SAMPLING DISTRIBUTION | CORRELATION & REGRESSION ANALYSIS

$(\bar{x})$ MEAN = DATA RANGE $\dfrac{(\text{MAXIMUM NUMBER} + \text{MINIMUM NUMBER})}{2}$

STANDARD DEVIATION = $S = \sqrt{\dfrac{\Sigma x^2 - n(\bar{x})^2}{n-1}}$

SAMPLE ERROR = $x - \mu$

= MAXIMUM NUMBER AT DATA RANGE − MEAN

MEAN OF DISTRIBUTION = $\dfrac{\Sigma \text{ MEAN}}{\text{TOTAL SAMPLE SIZE}}$

SAMPLE MEAN = $\dfrac{\Sigma \text{ TOTAL}}{\text{NO. OF EVENTS}}$

pb (23)

A POPULATION OF 5 SALES PERSONS SELLING CAR TELEPHONES TO BOTH PRIVATE AND COMMERCIAL CUSTOMERS OVER A PERIOD OF ONE MONTH. THE NUMBER OF TELEPHONES SOLD FOR EACH PERSON IS SHOWN IN FOLLOWING TABLE

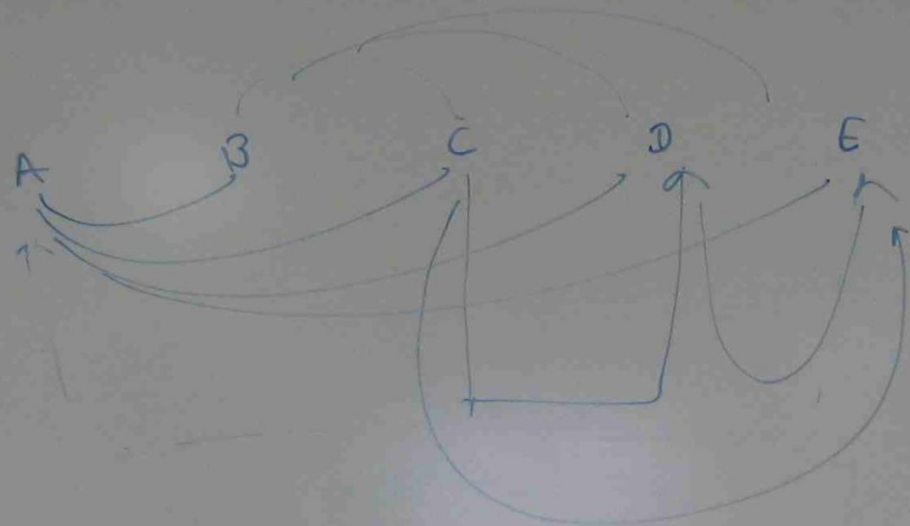
SALES PERSON	PHONES SOLD
ADAM (A)	14 ✓
BAKER (B)	20 ✓
COLLINS (C)	12
DAVID (D)	8
EDWARDS (E)	16

(a) LIST ALL POSSIBLE SAMPLES OF SIZE 2

(b) FOR EACH SAMPLE, CALCULATE DATA RANGE, SAMPLE MEAN, SAMPLE STANDARD DEVIATION

(c) CALCULATE MEAN, FREQUENCY, RELATIVE FREQUENCY

(d) GRAPH.



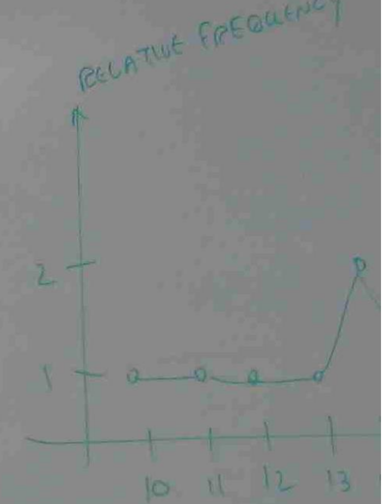
→ AB, AC, AD, AE, BC, BD, BE, CD, CE, DE

