Unit 3 :- Stastistical Methods

Book

1) by P.N. Arora
2) by S.G. Gupta

1) Measure of central tendency
2) Measure of dispersion
3) Correlation coefficient
4) Regression lines
5) Curve fitting
6) Moments, skewness, kurtosis

1) Measure of central tendency :-

A single expression, representing the whole group, is selected which may convey a fairly adequate idea about the group This single expression in statistics is known as average.

1. Arithmetic Mean
2. Median
3. Mode
5. Geometric Mean
6. Harmonic Mean.

Arithmetic Mean (A.M)

$x_1, x_2, \cdots x_n$ be n numbers then average or A.M of $\{x_i's\}$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$x_1, x_2, \ldots x_n$ occuring with frequencies $f_1, f_2, \ldots f_n$ resp.

0	f_1
1	f_2
2	⋮
⋮	
10	f_n

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{f_1 + f_2 + \cdots + f_n}$$

$$= \frac{\sum_{i=1}^{n} f_i x_i}{N}, \quad N = \sum f$$

Grouped data/ Class interval

Class interval	f_i	(Mid value of class interval x_i)
0 – 10	f_1	5
10 – 20	f_2	15
20 – 30	f_3	⋮
⋮	⋮	⋮
90 – 100	f_n	⋮

$$\bar{x} = \frac{\sum f_i x_i}{N}, \quad x_i = \text{Mid value of } i^{th} \text{ class interval.}$$

Ex 1. The following distribution gives the number of accidents met by 160 workers in a factory during a month

No. of accidents (x)	0	1	2	3	4
No. of workers (f)	70	52	34	3	1

Find the average number of accidents per worker.

→ $$\bar{x} = \frac{x_1f_1 + x_2f_2 + x_3f_3 + x_4f_4 + x_5f_5}{f_1 + f_2 + f_3 + f_4 + f_5}$$

$$= \frac{(0)(70) + (1)(52) + 2(34) + 3(3) + 4(1)}{70 + 52 + 34 + 3 + 1} = \frac{133}{160} = 0.83$$

Measure of dispersion :-

The meaning of dispersion is scatteredness. It helps out the variability of the data or scatteredness of individual items in a given distribution.

1. Mean deviation
2. Standard deviation

1. **Mean deviation** : It is the average of the modulus of the deviations of the observations in a series taken from mean or median or mode.

$$\text{Mean deviation} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

eg:- Calculate the mean deviation about the mean for the following series 15, 20, 17, 19, 21, 13, 12, 10, 17, 9, 12

→ $\bar{x} = \frac{15+20+17+19+21+13+12+10+17+9+12}{11} = 15$

Mean deviation =

x	$x-\bar{x}$	$\lvert x-\bar{x}\rvert$
15	0	0
20	5	5
17	2	2
19	4	4
21	6	6
13	-2	2
12	-3	3
10	-5	5
17	2	2
9	-6	6
12	-3	3
		38

$\therefore$ Mean deviation $= \frac{38}{11}$

$= 3.455$

$x_1, x_2, - - - x_n$ with frequencies $f_1, f_2, - - - f_n$

$$\text{Mean deviation} = \frac{\sum f_i |x_i - \bar{x}|}{\sum f_i} = \frac{\sum f |x - \bar{x}|}{N}$$

eg:- Find the mean deviation for the following data

Marks	20	18	16	14	12	10	8	6
# of students	2	4	9	18	27	25	14	1

Marks (x)	No. of students (f)	xf	$\lvert d\rvert = \lvert x - 12\rvert$	$f \times \lvert d\rvert$
6	1	6	6	6
8	14	112	4	56
10	25	250	2	50
12	27	324	0	0
14	18	252	2	36
16	9	144	4	36
18	4	72	6	24
20	2	40	8	16
	100	1200		224

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{1200}{100} = 12 \quad , \quad \text{A.M} = \frac{\sum f |x - \bar{x}|}{\sum f} = \frac{224}{100} = 2.24.$$

Standard deviation (σ) :- It measure the absolute dispersion of a distribution.

Std. deviation is a positive square root of the average of the squared deviations taken from the arithmetic mean.

