correlation we can't say that one variable is the cause and other the effect

(5) Regression coefficient are independent of change of origin ($+,-$) but not of scale ($\times, \div$), while correlation coeff. is independent of change of scale and origine

(6) ~~Rep~~ Objective of reg. Analysis is to study the 'nature of relationship' b/w variables while correlation coefficient is a measure of 'degree of covariability' b/w X and Y

Similarity → Corr. coeff (r) and reg. coeff (b_{xy} and b_{yx}) always have same sign

22

Regression equations

Regression lines

Regression line of y on x

① $y-\bar{y} = b_{yx}(x-\bar{x})$

reg. coeff of y on x

$$b_{yx} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

both lines intersect at point $(\bar{x}, \bar{y})$

② x → independent variable
y → dependent variable

③ used to calculate y for given x

④ slope of line = b_{yx}

coeff of x → b_{yx}

Regression line of x on y

① $x-\bar{x} = b_{xy}(y-\bar{y})$

reg. coeff of x on y

$$b_{xy} = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma y^2 - (\Sigma y)^2}$$

② x → dependent variable
y → independent variable

③ used to calculate x for given y.

④ slope of line = b_{xy}

Q. find ~~the~~ ~~a~~ both regression lines (x on y and y on x) for the following data.

x	y	x^2	y^2	xy
6	9	36	81	54
2	11	4	121	22
10	5	100	25	50
4	8	16	64	32
8	7	64	49	56
30	40	220	340	214

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

y on x → $y-\bar{y} = b_{yx}(x-\bar{x})$

$$b_{yx} = \frac{n\Sigma xy - \Sigma x\Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$b_{yx} = \frac{5(214)-(30)(40)}{5(220)-(30)^2}$$

$$= \frac{1070-1200}{1100-900}$$

$$= -0.65$$

$$b_{xy} = \frac{n\Sigma xy - \Sigma x\Sigma y}{n\Sigma y^2 - (\Sigma y)^2}$$

$$b_{xy} = \frac{5(214)-(30)(40)}{5(340)-(40)^2}$$

$$= \frac{1070-1200}{1700-1600}$$

$$= -1.3$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{6+2+10+4+8}{5} = 6$$

$$\bar{y} = 8$$

Date:
P. No:

x on y

$x-6 = -(1.3)(y-8)$

$x-6 = -1.3y + 1.3\times 8$

$x = -1.3y + 10.4 + 6$

$\boxed{x = -1.3y + 16.4}$

y on x

$y-8 = -0.65(x-6)$

$y = -0.65x + 3.9 + 8$

$\boxed{y = -0.65x + 11.9}$

Q. From the following data
(a) Obtain two regression lines (x on y and y on x) b_{xy}, b_{yx}
(b) Calculate correlation coeff. $r = \sqrt{b_{xy} \times b_{yx}}$
(c) Estimate the value of y for $x = 0.62$

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

y on x $\rightarrow (y-\bar{y}) = b_{yx}(x-\bar{x})$

x	y	xy	x^2	y^2
1	9	9	1	81
2	8	16	4	64
3	10	30	9	100
4	12	48	16	144
5	11	55	25	121
6	13	78	36	169
7	14	98	49	196
8	16	128	64	256
9	15	135	81	225
45	108	597	285	1356

$\bar{x} = \dfrac{1+2+3+4+5+6+7+8+9}{9}$

$\boxed{\bar{x} = 5}$

$\bar{y} = \dfrac{108}{9}$ 12

$\boxed{\bar{y} = 12}$

$b_{xy} = \dfrac{(9)(597)-(45)(108)}{(9)(1356)-(108)^2} = \dfrac{513}{540}$

$= 0.95$

$b_{yx} = \dfrac{(9)(597)-(45)(108)}{(9)(285)-(45)^2}$ ~~bxy~~

$= \dfrac{513}{540} = 0.95$

PTO

Date:
P. No:

<u>x on y</u>

$$x - \bar{x} = b_{xy}(y - \bar{y})$$
$$x - 5 = 0.95(y - 12)$$
$$\boxed{x = 0.95y - 6.4}$$

y on x

$$y - \bar{y} = b_{yx}(x - \bar{x})$$
$$y - 12 = 0.95(x - 5)$$
$$\boxed{y = 0.95x + 7.25}$$

<u>ATQ</u>

value of y for $x = 0.62$

$$y = 0.95 \times 0.62 + 7.25$$
$$\boxed{y = 7.839}$$ A

correlation coefficient

$$r = \sqrt{b_{xy}\, b_{yx}}$$
$$= \sqrt{0.95 \times 0.95}$$
$$\boxed{r = 0.95}$$

median for continous series

$$m = L_1 + \frac{\frac{N}{2} - c.f}{f} \times i$$

e.g.

marks	f	c.f
10-25	6	6
25-40	20	26 cf
40-55	44 f	70
55-70	26	96
70-85	3	99
85-100	1	100

median $= \frac{N}{2} = \frac{100}{2} = 50 \rightarrow$ lies 40-55

$$m = L_1 + \frac{\frac{N}{2} - cf}{f} \times i$$

$$= 40 + \left(\frac{50 - 26}{44}\right) \times 15 \Rightarrow 48.18$$ A

median of Discrete series:-

$S_1 \rightarrow$ arrange the data in ascending order

$S_2 \rightarrow$ find cummulative freq.

$S_3 \rightarrow$ Apply formula

Odd $\rightarrow$ median = size of $\left(\frac{N+1}{2}\right)^{th}$ item

even $\rightarrow$ " = size of $\left(\frac{N}{2}\right)^{th}$ item

e.g. locate median

size of shoes	freq.	cf
5	10	10
5.5	16	26
6	28	54
6.5	15	69
7	30	99
7.5	40	139
8	34	173

$$m = \frac{173+1}{2} \rightarrow \frac{174}{2} = 87^{th} \rightarrow 7$$ A

Some theory in laptop

Date:
P. No:

Unit-3
Testing of Hypothesis

Procedure of testing of Hypothesis :-
Steps :- (1) Set up a Null Hypothesis
(2) Set up a suitable level of significance
(3) Set up a suitable test of Hypothesis
(4) Calculation for Test statistics
(5) Compare calculated value of Test Statistics with Table value.
(6) Take suitable Decisions

Basics of Hypothesis :-
content
(1) Meaning of Hypothesis
(2) Concept of population parameters
(3) Concept of sample statistics
(4) Concept of Null Hypothesis
(5) Concept of Alternative Hypothesis
(6) How to set Hypothesis.

There are two types of Hypothesis :- (Theory in laptop)

(1) Null Hypothesis (2) Alternative Hypothesis

~~(1) Null Hypothesis~~ :-

e.g. If we required to test the statement that the mean height of population is 65". Then the two Hypothesis will be formed as:-

Null Hypothesis H_0 $\Rightarrow \mu = 65$

Alternative Hypothesis $H_1 \Rightarrow \mu \neq 65$ (or) $\mu > 65$ (or) $\mu < 65$

$p \to$ probability

	Accept Null Hypothesis Ho	Reject Null Hypothesis Ho
Ho is True	Correct Decision $p = 1-\alpha$	Type 1 Error $p = \alpha$
Ho is false	Type 2 Error $p = \beta$	Correction Decision $p = 1-\beta$

Example of level of significance :-

Q. The mean height of a random sample of 100 students is 64" and std. Deviation is 3". Test the statement that the mean height of population is 67" at 5% level of significance

$\Rightarrow$ 5% (or) 0.05 is the level of significance (probability of occuring Type 1 Errors)

level of confidence
$= 1 - 0.05 = 95\%$

Steps in Hypothesis Testing :-

- State the Null hypothesis (Ho) and Alternate Hypothesis (Ha)
- Choose the level of significance (α)
- Determine the rejection region
- Calculate Test statistics and compare it with critical value.
- State the conclusion (Reject the Null Hypothesis / Fail to Reject Null Hypothesis).

← Digital E-learning youtube

Case-1 Hypothesis Testing for single population mean.

Date:
P. No: 27

e.g.① Vendor claims that average weight of box is 1.84 kg. Customer randomly choose 64 parts and find sample weight as 1.88 kg. Suppose population std. deviation is 0.3 kg. Use $\alpha = 0.05$ and test for hypothesis that true mean is of shipment is 1.84 kg.

