QUERTILE DEVIATION:

Half of the difference b|t third and first quertile is called semi-interquartile range / quartile deviation

$$Q.D = \frac{Q_3 - Q_1}{2}$$

Coefficient of Quarite $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Interquartile Deviation is $Q_3 - Q_1$

For ungrouped Data:

$$Q_1 = \left(\frac{N}{4}\right)^{th} \text{item}$$

$$Q_2 = \left(\frac{N}{2}\right)^{th} \text{item}$$

$$Q_3 = \left(\frac{3N}{4}\right)^{th} \text{item.}$$

For Grouped Data:

$$Q_r = l_1 + \frac{\left[r\left(\frac{N}{4}\right) - C\right]}{f}(l_2 - l_1).$$

where, $Q_r \rightarrow$ the r^{th} quartile

l_1 = lower limit of quartile class

l_2 = upper limit of quartile class

L = Tot frequency

f = Frequency

c = Value preceding to quartile class in cumulative frequency

---

From the following table giving marks of students, calculate the interquartile range, quartile deviation and coefficient of quartile deviation.

Marks of Students (x)	60	62	68	70	75	80	88	90
No. of Students (f)	25	21	28	18	24	20	24	17
Cummulative Frequency	25	46	74	92	116	136	160	177

$$Q_1 = \frac{199}{4}$$

$$Q_1 = 49.75$$

$$\therefore Q_1 = 68$$

$$Q_3 = \frac{3 \times 199}{4}$$

$$Q_3 = \frac{59.70}{4}$$

$$Q_3 = 14.9.3 \qquad Q_3 = 88$$

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{88 - 68}{2}$$

$$= 10$$

$$\text{Coff. of } Q.D = \frac{20}{156}$$

$$= 0.13$$

Calculate inquartile range, quaritil deviation and coeffs. of quartile deviation from the following

Size x	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	9	15	29	24	3	20
Cummulative Frequency	9	24	53	77	80	100

$$Q_1 = \frac{100}{4} \qquad Q_3 = \frac{3 \times 100}{4}$$

$$Q_1 = 25 \qquad Q_3 = \frac{300}{4} = 75$$

$\therefore \boxed{Q_1 = 53}$ $\therefore \boxed{Q_3 = 77}$

$$Q_1 = l_1 + \left[\frac{\left(r\left(\frac{N}{4}\right) - c\right)}{f}\right](l_2 - l_1)$$

$$= 20 + \left[\frac{\left(\frac{100}{4}\right) - 24}{29}\right](30 - 20)$$

$$= 20 + \frac{[25 - 24]\ 10}{29}$$

$$= 20 + \frac{10}{29}$$

$$Q_1 = 20 + 0.34$$

$$\boxed{Q_1 = 20.34}$$

$$Q_3 = 30 + \left[\frac{3(25) - 53}{24}\right](40 - 30)$$

$$= 30 + \left(\frac{22}{24}\right) 10$$

$$= 30 + 9.17$$

$$= 39.17$$

Interquartile Range $= Q_3 - Q_1$

$= 39.17 - 20.34$

$= 18.83$

Quartile Deviation $= \frac{Q_3 - Q_1}{\frac{Q_3 + Q_1}{2}}$

$= \frac{18.83}{2}$

$\boxed{Q.D = 9.42}$ //

Coeff. of $Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{18.83}{59.51}$

$= 0.32$ //

Find interquartile range, Quartile deviation and coefficient for the following:

Class Interval	10-15	15-20	20-25	25-30	30-40	40-50
Frequency	4	12	16	22	10	8
Cummulat Frequency	4	16	32	54	64	72

50-60	60-70
6	4
78	82

$Q_1 = \frac{82}{4}$

$= 20.50$

$\therefore \quad Q_1 = 32$

$Q_3 = 3 \times 20.50$

$Q_3 = 61.50$

$\therefore \quad Q_3 = 64$

$$Q_1 = 20 + \left[\frac{20.50 - 16}{16}\right] 5$$

$$= 20 + 0.28 \times 5$$

$$= 20 + 1.40$$

$$\boxed{Q_1 = 21.40}$$

$$Q_3 = 30 + \left[\frac{61.50 - 54}{10}\right] \times 10$$

$$= 30 + 7.50$$

$$\boxed{Q_3 = 37.50}$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$

$$= \frac{37.50 - 21.40}{2}$$

$$= 8.05 //$$

Interquartile Range $= Q_3 - Q_1$

$= 37.50 - 21.40$

$= 16.10$ //

Coeff of Q.D $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \dfrac{16.10}{58.90}$