LAST − MEAN

SAMPLE COMBINATION	DATA RANGE	MEAN $= \frac{MAX + MIN}{2}$, $\bar{X}$	STANDARD DEVIATION	SAMPLE ERROR $x - \mu$
AB	14 − 20	$\bar{x} = \frac{14+20}{2} = 17$	$\sqrt{\frac{14^2+20^2-2\times17^2}{2-1}} = 4.24$	20 − 17 = 3
AC	14 − 12	$\bar{x} = \frac{14+12}{2} = 13$	$\sqrt{\frac{14^2+12^2-2\times13^2}{2-1}} = 1.414$	12 − 13 = −1
AD	14 − 8	11	4.243	−3
AE	14 − 16	15	1.414	+1
BC	20 − 12	16	5.657	+2
BD	20 − 8	14 ←	8.485	0
BE	20 − 16	18	2.828	4
CD	12 − 8	10	2.828	−4
CE	12 − 16	14 ←	2.828	0
[illegible]	[illegible]	[illegible]	5.657	−2

MEAN	FREQUENCY	RELATIVE FREQUENCY
10	1	$\frac{1}{\Sigma f} = \frac{1}{10} = 0.1$
11	1	$\frac{1}{10} = 0.1$
12	1	$\frac{1}{10} = 0.1$
13	1	$\frac{1}{10} = 0.1$
14	2	$\frac{2}{10} = 0.2$
15	1	$\frac{1}{10} = 0.1$
16	1	$\frac{1}{10} = 0.1$
17	1	$\frac{1}{10} = 0.1$
18	1	$\frac{1}{10} = 0.1$

$\Sigma f = 10$

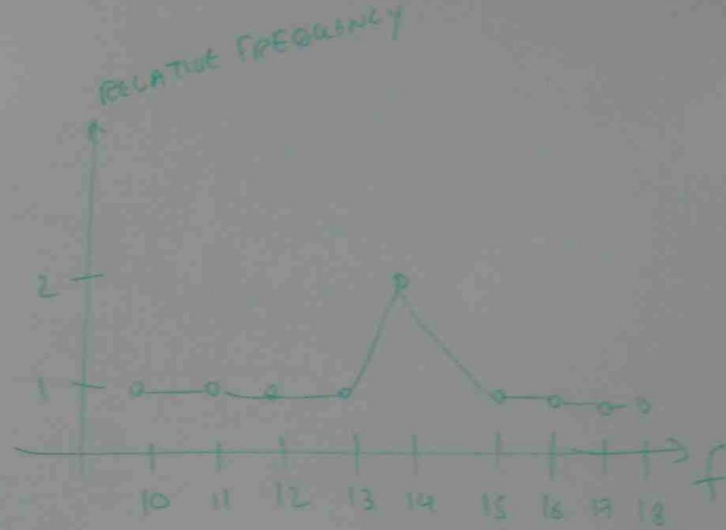


Pb (29) IN ABOVE
DISTRIBUT

(i) MEAN OF DISTRIBUTION

ACCORDING TO FREQUENCY DISTRIBUTION,
B+D, C+E CAN SELL MORE PRODUCTS.

(ii) SAMPLE MEAN =



Pb 29 IN ABOVE PROBLEM, CALCULATE (i) MEAN OF DISTRIBUTION μ_x, (ii) SAMPLE MEAN

(i) MEAN OF DISTRIBUTION $= \dfrac{\sum \text{MEAN}}{\text{TOTAL SAMPLE SIZE}} =$

$$= \frac{17+13+11+15+16+14+18+10+14+12}{10}$$

$$= 14$$

SALES

(ii) SAMPLE MEAN $= \dfrac{\sum \text{TOTAL EVENTS}}{\text{TOTAL PERSON}} = \dfrac{14+20+12+8+16}{5}$

$$= 14$$

STANDARD DISTRIBUTION OF SELECTED SAMPLE

STANDARD DISTRIBUTION OF SELECTED SAMPLE $\delta x = \dfrac{\text{PARENT STANDARD DISTRIBUTION}}{\sqrt{\text{NUMBER OF SAMPLE}}}$