⇒ Null Hypothesis (H_0) : $\mu = 1.84$

Alternate Hypothesis (H_a) : $\mu \neq 1.84$

Prob. of Type 1 error $\alpha = 0.05$, $Z_{\alpha/2} = Z_{0.025}$

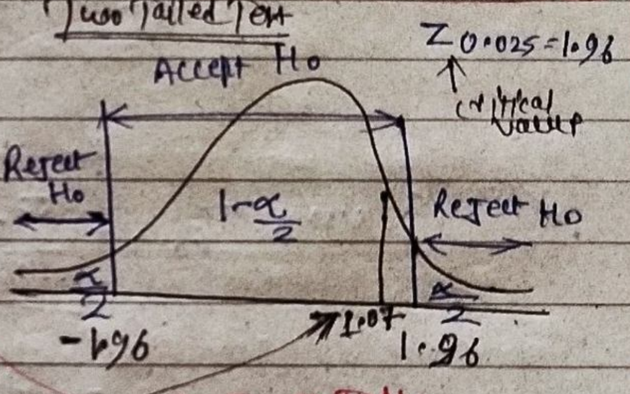
$\frac{\alpha}{2} = 0.025$

$1 - 0.025 = 0.975$

$Z_{0.025} = 1.96$ ← critical value

$$Z_{calculated} = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$$

$\bar{X}$ → sample mean, μ → population mean, $\sigma/\sqrt{N}$ → std. error, σ → popu. std deviation, N → sample size



$$= \frac{1.88 - 1.84}{0.3/\sqrt{64}}$$

$= 1.07$ → falls b/w $\boxed{-1.96 \text{ to } 1.96}$

→ In Z Table

	0.06
1.9	0.9750

adding → 1.96

Decision → Fails to reject Null Hypothesis (H_0)

or

Accept (H_0)

~~case-2~~ m/c produces parts with mean length of 4.125 mm.

eg. (2) (20) customer randomly choose 20 parts and find ~~sample length~~ as 4.120 mm. Suppose sample std. deviation is 0.008 mm. use $\alpha = 0.10$, and determine whether the mean length has decreased?

$\Rightarrow$

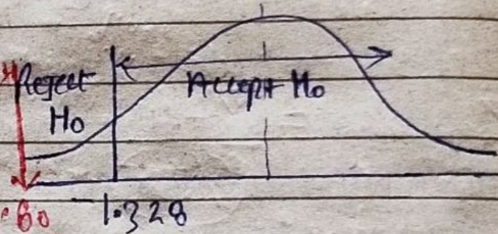
$$H_0 : \mu = 4.125 \qquad H_a : \mu < 4.125 \qquad \alpha = 0.10$$

Hyp Test for single populatn mean

$$T_{(critical)} = T_{(\alpha, N-1)} \quad (N = 20)$$

$$= T_{(0.1, 19)} \rightarrow \text{looking at T-Table}$$

$$T_{(critical)} = 1.328$$



Now $T_{(cal)} = \dfrac{\bar{x} - \mu}{S/\sqrt{N}}$

Sample std. devi.

$$= \frac{4.120 - 4.125}{0.008/\sqrt{20}} = -2.80$$

Decision $\Rightarrow$ Reject the Null Hypothesis.

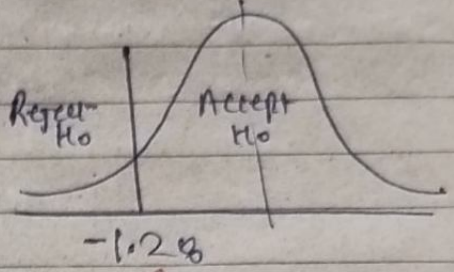
case-3 Hypothesis Testing for single population proportion

Billing statements in particular hotel has some error. There was error of 15%. out of 1000 billing statements which were randomly selected contains 102 errors. use $\alpha = 0.10$, and determine wheter the proportion of billing statement is less than 15%. ?

Date:
P. No:

→ $H_0: p = 0.15 \quad H_a: p < 0.15 \quad \alpha = 0.10$

$$\hat{p} = \frac{x}{n} = \frac{102}{1000} = 0.102$$



$Z(\alpha)$; $Z_{(0.1)} = ?$

$1 - 0.1 = 0.9$

→ looking at Z table

$Z_{(0.1)} = 1.28$

$$Z_{cal} \cong \frac{\hat{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

$$Z_{(cal)} \cong \frac{0.102 - 0.15}{\sqrt{\frac{0.15(1-0.15)}{1000}}} \cong -4.25$$

Decision = Reject the null hypothesis.

Case 4 Hypothesis Testing for population variance

m/c is running with variance of 6.25 for some critical dimension. After improving the process, sample of 13 items are drawn randomly. Now the variance is 6.82 use $\alpha = 0.05$, and determine whether the new process variance has increased ?

PTO

Date:
P. No:

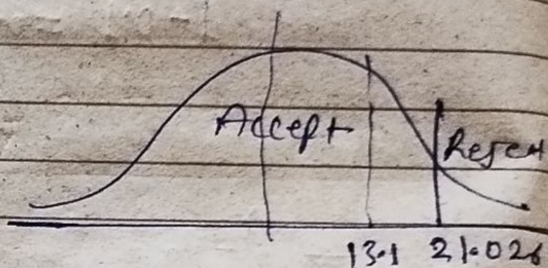
Ans $H_0 : \sigma^2 = 6.25$ $H_a : \sigma^2 > 6.25$ $\alpha = 0.05$

χ^2 (critical) $= \chi^2_{(\alpha, N-1)}$ $\chi^2_{(0.05, 12)}$

check in chi-square table

$$\chi^2_{(0.05, 12)} = 21.026$$

$$\chi^2_{(cal)} = \frac{(n-1)s^2}{\sigma_0^2}$$



$$\chi^2_{(cal)} = \frac{(13-1)\, 6.82}{6.25}$$

$$\chi^2_{(cal)} = 13.1$$

Decision ⇒ fails to reject the Null Hypothesis.

→ Simple Random Sampling with Replacement

Sampling Distribution and standard error :-

K. Sample

Population size = N

sample size = n

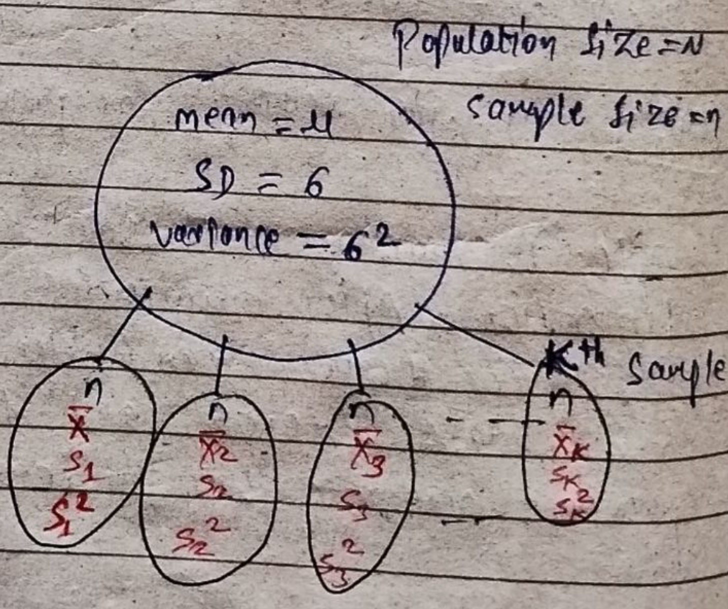
SRSWR $K = N^n$

SRSWOR $K = {}^N C_n$

eg. $N = 4$ (population size)

$n = 2$ (sample size)

Sample



$K = 4^2 = 16$

$K = {}^4 C_2 = \frac{4!}{2!\,2!} = 6$

Sampling Distribution of mean

mean of Sampling Dist. of mean: $\boxed{\mu_{\bar{x}} = \mu}$ → population mean

Date:
P. No: (31)

$\bar{X}_1$	$P(\bar{X}_1)$
$\bar{X}_2$	$P(\bar{X}_2)$
$\bar{X}_3$	$P(\bar{X}_3)$
$\vdots$	$\vdots$
$\bar{X}_K$	$P(\bar{X}_K)$

→ Std. deviation of Sampling distribution is called Std. error

S.D. of Sampling Dist. of mean (Std. error of mean) $\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$

$\sqrt{n}$ → Sample size; σ → Population S.D.

Statistic	Std. error
(1) Sample mean, $\bar{x}$	$\sigma/\sqrt{n}$
(2) Sample SD, s	$\sqrt{\sigma^2/2n}$
(3) Sample variance, s^2	$\sigma^2\sqrt{2/n}$
(4) Sample proportion, p	$\sqrt{PQ/n}$
(5) Sample correlation, r	$(1-\rho^2)/\sqrt{n}$
(6) Diff. of two sample mean $(\bar{x}_1 - \bar{x}_2)$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Numericals

A population consists of the four members 3, 7, 11, 15

case 1):- Consider all possible sample size $\boxed{\text{two}}$ which can be drawn with replacement from population, find the population mean (μ), population Std. deviation (σ), the mean of the sampling distribution of mean $(\mu_{\bar{x}})$ and Std. deviation of sampling distribution of mean $(\sigma_{\bar{x}})$

PTO

Q. 3 7 Population
11 15

N = 4

32

Date:
P. No:

Case I

SRSWR ← given

$K = N^n = 4^2 = 16$

S.No.	Sample value	Total	Sample mean
1	3, 7	10	10/2 = 5
2	3, 11	14	7
3	3, 15	18	9
4	3, 3	6	3
5	7, 7	14	7
6	7, 3	10	5
7	7, 11	18	9
8	7, 15	22	11
9	11, 11	22	11
10	11, 3	14	7
11	11, 7	18	9
12	11, 15	26	13
13	15, 15	30	15
14	15, 3	18	9
15	15, 7	22	11
16	15, 11	26	13

Sampling Dist of mean with Replacement

Sample mean, $\bar{x}$	3	5	7	9	11	13	15	Total
probability	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	1

Population

X	$X-\mu$ $X-9$	$(X-\mu)^2$
3	-6	36
7	-2	4
11	2	4
15	6	36
$\Sigma X = 36$		$\Sigma(X-\mu)^2 = 80$

Population mean, $\mu = \frac{36}{4} = 9$

Variance, $\sigma^2 = \frac{\Sigma(X-\mu)^2}{N} = \frac{80}{4} = 20$

S.D, $\sigma = \sqrt{20}$

↑ Population S.D.