$= 0.27$ //

Calculate the quartile deviation of the marks of 39 students in Statistics given:

Marks	0-5	5-10	10-15	15-20	20-25	25-
No. of students	4	6	8	12	7	2
Cummulative frequency	4	10	18	30	37	39

$$Q_1 = \frac{39}{4}$$

$$= 9.75$$

$Q_1 = 10$

$$Q_3 = 3 \times 9.75$$

$$= 29.25$$

$Q_3 = 30$

$$Q_1 = 5 + \left[\frac{9.75 - 4}{6}\right] \times 5$$

$$= 5 + \frac{5.75}{6} \times 5$$

$$\boxed{Q_1 = 9.79} \;//$$

$$Q_3 = 15 + \left[\frac{3 \times \frac{39}{4} - 18}{12}\right] \times 5$$

$$= 15 + \frac{29.25 - 18}{} \times 5$$

$$= 15 + \frac{11.25}{12} \times 5$$

$$= 15 + 4.69$$

$$\boxed{Q_3 = 19.69} \quad //$$

Quartile Deviation $= \frac{Q_3 - Q_1}{2}$

$$= \frac{19.69 - 9.79}{2}$$

$$\boxed{Q.D = 4.95} \quad //$$

Interquartile Range $= Q_3 - Q_1$

$$= 9.90 \quad //$$

Coeffecient of Quartile Deviation :

$$\Rightarrow \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$= \frac{9.90}{29.48}$$

$$= 0.34 \quad //$$

STANDARD DEVIATION:

It is the root mean square deviation of the values from arithemetic mean. SD → σ

The square of SD is called variance

FOR RAW DATA:

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

FOR DISCRETE DATA

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}}, \quad \bar{x} = \frac{\Sigma fx}{\Sigma f}$$

CONTINOUS DATA:

$$\sigma = \sqrt{\Sigma f \frac{(x-\bar{x})^2}{\Sigma f}}, \quad \bar{x} = \frac{\Sigma fx}{\Sigma f}$$

COMBINED STD DEVIATION:

Calculating std deviation of 2 or more series.

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1+N_2}}$$

σ_{12} - Combined Std deviation of 2 groups

σ_1 - Std deviation of 1st group

σ_2 - Std deviation of 2nd group.

$\bar{x}_{1,2}$: Comibined Arthemetic Mean of two groups

$$\bar{x}_{1,2} = \frac{N_1\bar{x} + N_2\bar{x}_2}{N_1 + N_2}$$

$\bar{x}_1$ = Arthemetic Mean of 1st group

$\bar{x}_2$ = Arthemetic Mean of 2nd group

N_1 = No. of Observation in the 1st grp

N_2 = No. of observation in 2nd group.

Find the mean and std deviation for the following data 75, 73, 70, 77, 72, 75, 76, 72, 74, 76.

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

$$= \frac{740}{10}$$

$$= 74$$

x	$d = x - \bar{x}$	$d = (x - \bar{x})$
75	1	1
73	-1	1
70	-4	16
77	3	9
72	-2	4
75	1	1
76	2	4
72	-2	4
74	0	0
76	2	4

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

$$= \sqrt{\frac{44}{10}}$$

$$= \sqrt{4.40}$$

$$S.D = 2.10$$

Find mean and std deviation of

Class	0-10	10-20	20-30	30-40	40-50
Frequency	15	18	14	40	13

Class	x	f	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
0-10	5	15	75	475.24	7128.60
10-20	15	18	270	139.24	2506.32
20-30	25	14	350	3.24	45.36
30-40	35	40	1400	67.24	2689.60
40-50	45	13	585	331.24	4306.12

$$\bar{x} = \frac{\Sigma f(x)}{\Sigma f} = \frac{2680}{100}$$

$$\bar{x} = 26.80$$

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

$$= \sqrt{\frac{16676}{100}}$$

$$= \sqrt{166.76}$$

$$\boxed{\sigma = 12.91}$$

Series A	53	54	32	30	60	46	28	35
Series B	55	41	48	49	27	25	23	25
48	42							
28	60							