pb 30 A CERTAIN BRAND OF ROPE IS KNOWN TO HAVE A MEAN BREAKING STRENGTH 25 kg WITH A STANDARD DEVIATION OF 0.5 kg. A RANDOM SAMPLE OF ROPE IS TESTED FOR BREAKING STRENGTH. WHAT IS THE POSSIBILITY THAT SAMPLE MEAN BREAKING STRENGTH OF 50 PIECES OF ROPE WILL BE (i) BETWEEN 24.9 AND 25.1 kg (ii) LESS THAN 24.8 kg.

MEAN = 25 kg $\quad \delta = 0.5$ kg ← PARENT DISTRIBUTION

(μ)

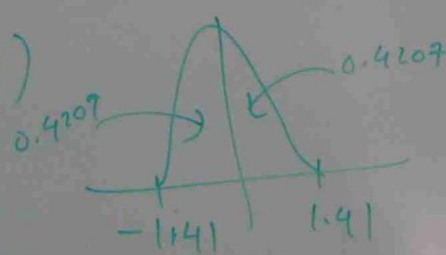
STANDARD DEVIATION OF SELECTED SAMPLE $(\delta_x) = \dfrac{\text{PARENT STANDARD DEVIATION}}{\sqrt{\text{NO. OF SAMPLE}}} = \dfrac{0.5}{\sqrt{50}} = 0.071$

(a) BETWEEN 24.9 AND 25.1 kg

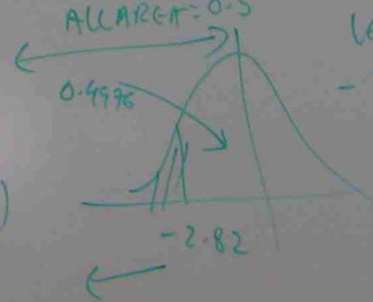
$$P(24.9 < x < 25.1) \longrightarrow P\left(\frac{x_1-\mu}{\delta x} < x < \frac{x_2-\mu}{\delta x}\right) = P\left(\frac{24.9-25}{0.071} < x < \frac{25.1-25}{0.071}\right)$$

TOTAL AREA = 0.4207 + 0.4207 = 0.8414

$$P(-1.41 < x < 1.41)$$



$$\text{(b) } P\left(\frac{x-\mu}{\delta x}\right) = P\left(\frac{24.8-25}{0.071}\right) = P(-2.82)$$