Date:
P. No:

mean of Sampling Dist of mean.

using mathematical expected value method because $\bar{x}$ is random variable

$$E(\bar{x}) = 3\times\frac{1}{16} + 5\times\frac{2}{16} + 7\times\frac{3}{16} + 9\times\frac{4}{16} + 11\times\frac{3}{16} + 13\times\frac{2}{16} + 15\times\frac{1}{16}$$

$$= \frac{3}{16} + \frac{10}{16} + \frac{21}{16} + \frac{36}{16} + \frac{33}{16} + \frac{26}{16} + \frac{15}{16}$$

mean $E(\bar{x}) = \frac{144}{16} = 9$

$$\boxed{\mu_{\bar{x}} = 9 = \mu}$$

$$Var(\bar{x}) = E(\bar{x})^2 - [E(\bar{x})]^2$$

Now

$$E(\bar{x})^2 = 3^2\times\frac{1}{16} + 5^2\times\frac{2}{16} + 7^2\times\frac{3}{16} + 9^2\times\frac{4}{16} + 11^2\times\frac{3}{16} + 13^2\times\frac{2}{16} + 15^2\times\frac{1}{16}$$

$$= 9\times\frac{1}{16} + \frac{50}{16} + \frac{147}{16} + \frac{324}{16} + \frac{363}{16} + \frac{338}{16} + \frac{225}{16}$$

$E(\bar{x})^2 = 91$

$$Var(\bar{x}) = E(\bar{x})^2 - [E(\bar{x})]^2$$

$Var(\bar{x}) = 91 - 9^2 = 91 - 81$

$Var(\bar{x}) = 10$

$SD(\bar{x}) = \text{Std. Error}(\bar{x}) = \sqrt{10}$

Std error of sampling Dist of mean

(3)

Date:
P. No:

Std. error of Sampling Dist. of mean

$$S.E(\bar{x}),\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{20}}{\sqrt{2}} = \sqrt{\frac{20}{2}} = \sqrt{10} = SD(\bar{x})$$

So, Here

$$S.E(\bar{X}) = \sqrt{10} = SD(\bar{X})$$

Case-2

SRS w/o Replacement

3 7 → popⁿ
11 15

N = 4

$$K = {}^{N}C_n = {}^{4}C_2 = \frac{4\times3\times2\times1}{2!(4-2)!} = \frac{4\times3\times2\times1}{2\times1\times2\times1} = 6$$

S.No.	Sample values	Total	Sample mean
1	3, 7	10	10/2 = 5
2	3, 11	14	7
3	3, 15	18	9
4	7, 11	18	9
5	7, 15	22	11
6	11, 15	26	13

Sampling Distribution of mean w/o replacement.

Sample mean, $\bar{x}$	5	7	9	11	13	Total
probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

$E(\bar{X}) = 5 \times \frac{1}{6} + 7 \times \frac{1}{6} + 9 \times \frac{2}{6} + 11 \times \frac{1}{6} + 13 \times \frac{1}{6}$

$= \frac{5}{6} + \frac{7}{6} + \frac{18}{6} + \frac{11}{6} + \frac{13}{6} = \frac{54}{6}$

$\boxed{E(\bar{X}) = 9}$

$\downarrow$

$\boxed{\mu_{\bar{x}} = \mu = 9}$

Now,

$E(\bar{X})^2 = 5^2 \times \frac{1}{6} + 7^2 \times \frac{1}{6} + 9^2 \times \frac{2}{6} + 11^2 \times \frac{1}{6} + 13^2 \times \frac{1}{6}$

$= \frac{25}{6} + \frac{49}{6} + \frac{81 \times 2}{6} + \frac{121}{6} + \frac{169}{6}$

$E(X^2) = \frac{526}{6} = \frac{263}{3}$

Now

$var(\bar{X}) = E(\bar{X})^2 - [E(\bar{X})]^2$

$= \frac{263}{3} - 81$

$var(\bar{X}) = \frac{263 - 243}{3} = \frac{20}{3}$

$\boxed{S.D.(\bar{X}) = \sqrt{\frac{20}{3}}}$ when using SRS w/o Replace

If $\frac{n}{N} > 0.05$, then we use finite correction factor $\sqrt{\frac{N-n}{N-1}}$

$\frac{2}{4} = 0.5 > 0.05$

So, $S.E(\bar{X}), \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$

PTO

(36)

central limit theorem

Date:
P. No:

$$\sigma_{\bar{x}} = \frac{\sqrt{20}}{\cancel{\sqrt{2}}} \times \sqrt{\frac{4-2}{4-1}}$$

$$\boxed{\sigma_{\bar{x}} = \sqrt{\frac{20}{3}}}$$

The Statistics Teacher youtube channel

Estimation

Suppose a researcher interested in obtaining an estimate of the average level of some enzyme in a certain human population, takes a sample of 10 individuals, determine the level of enzyme in each and compute a sample mean of $\bar{x} = 22$. Suppose further it is known that the variable of interest is approximately normally distributed with a variance of 45. we wish to estimate m.

$\Rightarrow$

$n = 10 \rightarrow$ Central Limit theorem (CLT)

$\bar{x} = 22$

$\sigma^2 = 45$

$$\bar{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$22 \pm 1.96 \times \frac{\sqrt{45}}{\sqrt{10}} = 22 \pm 4.158$$

$$= (22 - 4.158,\ 22 + 4.158)$$

$$m = (17.842,\ 26.158)$$

95% CI

Estimating means and proportions

e.g. $\rightarrow$ A homeowner randomly samples 64 homes similar to her own and finds that the average S.P. is Rs 250,000 with a std. of Rs 15,000. Estimate

The avg sp for all similar homes in the city.

$\Rightarrow$

$\bar{x}$ = Rs 250,000

σ = Rs 15,000

$n = 64 > 30 \rightarrow$ CLT

$\bar{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$

Rs $250{,}000 \pm 1.96 \times \frac{15000}{\sqrt{64}}$

Rs $250{,}000 \pm 3675$ Rs.

μ = (Rs 246,325, Rs. 253,675) 95% CI $\rightarrow$ confidence interval.

Student's t-test :- Puspendra classes YT

Application and formula

① To test the significance of the mean of a random sample

$t = \frac{(\bar{x} - \mu) \times \sqrt{n}}{s}$, $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{(n-1)}}$, $\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{n}}$ $\rightarrow$ for populatn

$(n-1) \rightarrow \nu$ $\uparrow$ DOF $\rightarrow$ Degree of freedom

$\bar{x}$ = mean of the sample

μ = —— " —— populatn

n = sample size

s = std deviation of the sample

Confidence interval estimate (for α level of significance)

One tailed test $\bar{x} \pm t_{\alpha} \frac{s}{\sqrt{n}}$ $\alpha = 0.05$

Two tailed test $\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$ $\alpha = \frac{0.05}{2} = 0.025$

3.8

Date:
P. No:

(2) To test the difference b/w means of the two samples (Independent samples)

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{S} \times \sqrt{\frac{n_1 \times n_2}{n_1 + n_2}}, \quad S = \sqrt{\frac{\Sigma(x_1 - \bar{x}_1)^2 + \Sigma(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2 \rightarrow \nu}}$$

$\bar{x}_1$ = mean of the sample (1), $\bar{x}_2$ = mean of sample (2)
n_1 = sample size of sample (1), n_2 = sample size of sample (2)
S = std deviation (combined)

(3) To test the difference b/w means of two samples (Dependent samples or matched pair observations).

$$t = \frac{\bar{d}\sqrt{n}}{S}, \quad S = \sqrt{\frac{\Sigma(d - \bar{d})^2}{n-1 \rightarrow \nu}}$$

$\bar{d}$ = mean of the differences
S = std. deviation of the differences
n = size of the sample

(4) Testing the significance of an observed correlation coefficient.

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

r = correlation coefficient
n = size of the sample

see in Laptop folder

39

Date:
P. No:

Q. 1) The manufacturer of a certain make of LED Bulb claims that his bulbs have a mean life of 20 months. A random sample of 7 such Bulbs gave the following values.

Life of bulbs in months : 19, 21, 25, 16, 17, 14, 21.

Can you regard the producer's claim to be valid at 1% level of significance?