$$\bar{x}_1 = \frac{\Sigma x_1}{N_1} = \frac{448}{10}$$

$$\boxed{\bar{x}_1 = 44.80}$$

$$\bar{x}_2 = \frac{376}{10}$$

$$\boxed{\bar{x}_2 = 37.60}$$

$$\bar{x}_{1,2} = \frac{N_1 \bar{x}_1 + N_2 \bar{x}_2}{N_1 + N_2}$$

$$= \frac{448 + 376}{20}$$

$$\boxed{\bar{x}_{1,2} = 41.20}$$

x	$d_1^2 = (x - \bar{x})^2$	x	$d_2^2 = (x - \bar{x})^2$
53	67.24	55	302.76
54	84.64	41	11.56
32	163.84	48	108.16
30	219.04	49	129.96
60	231.04	27	112.36
46	1.44	25	158.76
28	282.24	23	213.16
25	392.04	20	309.76
48	10.24	28	92.16
72	739.84	60	501.76

$$\sigma_1 = \sqrt{\frac{2191.6}{10}} \qquad \sigma_2 = \sqrt{\frac{1940.40}{10}}$$

$$\sigma_1 = 219.16 \qquad \sigma_2 = 194.04$$

$$d_1^2 = (\bar{x}_{10} - \bar{x}_{12})^2 = (144.80 - 141.20)^2$$

$$= 12.96$$

$$d_2^2 = (\bar{x}_2 - \bar{x}_{12})^2 = (37.60 - 41.2)^2$$

$$= 12.96.$$

$$\sigma = \sqrt{\frac{10 \times 219.16 + 10 \times 194.04 + 10 \times 12.96 + 10 \times 12.96}{20}}$$

$$= \sqrt{\frac{4391.20}{20}}$$

$$\sigma_{1,2} = \sqrt{219.560}$$

Coefficient Of Variation :

$$\text{Coff of variation} = \frac{S.D}{\text{mean}} \times 100$$

The measures of dispersion are expressed in the same units as orginal observation and are called absolute variability. These values cannot be used to compare two / more distribution given in different units.

In such cases the relative measures of variability can be used which are independent of units of measurement.

Find the coefficient of variation for the following data.

Weight	Frequency	x	fx	d	$f(x-\bar{x})^2$
110-119	5	114.5	572.25	1122.25	5611.25
120-129	7	124.5	871.5	552.25	3865.75
130-139	12	134.5	1614.00	182.25	2187
140-149	20	144.5	2890	12.25	245
150-159	16	154.5	2472	42.25	676
160-169	10	164.5	1645	272.25	2722.5
170-179	7	174.5	1221.5	702.25	4915.75
180-189	3	184.5	553.5	1332.25	3996.75

$$\text{Mean} = \bar{x} = \frac{\Sigma f(x)}{\Sigma f}$$

$$= \frac{11840}{80}$$

$$\boxed{\bar{x} = 148}$$

$$\sigma = \sqrt{\frac{64220}{80}}$$

$$= \sqrt{802.75}$$

$$= 17.4$$

$$\text{Coeff of variation} = \frac{17.4}{148} \times 100$$

$$= 11.75 //$$

From the data given bllow which series is more variable (use S.D).

Variable	f Series A	f Series B	x	fx_A	fx_B
10 - 20	10	18	5	150	270
20 - 30	18	22	25	450	550
30 - 40	32	40	35	1120	1400
40 - 50	40	32	45	1800	1440
50 - 60	22	18	55	1210	990
60 - 70	18	10	65	1170	650
70 - 80					

$$\bar{x}_A = \frac{\Sigma f x_A}{\Sigma f_A} \qquad \bar{x}_B = \frac{\Sigma f_B}{\Sigma f_B}$$

$$= \frac{5900}{140} \qquad = \frac{5300}{140}$$

$$\bar{x}_A = 42.14 \qquad \bar{x}_B = 37.85$$

$d_A = (x-\bar{x})^2$	$d_B = (x-\bar{x})^2$	$f(x-\bar{x})^2$	$f(x-\bar{x})^2$
736.57	522.12	7365.7	9398.16
293.78	165.12	5288.04	3632.64
50.98	8.12	1681.04	324.8
8.18	51.12	327.2	1635.84
165.38	594.12	9638.36	5294.16
522.58	737.12	9406.44	7371.2

$$\sigma_A = \sqrt{\frac{27656.78}{140}}$$

$$= \sqrt{197.54}$$

$$\boxed{\sigma_A = 140.5}$$

$$\sigma_B = \sqrt{\frac{27656.8}{140}}$$

$$= \sqrt{197.54}$$

$$\boxed{\sigma_B = 14.0}$$

Coff of variation $= \frac{14.0}{37.85} \times 100$

$= 37$ or 36.9 (approx).

Calculate appropriate measure of dispersion.