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

In case of freq. distribution,

x	x_1 x_2 - - - x_n
f	f_1 f_2 - - - f_n

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}} \quad \text{OR} \quad \sqrt{\frac{\Sigma fx^2}{N} - \left(\frac{\Sigma fx}{N}\right)^2}$$

where $N = f_1 + f_2 + - + f_n$.

Ex:- Find the standard deviation from the following data:

Size of the item	10	11	12	13	14	15	16
Frequency	2	7	11	15	10	4	1

Soln:-

x	f	xf	x^2	fx^2
10	2	20	100	200
11	7	77	121	847
12	11	132	144	1584
13	15	195	169	2535
14	10	140	196	1960
15	4	60	225	900
16	1	16	256	256
	50	640		8282

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N} - \left(\frac{\Sigma fx}{N}\right)^2} = \sqrt{\frac{8282}{50} - \left(\frac{640}{50}\right)^2}$$

$$= \sqrt{165.64 - 163.84} = 1.3416$$

To reduce calculation, we define

$$u = x - A, \quad \text{OR} \quad u = \frac{x-A}{h}$$

$u = x - A$

$$\sigma_x = \sigma_u = \sqrt{\frac{\Sigma fu^2}{N} - \left(\frac{\Sigma fu}{N}\right)^2} \quad \text{OR}$$

$u = \frac{x-A}{h}$

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

x	f	$u=x-13$	fu	u^2	fu^2
10	2	−3	−6	9	18
11	7	−2	−14	4	28
12	11	−1	−11	1	11
13	15	0	0	0	0
14	10	1	10	1	10
15	4	2	8	4	16
16	1	3	3	9	9
$N=$	50		−10		92

$$\sigma = \sqrt{\frac{\Sigma fu^2}{N} - \left(\frac{\Sigma fu}{N}\right)^2} = \sqrt{\frac{92}{50} - \left(\frac{-10}{50}\right)^2} = 1.342$$

Ex:- Find the std. deviation for the following data

C.I	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	6	14	10	8	1	3	8

Soln:-

Class interval	f	Mid value(x)	$u=\frac{x-35}{10}$	u^2	fu	fu^2
0-10	6	5	−3	9	−18	54
10-20	14	15	−2	4	−28	56
20-30	10	25	−1	1	−10	10
30-40	8	35	0	0	0	0
40-50	1	45	1	1	1	1
50-60	3	55	2	4	6	12
60-70	8	65	3	9	24	72
	50				−25	205

$$S.D = h\sqrt{\frac{\Sigma fu^2}{N} - \left(\frac{\Sigma fu}{N}\right)^2} = 10\sqrt{\frac{205}{50} - \left(\frac{-25}{50}\right)^2}$$

$$S.D = 19.62$$

Coefficient of variation [C.V] $= \frac{\sigma}{\bar{x}} \times 100$

Ex:- The following are the runs scored by two batsman A & B in ten innings.

(x) A:	101	27	0	36	82	45	7	13	65	14
(y) B	97	12	40	96	13	8	85	8	56	15

Who is more consistent?

→

C.V for $x = \frac{\sigma_x}{\bar{x}} \times 100$

C.V for $y = \frac{\sigma_y}{\bar{y}} \times 100$

x	$u = x - 39$	u^2
101	62	3844
27	-12	144
0	-39	1521
36	3	9
82	43	1849
45	6	36
7	-32	1024
13	-26	676
65	26	676
14	-25	625
	0	10404

$$\sigma_x = \sigma_u = \sqrt{\frac{\Sigma u^2}{10} - \left(\frac{\Sigma u}{10}\right)^2}$$

$$= 32.26$$

$\therefore \bar{x} = 39$

$\therefore$ C.V for $x = \frac{\sigma_x}{\bar{x}} \times 100 = 82.72\%$

Similarly $\sigma_y = \sigma_v = 35.72 = \sqrt{\frac{\Sigma v^2}{10} - \left(\frac{\Sigma v}{10}\right)^2}$

y	$v = y - 43$	v^2

C.V for $y = \frac{\sigma_y}{\bar{y}} \times 100 = 83.07\%$

C.V of A < C.V of B $\Rightarrow$ Batsman A is more consistent than batsman B.