Testing the significance of the mean (for Random sample)

Soln :-

Given data,

Population mean $\mu = 20$ months

Life of Bulbs in months = 19, 21, 25, 16, 17, 14, 21.

$n = 7$

Level of significance = 1%

$H_0 : \mu = \bar{x}$

$H_a : \mu \neq \bar{x}$

$$t = \frac{|\bar{x} - \mu|}{S} \times \sqrt{n} \; ; \qquad S = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n - 1}}$$

PTO

(40)

Date:
P. No:

Calculation of $\bar{x}$ and s

x	$x-19$ $(x-\bar{x})$	$(x-\bar{x})^2$
19	0	0
21	+2	4
25	+6	36
16	-3	9
17	-2	4
14	-5	25
21	+2	4
$\Sigma x = 133$		$\Sigma(x-\bar{x}) = 82$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{133}{7} = 19$$

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

$$s = \sqrt{\frac{82}{6}} = 3.7$$

Now

$$t = \frac{|\bar{x} - \mu|}{s} \times \sqrt{n}$$

$$t = \frac{|19-20|}{3.7} \times \sqrt{7} = \underline{0.716} \text{ (calculated } t \text{ value)}$$

(DOF) $v = n-1$

$v = \cancel{7}-1 = \boxed{6}$ and ATQ 1% LOsignificance = 0.01

after checking t table

	Two tailed test 0.01
6	3.707

So, $t_{0.01} = \underline{3.707}$

So,

Ho is passed and accepted. There is no difference b/w the sample mean and population mean life of bulb. claim of the producer is correct.

Q.2 A random sample of size 15 has 50 as mean, the sum of the squares of the deviation taken from mean is 130. Can this sample be regarded as taken from the population having 53 as mean? obtain 95% and 99% confidence limits of the mean for the population.

$\bar{x} = 50$, $\mu = 53$
$n = 15$, $\Sigma(x-\bar{x})^2 = 130$

$H_0 : \mu = \bar{x}$
$H_a : \mu \neq \bar{x}$

$t = \frac{|\bar{x}-\mu|}{s} \times \sqrt{n}$, $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$

$S = \sqrt{\frac{130}{15-1}} = \sqrt{\frac{130}{14}} = 3.05$

$t = \frac{|50-53|}{3.05} \times \sqrt{15} = \frac{3 \times 3.87}{3.05}$

$t = \frac{11.61}{3.05} = 3.81$ (calculated)

$\nu = n-1 = 15-1 = 14$

$t_{0.05} = 2.145$

$t_{0.01} = 2.977$

42

Date:
P. No:

Sample mean for 95% confidence limit.

$$\bar{x} \pm \frac{s}{\sqrt{n}} \times t_{0.05}$$

$$= 50 \pm \frac{3.05}{\sqrt{15}} \times 2.145 = 50 \pm \frac{3.05}{3.87} \times 2.145$$

$$= 50 \pm \frac{6.54}{3.87} = 50 \pm 1.69$$

limit = 48.31 to 51.69

Sample mean for 99% confidence limit

$$\bar{x} \pm \frac{s}{\sqrt{n}} \times t_{0.01}$$

$$= 50 \pm \frac{3.05}{\sqrt{15}} \times 2.977$$

$$= 50 \pm \frac{3.05}{3.87} \times 2.977$$

$$= 50 \pm \frac{9.08}{3.37} = 50 \pm 2.35$$

limit = 47.65 to 52.35.

Testing the difference b/w means of the two independent samples.

Date:
P. No: 43

Q.1) Two type of drugs were used on 6 and 5 patients for reducing their weight. Drug A was imported and Drug B was indegenous. The increase in the weight (in kg) after using the drugs for 90 days was given below:

Drug A: 8, 10, 12, 9, 14, 13
Drug B: 7, 9, 14, 12, 8

is there a significant difference in the efficacy of the drug?

⇒ H_0: Drug A = Drug B
H_a: Drug A ≠ Drug B

Given Data

$n_1 = 6$, $n_2 = 5$

Calculation of $\bar{x}_1$, $\bar{x}_2$ and s

x_1	$(x_1 - \bar{x}_1)$ $x_1 - 11$	$(x_1 - \bar{x}_1)^2$	x_2	$(x_2 - \bar{x}_2)$	$(x_2 - \bar{x}_2)^2$
8	-3	9	7	-3	9
10	-1	1	9	-1	1
12	+1	1	14	+4	16
09	-2	4	12	+2	4
14	+3	9	8	-2	4
13	+2	4			
66		$\Sigma(x_1-\bar{x}_1)^2 = 28$	50		$\Sigma(x_2-\bar{x}_2)^2 = 35$

$\bar{x}_1 = \frac{\Sigma x_1}{n} = \frac{66}{6} = 11$ $\quad x_2 = \frac{\Sigma x_2}{n} = \frac{50}{5} = 10$

PTO

(44)

Date:
P. No:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S} \times \sqrt{\frac{n_1 \times n_2}{n_1 + n_2}} \quad , \quad S = \sqrt{\frac{\Sigma(x_1-\bar{x}_1)^2 + \Sigma(x_2-\bar{x}_2)^2}{(n_1+n_2-2)}}$$

$\leftarrow \nu$ DoF

$$= \frac{|\cancel{11}-10|}{2.62} \times \sqrt{\frac{6\times5}{6+5}} \quad \xleftarrow{\text{Put}} \quad S = \sqrt{\frac{28+34}{6+5-2}} = \sqrt{\frac{66}{9}} = \sqrt{6.88} = 2.62$$

$$= \frac{1}{2.62} \times \sqrt{\frac{30}{11}}$$

$$t = \frac{1}{2.62} \times \sqrt{2.72} = \frac{1}{2.62} \times 1.65 = \frac{1.65}{2.62} = 0.63 \text{ (Calculated)}$$

$$\nu = n_1 + n_2 - 2 = 6+5-2 = 9$$

$\boxed{\nu = 9}$, $t_{0.05}$ $\Rightarrow$ after checking in t table $= 2.262$

$$\boxed{t_{0.05} = 2.262}$$

So, Here calculated $t = 0.630$
and $t_{0.05} = 2.262$

So, H_0 is passed and accepted. So there is no significance difference b/w the efficacy of Drug A and Drug B

Q. Two laboratories A and B carry out independent estimates of sugar content in a chocolate made by a firm named X. A sample is taken from each batch and divided into two parts, and sent to the two laboratories. The sugar content obtained by the laboratories are regarded below (in mg/g)

Date:
P. No:

Lab A: 8, 9, 9, 6, 4, 6

Lab B: 7, 8, 6, 4, 5, 6

is there a significant diff. b/w the mean sugar content obtained by two laboratories A and B

⇒ H_0: Lab A = Lab A

H_a: Lab A ≠ Lab B

$n_1 = 6$, $n_2 = 6$

x_1	$(x_1-\bar{x}_1)$	$(x_1-\bar{x}_1)^2$	x_2	$(x_2-\bar{x}_2)$	$(x_2-\bar{x}_2)^2$
8	1	1	7	1	1
9	2	4	8	2	4
9	2	4	6	0	0
6	-1	1	4	-2	4
4	-3	9	5	-1	1
6	-1	1	6	0	0
42		20	36		10

$\bar{x}_1 = \frac{42}{6} = 7$ $\bar{x}_2 = \frac{36}{6} = 6$

$S = \sqrt{\frac{20+10}{6+6-2}} = \sqrt{\frac{30}{10}} = 1.73$

$t = \frac{|\bar{x}_1 - \bar{x}_2|}{S} \times \sqrt{\frac{n_1 \times n_2}{n_1+n_2}} \Rightarrow \frac{7-6}{1.73} \times \sqrt{\frac{6\times 6}{12}} = \frac{1}{1.73} \times 1.73$

$t = 1$ (calculated)

$v = n_1+n_2-2 = 6+6-2 = 10$

$t_{0.05} = 2.228$

H_0 is correct and accepted so, there is no significant diff. b/w the sugar content obtained by Lab A and LAB B.