LESS THAN = ALL − 2.82 AREA

− 2.82

= 0.5 − 0.4976

= 0.0024

CORRELATION AND REGRESSION ANALYSIS

POSSIBLE RELATIONSHIPS BETWEEN X AND Y SCATTER PLOT

DIRECT LINEAR

INVERSE LINEAR

INVERSE CURVILINEAR

DIRECT CURVILINEAR

NON RELATIONSHIP

CORRELATION ANALYSIS → ACTIVITY & EFFECTIVENESS

INPUT

$$\bar{X} = \frac{X_1 + Y_2 + - -}{n}$$

AVERAGE

STANDARD DEVIATION

$$S_X = \sqrt{\frac{X_1^2 + X_2^2 + \cdots - n(\bar{X})^2}{n-1}}$$

EFFECT

OUTPUT

$$\bar{Y}_{AVERAGE} = \frac{Y_1 + Y_2 + Y_3 + \cdots}{n}$$

$$S_y = \sqrt{\frac{y_1^2 + y_2^2 + \cdots - n(\bar{y})^2}{n-1}}$$

Σxy

$SS_{xy} = \Sigma xy - n\bar{x}\bar{y}$

$SS_x = (n-1)s_x^2$

$SS_y = (n-1)s_y^2$

CORRELATION $r = \dfrac{SS_{xy}}{\sqrt{SS_x \times SS_y}}$

IF r IS POSITIVE
CLOSE TO +1
THEN INPUT IS STRONGLY RELATED TO OUT PUT

IF r IS NEGATIVE, INPUT DOES NOT EFFECTED ON OUT PUT

pb 32

A CERTAIN LARGE COMPANY IS INTERESTED IN WHETHER OR NOT THERE IS A RELATIONSHIP BETWEEN THE AMOUNT SPENT ON ADVERTISING PER MONTH AND THE GROSS MONTHLY SALE VOLUME.
TO EXAMINE THIS PROBLEM, THE FOLLOWING SAMPLE OF ADVERTISING EXPENDITURES AND ASSOCIATED SALE VOLUMES.
VOLUMES ARE RANDAMLY SELECTED FOR MONTHLY BASIS

MONTH	ADVERTISING EXPENDITURE (X) \$ 10000	SALE (Y) (X \$ 10000)
1	1.2	101
2	0.8	92
3	1.0	110
4	1.3	120
5	0.7	90
6	0.8	82
7	1.0	93
8	0.6	75
9	0.9	91
10	1.1	108

ADVERTISING

MEAN $\bar{x} = \dfrac{1.2+0.8+1.0+1.3+0.7+0.8+1.0+0.6+0.9+1.1}{10} = 0.94$

$S_x = \sqrt{\dfrac{1.2^2+0.8^2+1.0^2+1.3^2+0.7^2+0.8^2+1.0^2+0.6^2+0.9^2+1.1^2-10\,(0.94)^2}{10-1}} = 0.22211$

SALE

MEAN = $\bar{y} = \dfrac{101+92+110+120+90+82+93+75+91+108}{10} = 95.9$

$S_y = \sqrt{\dfrac{101^2+92^2+110^2+120^2+90^2+82^2+93^2+75^2+91^2+108^2-10\times 95.9^2}{10-1}} = 13.37$

$\sum xy = 1.2\times101+0.8\times92+1.0\times110+1.3\times120+0.7\times90+.8\times82+1.0\times93+0.6\times75+0.9\times91+1.1\times108$
$= 924.8$

$SS_{xy} = \sum xy - n\bar{x}\bar{y} = 924.8-10\times0.94\times95.9 = 23.34$

$SS_x = (n-1)S_x^2 = (10-1)\times(0.22211)^2 = 0.444$

$SS_y = (n-1)S_y^2 = (10-1)\times(13.37)^2 = 1600.9$

$r = \dfrac{SS_{xy}}{\sqrt{SS_x \times SS_y}}$

$= \dfrac{23.34}{\sqrt{0.444\times1600.9}} = 0.875$

POSITIVE

CLOSE TO +1

THUS ADVERTISEMENT AFFECTS THE SALE

$\alpha = 0.05$

v_2 \ v_1	1	2	3	4	5	6	7	8
1								
2								
3								
4								
5								
6								
7								

CATEGORY	OBSERVED FREQUENCY (O)	E[illegible] CUST[illegible]
A	48	200[illegible]
B	98	20[illegible]
C	54	20[illegible]

$\chi^2_{\alpha, v} =$

CATEGORY	OBSERVED FREQUENCY (O)	EXPECTED FREQUENCY (E) CUSTOMER × % SHARE	O − E	$\frac{(O-E)^2}{E}$
A	48	$200 \times \frac{30}{100} = 60$	$O-E = 48-60 = -12$	$\frac{(-12)^2}{60} = 2.4$
B	98	$200 \times \frac{50}{100} = 100$	$O-E = 98-100 = -2$	$\frac{(-2)^2}{100} = 0.04$
C	54	$200 \times \frac{20}{100} = 40$	$54-40 = 14$	$\frac{(14)^2}{40} = 4.9$
			$\sum O-E = -12+(-2)+14 = 0$	$\sum \frac{(O-E)^2}{E} = 7.34$