Wages in Rs per week	No. of wage earners
less than 35	14
35 - 37	62
38 - 40	99
41 - 43	18
over 43.	7.

Sol:

Wages per week	x	f	fx	$x-\bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
32.5	33.5	14	462	-5.13	26.32	368.48
34.5-37.5	36	162	2232	-2.13	4.54	281.48
37.5-40.5	39	99	3861	0.87	0.76	75.24
40.5-43.5	42	18	756	3.87	14.98	269.64
43.5-46.5	45	7	313	6.87	47.22	330.54
						1325.38

$$\bar{x} = \frac{\sum fx}{\sum x} = \frac{7626}{200} = 36.13.$$

$$\sigma = \sqrt{\frac{1325.38}{200}} = 2.58 \text{ rupees}$$

Prob 2:

Find std deviation.

Height in inches	No. of students
Over 60 but not more than 62	35
" 62 " " 64	27
" 64 " " 66	20
" 66 " " 68	13
" 68 " " 70	5
	100

Sol:

Class Interval	frequency	Mid (x)	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
60 - 62	35	61	2135	6.3504	222.264
62 - 64	27	63	1701	0.2704	7.3008
64 - 66	20	65	1300	2.1904	43.80
66 - 68	13	67	871	12.1104	157.435
68 - 70	5	69	345	30.0304	150.15

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{6852}{100} = 68.5$$

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}} = \sqrt{\frac{580.96}{100}} = 2.4103$$

BIVARIATE DATA:

Sammarization Of Data:

1. TYPES OF DATA

Quantitative
- Discreate
- continous

QUALITATIVE
- Nominal
- Ordiral

2. DATA COLLECTION METHOD.

- Survey & Questionaires
- Experiments
- Observational Studies
- Interviews
- Soundary data Source.

3. DATA SUMMARIZATION TECHNIQUES:

Tabular

Graphical

Numerical

Mean, Median, Mode

Range, Varience, Std deviation

4. DATA DISTRIBUTION

5. BIVARIATE AND MULTIVARIATE

- Correlation
Covariance
- Scatterplots

Principal Component (PCA)

Multipl Regression Models

6. DATA CLEANING & PROCESSING

Handle Missing Values

7. DESCRIPTIVE VS Infrartial Statistic.

Summarize Data Conclusion

Present Data

Bivariate data analyze defined as the relation b/t two variables often represented graphically as a scatter plot. It helps to determine the correlation b/t two variables.

SCATTER PLOT

It is commonly used to visualize bivariate data represents relation b/t the data variables on x axis & y axis.

FREQUENCY DISTRIBUTION

The No. of times the value of x repeated in a distribution.

1. The No. of calls from motorists per day for road side service was recorded from the month of Dec 2024.

The result were as follows

28, 122, 217, 130, 120, 86, 80, 90, 120, 140, 70, 40, 145, 187, 1813, 90, 68, 174, 194, 170, 100, 75, 104, 97, 75, 123, 100, 82, 109, 120, 801.

Find the frequency distribution.

Smallest value -> 28

Largest value -> 217

Diff -> 189

Class = $\frac{189}{5}$ = 37.8

Let us fix class width as 40.

	Frequency
0 - 40	9
41 - 80	5
81 - 120	14
121 - 160	5
161 - 200	4
201 - 240	1

---

2. Consider the dataset of blood groups of 20 blood donate in blood donation camp.

O, A, A, A, AB, B, O, AB, B, O, AB, A, B, B, O, B, A, A, AB, A

Constract a frequency distribution for the above data set.

Blood	Frequency
O	4
A	7
B	5
AB	4

3. Find relative frequency and percentage frequency of a class interval table.

Mark Obtained	f No. of students	Relative Freq	% Freq
20-40	6	6/25 = 0.24	0.24 × 100 ≈24
40-60	12	0.48	48
60-80	4	0.160	16
80-100	3	0.12	12
	≈25		

LINEAR REGRESSION MODEL:

Regression Analysis is a measure to analyse the corelation (relation b/t 2/more more variables) in terms of the orginal units of the data.

Regression lines are the lines of best fit as a straight line.

A linear regression model is a statical model that examines relationship b/t 2/more variables by fitting a linear equation to the observed data

$$y = ax + b$$

y → dependent variable

x → independent variable

a, b → correlation variable.

1. Construct a scatter plot for the data shown

Company	Cars	Revenew
A	63.0	7
B	29.0	8.9
C	20.8	2.1
D	19.1	2.8
E	13.4	1.4
F	8.5	1.5