See Basic formula and condition and various application in laptop

Date:
P. No:

chi-square test for goodness of fit

Q. In an antimalarial campaign in India, Quinine was administered to 500 persons out of a total population of 2000. The no. of fever cases is shown below:

Treatment	Fever	No fever	total
Quinine	20	480	500
No Quinine	100	1400	1500
Total	120	1880	2000

Discuss the usefulness of quinine in checking malaria

Soln

	Trtment	Fever	Expect ① value	No fever	expected value ②	Total
①	Quinine	20	30	480	470	500
②	No Quinine	100	90	1400	1410	1500
	Total	120		1880		2000

$$x^2 = \Sigma \frac{(O-E)^2}{E}, \quad E = \frac{RT \times CT}{N}, \quad N = 2000$$

O → observed frequency; E → expected frequency

RT → Row total for Row containing cell; CT → Coloum total for coloumn containing cell; N → Total no. of observations

$$E_{11} = \frac{RT \times CT}{N} = \frac{500 \times 120}{2000} = 30$$

$$E_{21} = \frac{RT \times CT}{N} = \frac{1500 \times 120}{2000} = 90$$

$$E_{21} = \frac{RT \times CT}{N} = \frac{500 \times 1880}{2000} = 470$$

$$E_{22} = \frac{RT \times CT}{N} = \frac{1500 \times 1880}{2000} = 1410$$

Date:
P. No: (47)

calculation of x^2

O	E	(O-E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
20	30	-10	100	3.33
100	90	+10	100	1.11
480	470	+10	100	0.21
1400	1410	-10	100	0.07
			$\Sigma \frac{(O-E)^2}{E} = 4.72$	

$$x^2 = \Sigma \frac{(O-E)^2}{E} = 4.72$$

no. of colouny   no. of rows

$$v(DOF) = (c-1)(r-1)$$
$$= (2-1)(2-1) = 1 \times 1 = 1$$

$$x^2_{0.05} = 3.84$$

Ho is failed and Rejected Hence Quinine is usefull in checking the malaria.

F-test

Introduction, Formula, Properties of F-distribution
See Laptop folder

Q. Two random samples were drawn from two normal populatn and their values are

A : 16, 17, 25, 26, 32, 34, 38, 40, 42

B : 14, 16, 24, 28, 32, 35, 37, 42, 43, 45, 47

Test whether the two populations have the same variance at 5% level of significance

(48)

Date:
P. No:

★★★

Formulas to carry out the test of significance we have to calculate the ratio of. F

$$F = \frac{\sigma_1^2}{\sigma_2^2} \quad OR \quad F = \frac{S_1^2}{S_2^2}$$

$$\sigma_1^2 > \sigma_2^2 \qquad S_1^2 > S_2^2$$

Variance is square of std. deviation

$$\sigma^2 = \frac{\Sigma(x-\mu)^2}{n} \quad , \quad S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

- In simplified way F can be written as

$$F = \frac{\text{Larger estimate of variance}}{\text{Smaller estimate of variance}}$$

DOF (ν)

ν_1 (numerator) = $n_1 - 1$ (Larger variance)
ν_2 (denominator) = $n_2 - 1$ (smaller variance)

Soln

A :- 16, 17, 25, 26, 32, 34, 38, 40, 42,
$n = 9$

B : 14, 16, 24, 28, 32, 35, 37, 42, 43, 45, 47.
$n = 11$

Let's take the Hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$
$H_a : \sigma_1^2 > \sigma_2^2$

PTO

Date:
P. No:

A	$(x-\bar{x}_A)$ $x-30$	$(x-\bar{x}_A)^2$	B	$(x-\bar{x}_B)$ $x-33$	$(x-\bar{x}_B)^2$
16	-14	196	14	-19	361
17	-13	169	16	-17	289
25	-5	25	24	-9	81
26	-4	16	28	-5	25
32	+2	04	32	-1	01
34	+4	16	35	+2	04
38	+8	64	37	+4	16
40	+10	100	42	+9	81
42	+12	144	43	+10	100
			45	+12	144
			47	+14	196
270	$\Sigma(x-\bar{x}_A)^2 = 734$		363	$\Sigma(x-\bar{x}_B)^2 = 1298$	

$$\bar{x}_A = \frac{270}{9} = 30 \quad , \quad \bar{x}_B = \frac{363}{11} = 33$$

$$F = \frac{S_1^2}{S_2^2} \, , \quad S_1^2 > S_2^2 \quad . \quad S^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

$$S_2^2 = \frac{734}{9-1} = 91.75$$

$$S_1^2 = \frac{1298}{11-1} = 129.8$$

$$F = \frac{129.8}{91.75} = 1.4147 \text{ (calculated F value)}$$

$$\upsilon_1 = 11-1 = 10 \, .$$

$$\upsilon_2 = 9-1 = 8$$

$$F_{0.05} = 3.35 \text{ (Tabulated value)}$$

Calculated F value < Tab. F val. So, Null Hypothesis is accepted, Hense, we can say that two populations have same variance

Date:
P. No:

Analysis of variance (ANOVA)

Introduction, Assumptions, (see laptop)

One way ANOVA

Q. To assess the significance of possible variation in performance in a certain test b/w the convent schools of a city, a common test was given to a number of students taken at random from the fifth class of the 3 schools concerned. The result given below:

A	B	C
9	13	14
11	12	13
13	10	17
9	15	7
8	5	9
50	55	60

make the analysis of variance for the given data.

$\bar{X}_A = \frac{50}{5} = 10$, $\bar{X}_B = \frac{55}{5} = 11$, $\bar{X}_C = \frac{60}{5} = 12$, $\bar{\bar{X}} = \frac{\bar{X}_A + \bar{X}_B + \bar{X}_C}{3} = \frac{10+11+12}{3} = \frac{33}{3} = 11$

SSC = Sum of sq. b/w samples (columns)

SSE = —"— within samples (rows)

MSC = mean sum of sq. b/w the samples

MSE = —"— within the samples

TSS ⇒ Total sum of squares

Calculation of SSC

$(\bar{X}_A - \bar{\bar{X}})$	$(\bar{X}_A - \bar{\bar{X}})^2$	$(\bar{X}_B - \bar{\bar{X}})$	$(\bar{X}_B - \bar{\bar{X}})^2$	$(\bar{X}_C - \bar{\bar{X}})^2$
(10−11)=−1	1	(11−11)=0	0	1
(10−11)=−1	1	(11−11)=0	0	1
(10−11)=−1	1	(11−11)=0	0	1
(10−11)=−1	1	(11−11)=0	0	1
(10−11)=−1	1	(11−11)=0	0	4
$\Sigma(\bar{X} - \bar{\bar{X}})$	5		0	5

$$SSC = \Sigma(\bar{X}_A - \bar{\bar{X}})^2 + \Sigma(\bar{X}_B - \bar{\bar{X}})^2 + \Sigma(\bar{X}_C - \bar{\bar{X}})^2$$
$$= 5 + 0 + 5 = 10$$

Source of variation	Sum of square	Degree of freedom	mean square	f Calculated
B/w the sample	SSC = 10	$V_1 = c-1 = 3-1 = 2$	$MSC = SSC/V_1 = \frac{10}{2} = 5$	$\frac{MSC}{MSE} = \frac{5}{11.5}$ = 0.435
within sample	SSE = 138	$V_2 = n-c = 15-3 = 12$	$MSE = \frac{SSE}{V_2} = \frac{138}{12} = 11.5$	

Calculation of SSE

A-10, B-11, C-12

$(A-\bar{X}_A)$	$(A-\bar{X}_A)^2$	$(B-\bar{X}_B)$	$(B-\bar{X}_B)^2$	$(C-\bar{X}_C)$	$(C-\bar{X}_C)^2$
9-10 = -1	1	13-11 = 2	4	2	4
11-10 = 1	1	1	1	1	1
13-10 = 3	9	-1	1	5	25
9-10 = -1	1	4	16	-5	25
8-10 = -2	4	6	36	-3	9
$\Sigma(A-\bar{X}_A)$ = 16			58		64

$$SSE = \Sigma(A-\bar{X}_A)^2 + \Sigma(B-\bar{X}_B)^2 + \Sigma(C-\bar{X}_C)^2$$
$$= 16 + 58 + 64 = 138$$

Calculated f value = 0.435

Tabulated f value $f_{0.05} = 3.89$

Null Hypothesis is passed and no significant variation in the schools

Date:
P. No:

Two way ANOVA :-

Q. The following data represents the no. of units of tablet production (in thousands) per day by five different technicians by using four different type of machines.

Workers	A	B	C	D
P	54	48	57	46
Q	56	50	62	53
R	44	46	54	42
S	53	48	56	44
T	48	52	59	48

(a) Test whether the mean productivity of the different machines are same.?
(b) Test whether the 5 technicians differ w.r.t the mean productivity?

⇒ ① Calculation of Grand total and correction factor.