$$\chi^2_{\alpha, v} = \chi^2_{0.05,\ (\text{NO. OF PRODUCTS} - 1)}$$

K

$$= \chi^2_{0.05,\ (3-1)} = \chi^2_{0.05,\ 2}$$

v_2 $v_1 = 2$

4 — 6.44

REJECTION POINT

FIND FIRST NUMBER LESS THAN 7.34 IN χ^2 TABLE

ph

pe

T

E

STATISTICS

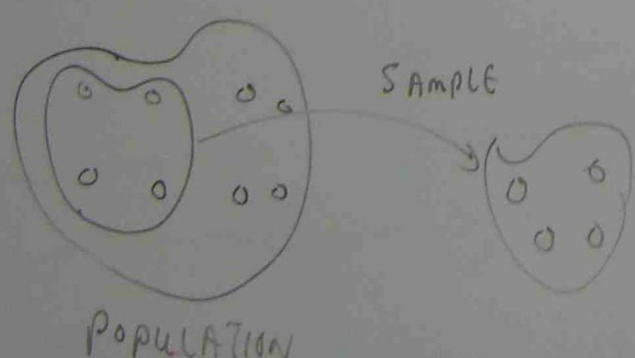
STATISTICS INVOLVES COLLECTING, SUMMARISING ANALYZING AND INTERPRETING DATA.

TYPES OF DATA

CATEGORICAL DATA

NUMERICAL DATA

DISCRETE CONTINUOUS



DISCRETE – COUNTABLE SUCH AS NUMBER OF DAYS.

CONTINUOUS – INFINITE SUCH AS DAILY TEMPERATURE

PARAMETER

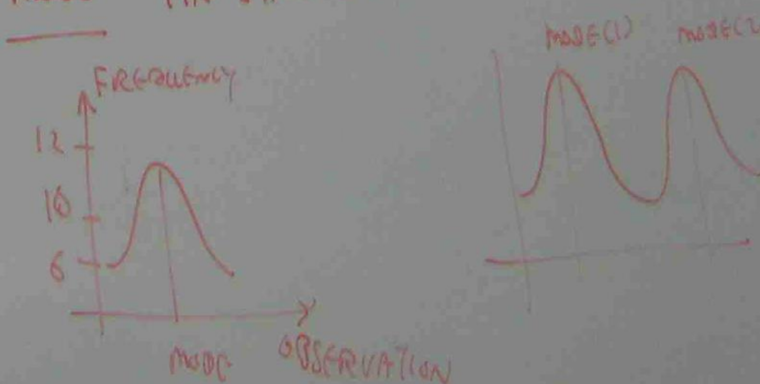
A NUMERICAL MEASURE THAT DESCRIBES SOME CHARACTERISTICS OF POPULATION

STATISTICS

· NUMERICAL MEASURE THE DESCRIBES SOME CHARACTERISTICS OF SAMPLE

MODE

THE DATA SET THAT OCCURS MOST OFTEN



NO MODE

pb (1) FIND THE MODE OF THE FOLLOWING SAMPLES OF DATA

(a) 18, 19, 18, 20, 18, 18, 20, 21, 37, 18

(b) 27, 28, 29, 36, 14, 24

(c) 19, 18, 20, 39, 18, 18, 19, 46, 19

(a) MODE = 18

(b) NO MODE

(c) 19, 18 = MODES

| 18 = MODE 1
| 19 = MODE 2

MEDIAN

THE MIDDLE NUMBER IN AN ORDERED SET.

pb (2) FIND THE MEDIAN OF EACH OF THE FOLLOWING SAMPLES OF DATA

(a) X : 7, 3, 2, 8, 11

(b) X : 18, 19, 18, 20, 18, 18, 20, 21, 37, 18

(a) 2, 3, 7, 8, 11

MEDIAN = 7

(b) X = 18, 18, 18, 18, 18, 19, 20, 20, 21, 37

MEDIAN =

MEAN

MEASURE

BY ADD

DIVID

THE

$$\text{MEDIAN} = \frac{18+19}{2}$$

$$= 18.5$$

MEAN: THE ARITHMETIC AVERAGE OF ALL MEASUREMENTS IN THE DATA SET OBTAINED BY ADDING THEM ALL UP AND DIVIDING THIS SUM BY THE SIZE OF THE DATA SET.