A-50 B-50 C-50 D-50 ← To make calculation easy

	A	B	C	D	Total
P	+4	-2	+7	-4	+5
Q	+6	0	+12	+3	+21
R	-6	-4	+4	-8	-14
S	+3	-2	+6	-6	+1
T	-2	+2	+9	-2	+7
Total	+5	-6	+38	-17	20 → T

$N \Rightarrow 5A + 5B + 5C + 5D$
$= 20$

$$\text{Correction factor} = \frac{T^2}{N} = \frac{(20)^2}{20} = 20$$

② Calculation of SSC

$$SSC = \frac{(\Sigma A)^2}{n_A} + \frac{(\Sigma B)^2}{n_B} + \frac{(\Sigma C)^2}{n_C} + \frac{(\Sigma D)^2}{n_D} - \text{correction factor}\left(\frac{T^2}{N}\right)$$

$$= \frac{25}{5} + \frac{36}{5} + \frac{1444}{5} + \frac{289}{5} - 20 = 338.8$$

③ Calculation of SSR

$$SSR = \frac{P^2}{n_P} + \frac{Q^2}{n_Q} + \frac{R^2}{n_R} + \frac{S^2}{n_S} + \frac{T^2}{n_T} - \frac{T^2}{N}$$

$$= \frac{25}{4} + \frac{441}{4} + \frac{196}{4} + \frac{1}{4} + \frac{49}{4} - 20 = 158$$

③ Calculation of TSS (Total sum of square)

$$SST = (4)^2 + (6)^2 + (6)^2 + (3)^2 + (2)^2 + (-2)^2 + (0)^2 \ldots (-2)^2 - 20$$

$$SST = 564$$

$$\left(\frac{T^2}{N}\right)$$

④ Calculation of SSE (Total sum of squre due to Error)

$$SSE = SST - (SSC + SSR)$$

$$= 564 - (338.8 + 158) = 67.2$$

Source of variation	Sum of Squares	Degree of freedom	mean sum of squares	Ratio of F
B/w the column	$SSC = 338.8$	$v = c-1 = 4-1 = 3$	$MSC = SSC/c-1 = \frac{338.8}{3} = 112.6$	$MSC/MSE = \frac{112.67}{5.6} = 20.1$
B/w the rows	$SSR = 158$	$v = r-1 = 5-1 = 4$	$MSR = SSR/(r-1) = \frac{158}{4} = 39.5$	$MSR/MSE = \frac{39.5}{5.6} = 7.05$
Residual or Errors	$SSE = 67.2$	$v = (c-1)(r-1) = 3 \times 4 = 12$	$MSE = \frac{SSE}{(C-1)(R-1)} = \frac{67.2}{12} = 5.6$	
//////	$SST = 564$	$v = n-1$	//////	//////

when, $v_1 = 12$, $v_2 = 3$, $F_{0.05} = 3.49$

and when $v_1 = 12$, $v_3 = 4$, $F_{0.05} = 3.20$

(54)

Date:
P. No:

* How to find p-chart ?
→ whether it will be given in que that "find p chart"

and if it is not given then
then it will be given like "find appropriate chart",

If appropriate chart is given the check information in que.

If Item inspected is given different different then 100% we have to choose p-chart.

---

p-chart and np chart only tells that wheather any product is defective or not

---

np chart → To use it see following condition
① It will be directly mentioned in the que to find np-chart
② (OR) if the no. of inspected items will be given same then in that condition we can prefer np chart
③ (OR) if "no. of defective" word is used so go for np chart.

NOTE → p-chart, c-chart, np-chart all these are quality control chart, whereas $\bar{X}$-chart and R-chart are quantity control chart.

UNIT 4 Statistical Quality control.

SQC

Youtube → Amit kumar Thwari

Date:
P. No: 55

(1) $\bar{x}$-chart, R-chart, p-chart, np-chart, c-chart

$\bar{x}$-chart → (mean, Average)

R-chart → (Range)

p-chart → (fraction/proportion)

np-chart → (number of defective)

c-chart → (Defects)

P-chart

Q.

Sample (m) Number	Item inspected (n)	Defective (d) Items	$p = \frac{d}{n}$
1	200	8	8/200 ⇒ 0.040
2	"	12	12/200 ⇒ 0.060
3	"	2	0.10
4	"	20	0.100
5	"	10	0.050
6	"	15	0.075
7	"	6	0.030
8	"	20	0.100
9	"	13	0.065
10	"	9	0.045
11	"	16	0.080
12	"	10	0.050
13	"	13	0.065
14	"	6	0.030
15	"	8	0.040

$\Sigma p = 0.840$

Central Limit

$C.L = \bar{p} = \frac{\Sigma p}{m}$ → total no. of defective; m → total no. of sample

P (fraction defective)

$= \frac{d}{n}$ d → No. of def. Items; n → No. of item inspected

upper central limit

$U.C.L = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n \text{ or } \bar{n}}}$

Lower central limit.

$L.C.L = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n \text{ or } \bar{n}}}$

$\bar{n}$ → when item inspected will be given different then, we will take avg of ⇒ $\frac{\Sigma n}{15}$ ← Total no. of sample

56

Date:
P. No:

$C.L = \bar{p} = \frac{\Sigma p \rightarrow \text{total no. of defectives}}{m \rightarrow \text{total no. of samples}} = \frac{0.840}{15}$

$\boxed{\bar{p} = 0.056}$

$U.C.L = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n \text{ or } \bar{n}}}$

$= 0.056 + 3\sqrt{\frac{0.056(1-0.056)}{200}}$

$= 0.056 + 3\sqrt{0.0002643}$

$= 0.056 + 3 \times (0.01625)$

$\boxed{UCL = 0.10475}$

$L.C.L = \bar{p} - 3 \times \sqrt{\frac{\bar{p}(1-\bar{p})}{n \text{ or } \bar{n}}}$

$LCL = 0.056 - 3(0.01625)$

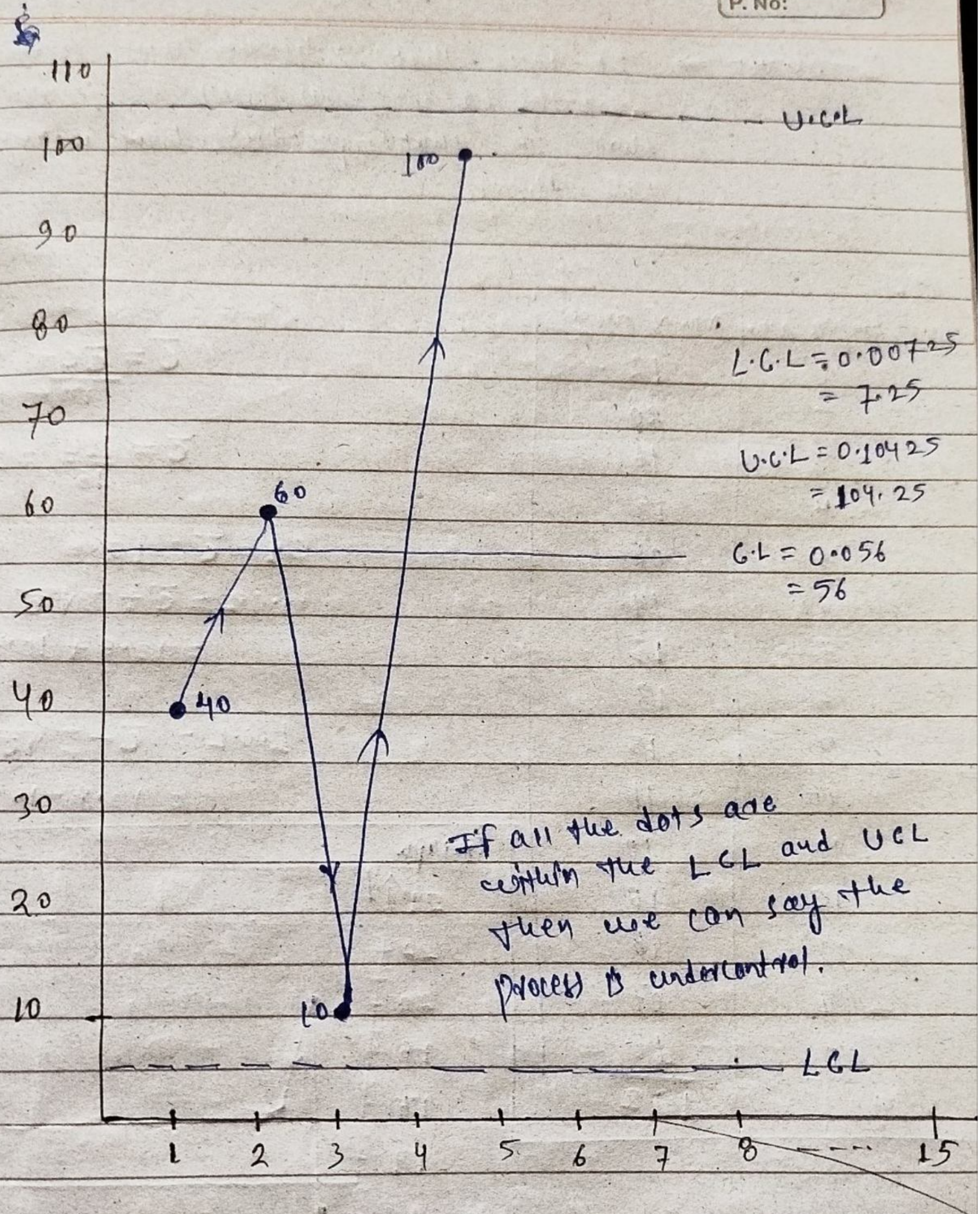
$\boxed{LCL = 0.00725}$

		$\times 1000$
1	0.040	40
2	0.060	60
3	0.010	10
4	0.100	100
5	0.050	50
6	0.075	75
7	0.030	30
8	0.100	100
9	0.065	65
10	0.045	45
11	0.080	80
12	0.050	50
13	0.065	65
14	0.030	30
15	0.040	40

PTO

57

Date:
P. No:



L.C.L = 0.00725
= 7.25

U.C.L = 0.10425
= 104.25

C.L = 0.056
= 56

If all the dots are within the LCL and UCL then we can say the process is undercontrol.

P-chart

when you've drawn the graph

C-chart → It tells that wheather any product is defected or not as well as it also tells that in that product how many defects are there.

Q.

(m) Sample no.	No. of defects (C)
1	12
2	20
3	16
4	15
5	08
6	14
7	13
8	13
9	15
10	11
11	10
12	10
13	18
14	07
15	10
	192

$\bar{C} = \frac{\Sigma C}{m}$ ← Total no. of defects / ← total no. of samples

$\bar{C} = \frac{192}{15} = 12.8$

$C.L = \bar{C} = 12.8$

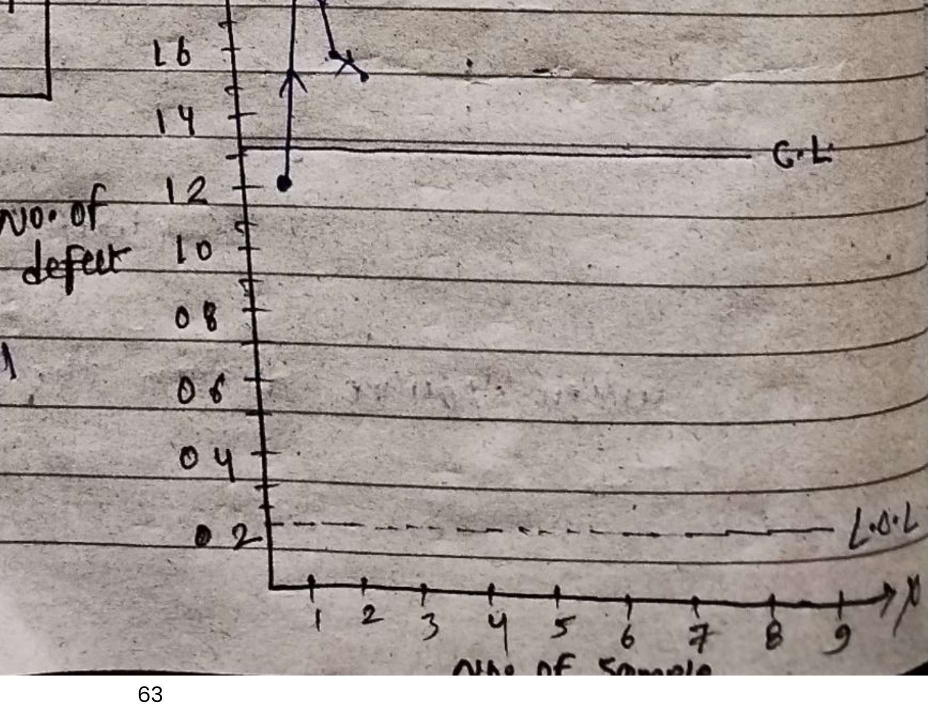
$U.C.L = \bar{C} + 3\sqrt{\bar{C}}$

$= 12.8 + 10.73 = 23.53$

$L.C.L = \bar{C} - 3\sqrt{\bar{C}}$

$= 12.8 - 10.73 = 2.07$

Graph



our product is under control

59

Date:
P. No:

np chart :-

Sample	No. of defective items
1	25
2	47
3	23
4	30
5	34
6	24
7	39
8	32
9	35
10	22
11	45
12	40

$\Sigma np = 396$

Sample each of 250

$n\bar{p} = \frac{\Sigma np}{m}$ → no. of defective items; m → total no. of sample

$= \frac{396}{12} = 33$

$\bar{p} = \frac{n\bar{p}}{n}$ n — inspected items

$= \frac{33}{250} = 0.132$

$\boxed{C.L = n\bar{p} = 33}$

$U.C.L = n\bar{p} + 3\times\sqrt{n\bar{p}(1-\bar{p})}$

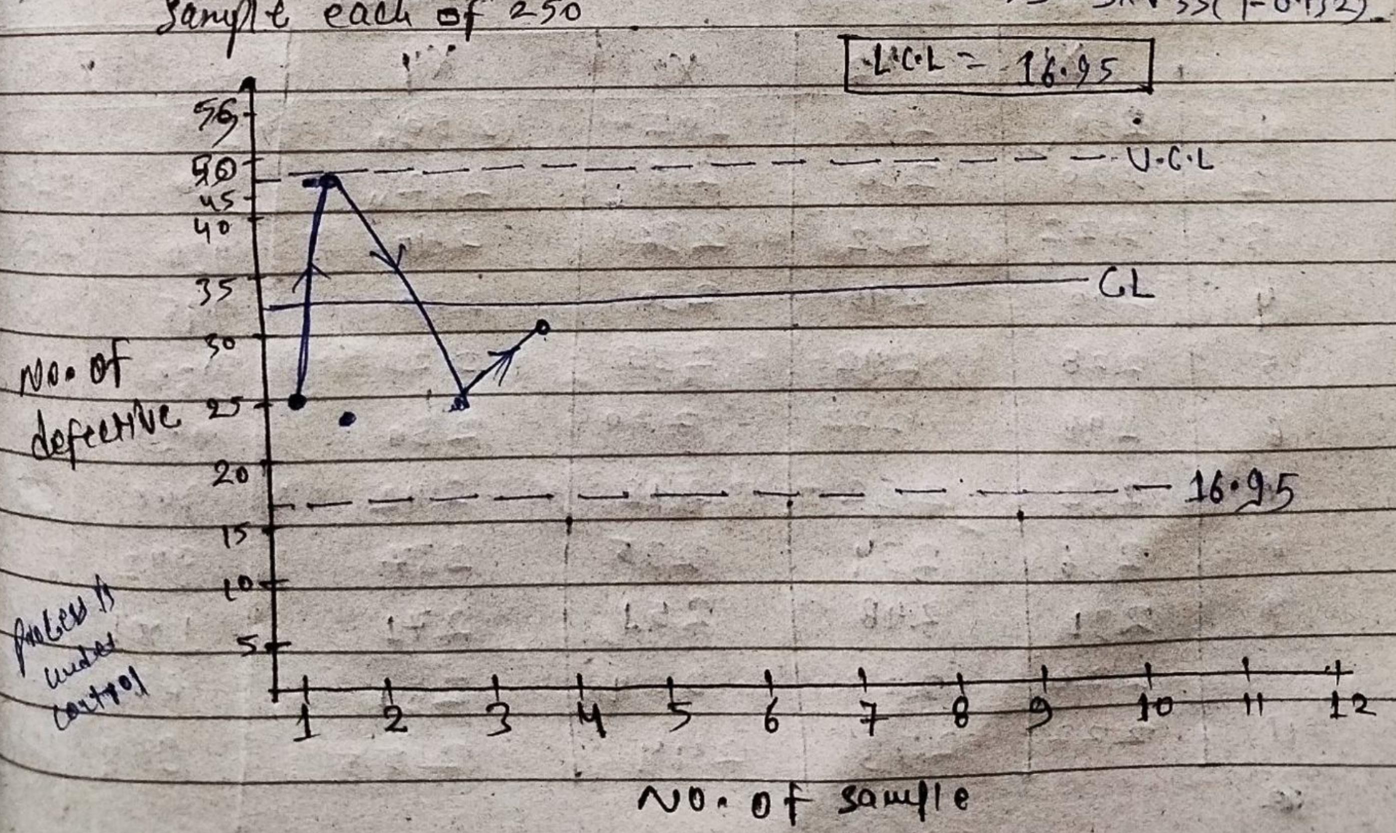
$= 33 + 3\times\sqrt{33(1-0.132)}$

$\boxed{U.C.L = 49.05}$

$L.C.L = n\bar{p} - 3\times\sqrt{n\bar{p}(1-\bar{p})}$

$= 33 - 3\times\sqrt{33(1-0.132)}$

$\boxed{L.C.L = 16.95}$



(60)

$\bar{X}$ and $\bar{R}$ charts

Date:
P. No:

$\bar{X}$-chart

$C.L = \bar{\bar{X}}$, $L.C.L = \bar{\bar{X}} - A_2\bar{R}$

$U.C.L = \bar{\bar{X}} + A_2\bar{R}$

$\bar{R}$-chart → is also a range chart

$C.L = \bar{R}$

$L.C.L = D_3\bar{R}$

$U.C.L = D_4 R$

$\bar{R} = \frac{\Sigma R}{m}$, $R \rightarrow (x_H - x_L)$

Higher Value (x_H), Lower Value (x_L), Range

Q. A factory produces 50 cylinders per hour. Samples of 4 cylinders are taken at random from the production at every hour. And the diameter of cylinders are measured. Draw $\bar{X}$ (mean) chart and r chart and decide whether the process is under control or not.