$$\bar{X} = \frac{\sum X}{n}$$

, 37

pb (3) FIND THE MEAN OF THE FOLLOWING SAMPLES OF DATA

(i) X : 11, 14, 10, 5

(ii) X : 2, 2, 4, 5, 5, 5, 6, 7, 8, 9, 11

(i) $\text{MEAN} = \bar{x} = \frac{11+14+10+5}{4} = \frac{40}{4} = 10$

(ii) $\text{MEAN} = \bar{x} = \frac{2+2+4+5+5+5+6+7+8+9+11}{11}$

$$= \frac{64}{11} = 5.818$$

pb (4) FIND THE MEAN OF THE FOLLOWING GROUP FREQUENCY DISTRIBUTION (GFD)

INTERVAL	FREQUENCY	INTERVAL	FREQUENCY
$0 \to 9$	1	$50 \to 59$	1
$10 \to 19$	2		
$20 \to 29$	4		
$30 \to 39$	8		
$40 \to 49$	4		

ency

INTERVAL	MEAN	f	f x
$0 \longrightarrow 9$	$\frac{0+9}{2} = 4.5$	1	$4.5 \times 1 = 4.5$
$10 \longrightarrow 19$	$\frac{10+19}{2} = 14.5$	2	$14.5 \times 2 = 29$
$20 \longrightarrow 29$	$\frac{20+29}{2} = 24.5$	4	$24.5 \times 4 = 98$
$30 \longrightarrow 39$	$\frac{30+39}{2} = 34.5$	8	$34.5 \times 8 = 276$
$40 \longrightarrow 49$	$\frac{40+49}{2} = 44.5$	4	$44.5 \times 4 = 178$
$50 \longrightarrow 59$	$\frac{50+59}{2} = 54.5$	1	$54.5 \times 1 = 54.5$
		$\Sigma f = 20$	$\Sigma fx = 640$

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{640}{20} = 32$$

RANGE

THE DIFFERENCE BETWEEN THE LARGEST AND SMALLEST VALUES IN THE SAMPLE.

Pb 5 FIND THE RANGE OF THE FOLLOWING SAMPLE OF DATA.

(i) 18, 19, 18, 20, 18, 37, 21, 20, 18, 18

(ii) 18, 19, 18, 20, 18, 21, 20, 18, 18

(i) RANGE = LARGEST − SMALLEST

$= 37 - 18 = 19$

(ii) RANGE = LARGEST − SMALLEST

$= 21 - 18 = 3$

RESIDUAL

THE DIFFERENCE BETWEEN EACH SAMPLE x AND THE MEAN $\overline{x}$ IS CALLED THE DEVIATION FROM THE MEAN OR RESIDUAL.

$$\text{RESIDUAL} = x - \overline{x}$$

Pb (6) 10, 12, 15, 17, 21

CALCULATE (i) MEAN (ii) RESIDUAL (DEVIATION)

(i) $\overline{x} = \dfrac{10+12+15+17+21}{5} = \dfrac{75}{5} = 15$

(ii) RESIDUAL $= (x - \overline{x}) = (10-15), (12-15), (15-15), (17-15)$

$= -5, -3, 0, 2, 6$

VARIANCE

$$\text{VARIANCE} = \frac{\sum (x - \overline{x})^2}{n-1}$$

pb 7

CALCULATE VARIANCE OF 10, 12, 15, 17, 21

$$\bar{X} = \text{MEAN} = \frac{\Sigma x}{n} = \frac{10+12+15+17+21}{5} = 15$$

$$\text{VARIANCE} = s^2 = \frac{\Sigma (x-\bar{x})^2}{n-1} = \frac{(10-15)^2 + (12-15)^2 + (15-15)^2 + (17-15)^2 + (21-15)^2}{5-1}$$