[for $n=4$, $A2 = 0.73$, $D3 = 0$, $D4 = 2.28$]

Sample no.	Diameters of Cylinders						
	X1	X2	X3	X4	Σx_i	$\bar{x} = \frac{\Sigma x_i}{4}$	R
1	230 L	+ 238	+ 242 +	+ 250 H	⇒ 960	240	20
2	220	230	218 L	242 H	910	227.50	24
3	222	232	236	240	930	232.5	18
4	250	240	230	225	945	236.25	25
5	228	242	235	225	930	232.5	17
6	248	222	220	230	920	230	28
7	232	232	242	242	948	237	10
8	236	234	235	237	942	235.5	3
9	231	248	251	271	1001	250.25	40
10	220	222	224	231	897	224.25	11
11	222	233	244	255	954	238.5	33

	X_1	X_2	X_3	X_4	Σx_i	$\bar{X} \Rightarrow \frac{\Sigma x_i}{4}$	R
12	272	262	265	225	1024	256	47
13	218	268	274	250	1010	252.5	56
14	214	218	252	262	946	236.5	48
15	260	262	265	263	1050	262.5	05

$\Sigma \bar{X} = 3591.75$

$\Sigma R = 385$

$$\bar{\bar{X}} = \frac{\Sigma \bar{X}}{m} = \frac{3591.75}{15} = 239.45$$

$\bar{X}$ - chart

$C.L = \bar{\bar{X}} = 239.45$

Given

$U.C.L = \bar{\bar{X}} + A_2\bar{R} = 239.45 + 0.73 \times \bar{R}$ — (1)

$\bar{R}$ (Range) $\to$ 385 , $\bar{R} \to \frac{385}{15} \Rightarrow 25.66$

(Highest 250) - (lowest 230)

Now from eqn (1)

$U.C.L = \bar{\bar{X}} + A_2\bar{R}$

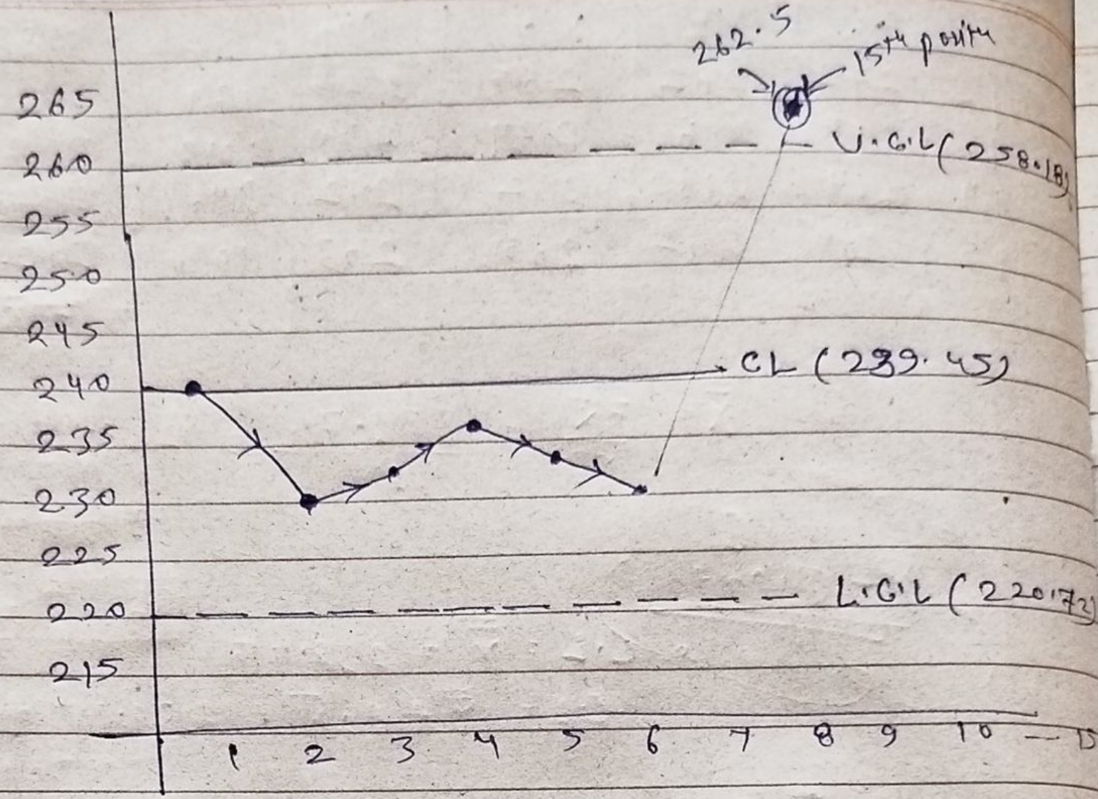
$= 239.45 + 0.73 \times 25.66 \Rightarrow \boxed{258.18}$

$L.C.L = \bar{\bar{X}} - A_2\bar{R} \Rightarrow \boxed{220.72}$

PTO for $\bar{X}$ - chart

(62)

Date:
P. No:



So the value of $\bar{x}$ for sample 15th falls outside the upper control limit hence the process is not under control.

$\bar{R}$-Chart

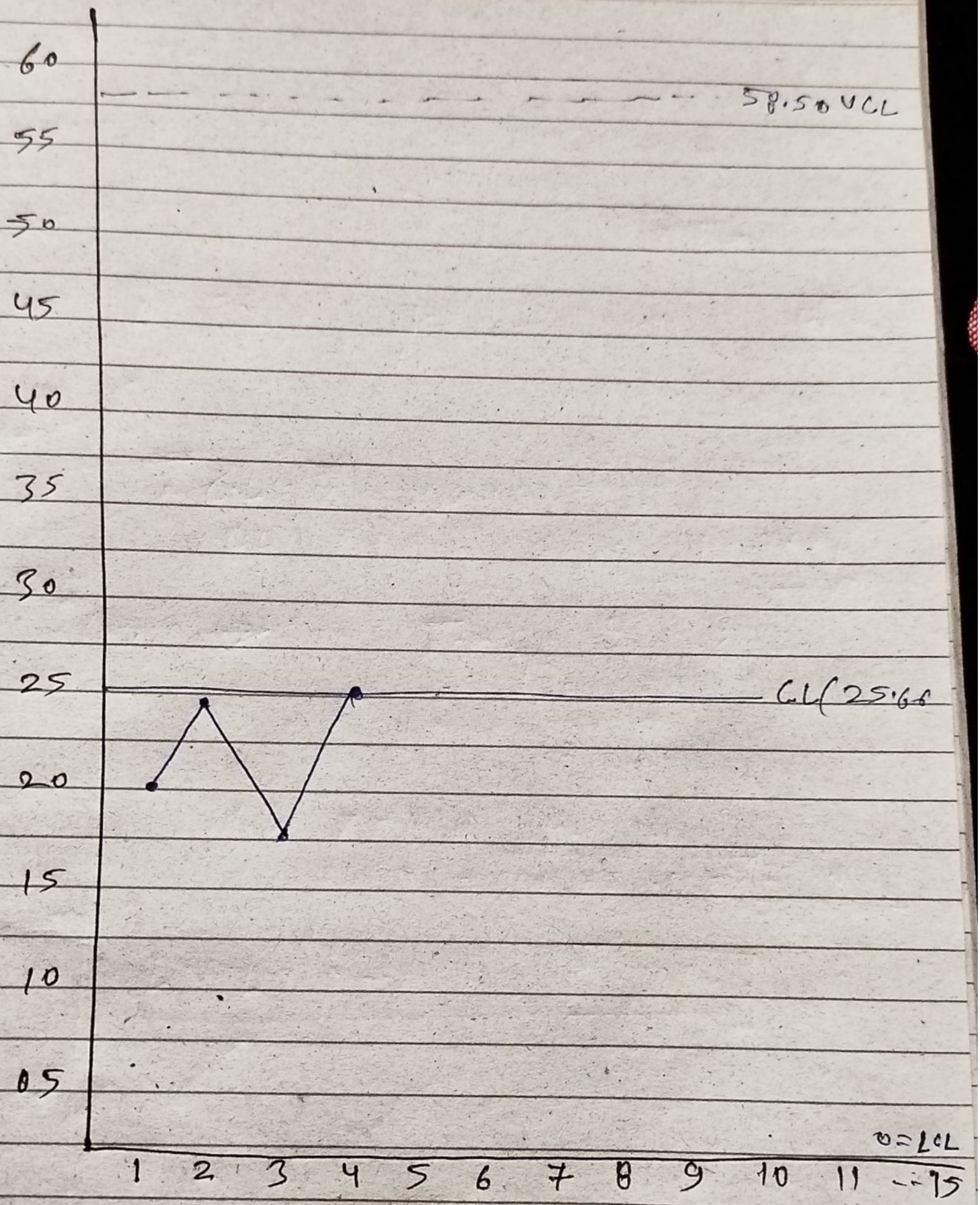
$$C.L = \bar{R} = 25.66$$

$$U.C.L = D_4\bar{R} = 2.28 \times 25.66 = 58.50$$

$$L.C.L = D_3\bar{R} = 0 \times 25.66 = 0$$

PTO for $\bar{R}$-Chart

Date:
P. No:



the process is under control.