$$= \frac{(-5)^2 + (-3)^2 + (0)^2 + (2)^2 + (6)^2}{4}$$

$$= \frac{25+9+4+36}{4} = \frac{74}{4} = 18.5$$

STANDARD DEVIATION

THE STANDARD DEVIATION OF A SAMPLE OF DATA IS DEFINED TO THE POSITIVE SQUARE ROOT OF THE VARIANCE

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

$(21-15)^2$

pb ⑧ CALCULATE STANDARD DEVIATION OF 10, 12, 15, 17, 21

$$S = \sqrt{\frac{\Sigma (x-\bar{x})^2}{n-1}} = \sqrt{\frac{(10-15)^2+(12-15)^2+(15-15)^2+(17-15)^2+(21-15)^2}{5-1}}$$

$$\bar{x} = \frac{10+12+15+17+21}{5} = 15$$

$$= \sqrt{18.5} = 4.3$$

CALCULATION OF S^2 (VARIANCE) AND S (STANDARD DEVIATION) FOR A SAMPLE OF RAW DATA

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

pb 9

CALCULATE THE VARIANCE AND STANDARD DEVIATION FOR THE FOL

SAMPLE OF DATA.

$$X: 2, 3, 5, 5, 9$$

$$\bar{X} = \frac{\sum X}{n} = \frac{2+3+5+5+9}{5} = \frac{24}{5} = 4.8$$

$$\sum X^2 = 2^2 + 3^2 + 5^2 + 5^2 + 9^2 = 144$$

$$n(\bar{X})^2 = 5 \times (4.8)^2 = 115.2$$

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

$$s^2 = \frac{144 - 115.2}{5-1} = 7.2$$

$$s = \sqrt{7.2} = 2.68$$

VARIAN

STAND
DEVIA

pb 1

$\bar{X} =$

$$s^2 = \frac{\sum f x^2 - n(}{n-1}$$

$s = \sqrt{32.85} = 5.$

...VIATION FOR THE FOLLOWING

...4.8

...$)^2$

CALCULATION OF s^2 AND s FOR SAMPLE DATA PRESENTED AS GFD (GROUP FREQUENCY DISTRIBUTION).

VARIANCE: $s^2 = \dfrac{\sum f(x-\bar{x})^2}{n-1} = \dfrac{\sum f(x-\bar{x})^2}{\sum f - 1} = \dfrac{\sum fx^2 - n(\bar{x})^2}{n-1}$

STANDARD DEVIATION $= s = \sqrt{\dfrac{\sum fx^2 - n(\bar{x})^2}{n-1}}$

Pb 10 FIND THE VARIANCE AND STANDARD DEVIATION OF THE FOLLOWING SAMPLE OF GROUPED DATA.

INTERVAL	FREQUENCY
10 – 14	4
15 – 19	7
20 – 24	5
25 – 29	3
30 – 34	1

INTERVAL	x (MID POINT)	f	fx	fx^2
10 – 14	$\frac{10+14}{2} = \frac{24}{2} = 12$	4	$4 \times 12 = 48$	$4 \times 12^2 = 576$
15 – 19	$\frac{15+19}{2} = 17$	7	$17 \times 7 = 119$	$7 \times 17^2 = 2023$
20 – 24	$\frac{20+24}{2} = 22$	5	$22 \times 5 = 110$	$5 \times 22^2 = 2420$
25 – 29	$\frac{25+29}{2} = 27$	3	$27 \times 3 = 81$	$3 \times 27^2 = 2187$
30 – 34	$\frac{30+34}{2} = 32$	1	$32 \times 1 = 32$	$1 \times 32^2 = 1024$
		$\sum f = 20$	$\sum fx = 390$	$\sum fx^2 = 8230$

$\bar{x} = \dfrac{\sum fx}{\sum f} = \dfrac{390}{20} = 19.5$

$s^2 = \dfrac{\sum fx^2 - n(\bar{x})^2}{n-1} = \dfrac{8230 - 20 \times (19.5)^2}{20-1} = \dfrac{8230 - 7605}{19} = 32.89$

$s = \sqrt{32.89} = 5.739$