STATISTICS

1. ARITHMETIC MEAN

Individual observation or unclassified data :

If x_1, x_2,.....x_n be n observations, then their arithmetic mean is given by

$$\bar{x} = \frac{x_1 + x_2 + \ldots\ldots + x_n}{n} \quad \text{or} \quad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Arithmetic mean of discrete frequency distribution :

Let x_1, x_2,.......,x_n be n observation and let f_1, f_2, ..., f_n be their corresponding frequencies, then their mean

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots\ldots + f_nx_n}{f_1 + f_2 + \ldots\ldots + f_n} \quad \text{or} \quad \bar{x} = \frac{\sum_{i=1}^{n} f_i\ x_i}{\sum_{i=1}^{n} f_i}$$

2. MEDIAN

Median is the middle most or the central value of the variate in a set of observations, when the observations are arranged either in ascending or in descending order of their magnitudes. It divides the arranged series in two equal parts.

Median of an individual series :

Let n be the number of observations-

(i) arrange the data in ascending or descending order.

(ii) **(a) if n is odd then-**

Median (M) = Value of $\left(\frac{n+1}{2}\right)^{th}$ observation

(b) If n is even then

Median (M) = mean of $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2}+1\right)^{th}$ observation

$$\text{i.e. } M = \frac{\left(\frac{n}{2}\right)^{th} \text{observation} + \left(\frac{n}{2}+1\right)^{th} \text{observation}}{2}$$

Median of the discrete frequency distribution:

Algorithm to find the median :

Step - I : Find the cumulative frequency (C. F.)

Step- II : Find $\frac{N}{2}$, where $N = \sum_{i=1}^{n} f_i$

Step -III : See the cumulative frequency (C.F.) just greater than $\frac{N}{2}$ and determine the corresponding value of the variable.

Step-IV : The value obtained in step III is the median.

Median of grouped data or continuous series:

Let the no. of observations be n

(i) Prepare the cumulative frequency table

(ii) Find the median class i.e. the class in which the $\left(\frac{N}{2}\right)^{th}$ observation lies

(iii) The median value is given by the formulae

$$\text{Median (M)} = \ell + \left[\frac{\left(\frac{N}{2}\right) - F}{f}\right] \times h$$

N = total frequency = $\sum f_i$

l = lower limit of median class

f = frequency of the median class

F = cumulative frequency of the class preceding the median class

h = class interval (width) of the median class

3. MEASURES OF DISPERSION

Dispersion is the measure of the variations. The degree to which numerical data tend to spread about an average value is called the dispersion of the data.

The measures of dispersion commonly used are:

(i) Range

(ii) Quartile deviation or the semi- interquartile range

(iii) Mean Deviation

(iv) Standard Deviation

Here we will discuss the mean deviation and standard deviation.

Mean Deviation : Mean deviation is defined as the arithmetic mean of the absolute deviations of all the values taken about any central value.

(i) Mean deviation of individual observations:

If $x_1, x_2, \ldots\ldots, x_n$ are n values of a variable x, then the mean deviation from an average A (median or AM) is given by

$$\text{M.D.} = \frac{1}{n}\sum_{i=1}^{n}|x_i - A|$$

$$= \frac{1}{n}\sum|d_i|, \text{ where } d_i = x_i - A$$

(ii) Mean deviation of a discrete frequency distribution :

If $x_1, x_2, \ldots\ldots, x_n$ are n observation with frequencies $f_1, f_2, \ldots\ldots, f_n$, then mean deviation from an average A is given by -

$$\text{Mean Deviation} = \frac{1}{N}\sum f_i|x_i - A|$$

$$\text{where } N = \sum_{i=1}^{n} f_i$$

(iii) Mean deviation of a grouped or continuous frequency distribution:

For calculating mean deviation of a continuous frequency distribution the procedure is same as for a discrete frequency distribution. The only difference is that here we have to obtain the mid-point of the various classes and take the deviations of these mid points from the given central value (median or mean)

4. VARIANCE AND STANDARD DEVIATION

The variance of a variate x is the arithmetic mean of the squares of all deviations of x from the arithmetic mean of the observations and is denoted by var (x) or σ^2

The positive square root of the variance of a variate x is known as standard deviation i.e. standard deviation = $+\sqrt{\text{var}(x)}$

(i) Variance of Individual observations :

If $x_1, x_2, \ldots\ldots, x_n$ are n values of a variable x, then by definition

$$\text{var}(x) = \frac{1}{n}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right] = \sigma^2 \quad \ldots(i)$$

$$\text{or var}(x) = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2 \quad \ldots(ii)$$

If the values of variable x are large, the calculation of variance from the above formulae is quite tedious and time consuming. In that case, we take deviation from an arbitrary point A (say) then

$$\text{var}(x) = \frac{1}{n}\sum_{i=1}^{n}d_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}d_i\right)^2 \quad \ldots(iii)$$

(ii) Variance of a discrete frequency distribution:

If $x_1, x_2, \ldots\ldots, x_n$ are n observations with frequencies $f_1, f_2, \ldots\ldots, f_n$ then var (x) =

$$\frac{1}{N}\left\{\sum_{i=1}^{n}f_i(x_i - \bar{x})^2\right\} \quad \ldots(i)$$

$$\text{or var}(x) = \frac{1}{N}\sum_{i=1}^{n}f_i x_i^2 - \bar{x}^2 \quad \ldots(ii)$$

If the values of x or f are large, we take the deviations of the values of variable x from an arbitrary point A. (say)

$\therefore d_i = x_i - A$; i = 1, 2, ...n

$$\therefore \text{Var}(x) = \frac{1}{N}\left(\sum_{i=1}^{n}f_i d_i^2\right) - \left(\frac{1}{N}\sum_{i=1}^{n}f_i d_i\right)^2 \ldots(iii)$$

$$\text{where } N = \sum_{i=1}^{n}f_i$$

Sometime $d_i = x_i - A$ are divisible by a common number h (say)

then

$$u_i = \frac{x_i - A}{h} = \frac{d_i}{h}, \; i = 1, 2, \ldots\ldots, n$$

then

$$\text{var}(x) = h^2\left[\frac{1}{N}\sum_{i=1}^{n}f_i u_i^2 - \left(\frac{1}{N}\sum_{i=1}^{n}f_i u_i\right)^2\right] \quad \ldots(iv)$$

(iii) Variance of a grouped or continuous frequency distribution :

In a grouped or continuous frequency distribution any of the formulae discussed in discrete frequency distribution can be used.

SOLVED PROBLEMS

Ex.1 Mean of 25 observations was found to be 78.4. But later on it was found that 96 was misread 69. Find the correct mean

Sol. Mean $\bar{x} = \frac{\sum x}{n}$ or $\sum x = n\bar{x}$

$\sum x = 25 \times 78.4 = 1960$

But this $\sum x$ is incorrect as 96 was misread as 69.

$\therefore$ correct $\sum x = 1960 + (96 - 69) = 1987$

$\therefore$ correct mean $= \frac{1987}{25} = 79.48$

Ex.2 Find mean deviation about mean from the following data :

x_i :	3	9	17	23	27
f_i :	8	10	12	9	5

Sol. Calculation of mean deviation about mean.

x_i	f_i	f_ix_i	$\lvert x_i - 15\rvert$	$f_i \lvert x_i - 15\rvert$
3	8	24	12	96
9	10	90	6	60
17	12	204	2	24
23	9	207	8	72
27	5	135	12	60

$N = \sum f_i = 44$ $\quad \sum f_i x_i = 660$ $\quad \sum f_i \lvert x_i - 15 \rvert = 312$

Mean $= \bar{X} = \frac{1}{N}(\sum f_i x_i) = \frac{660}{44} = 15$

Mean deviation = M.D. $= \frac{1}{N}\sum f_i \lvert x_i - 15 \rvert$

$= \frac{312}{44} = 7.09.$

Ex.3 Find the variance of the data given below

size of item:	*3.5*	*4.5*	*5.5*	*6.5*	*7.5*	*8.5*	*9.5*
frequency	3	7	22	60	85	32	8

Sol. Let the assumed mean be A = 6.5

Calculation of variance

size of item x_i	**f_i**	**$d_i = x_i - 6.5$**	**d_i^2**	**f_id_i**	**$f_id_i^2$**
3.5	3	−3	9	− 9	27
4.5	7	−2	4	−14	28
5.5	22	−1	1	−22	22
6.5	60	0	0	0	0
7.5	85	1	1	85	85
8.5	32	2	4	64	128
9.5	8	3	9	24	72

$N = \sum f_i = 217$

$\sum f_i d_i = 128$ $\qquad \sum f_i d_i^2 = 362$

Here, $N = 217$, $\sum f_i d_i = 128$ and $\sum f_i d_i^2 = 362$

$\therefore$ Var (X) $= \left(\frac{1}{N}\sum f_i d_i^2\right) - \left(\frac{1}{N}\sum f_i d_i\right)^2$

$= \frac{362}{217} - \left(\frac{128}{217}\right)^2 = 1.668 - 0.347 = 1.321$

Ex.4 Find the mean deviation about median from the following data : 340, 150, 210, 240, 300, 310, 320,

Sol. Arranging the observations in ascending order of magnitude, we have 150, 210, 240, 300, 310, 320, 340. Clearly, the middle observation is 300. So, median = 300

Calculation of Mean deviation

x_i	$\|d_i\| = \|x_i - 300\|$
340	40
150	150
210	90
240	60
300	0
310	10
320	20
Total	$\sum\|d_i\| = \sum\|x_i - 300\| = 370$

Mean deviation

$$= \frac{1}{n}\sum|d_i| = \frac{1}{7}\sum|x_i - 300|$$

$$= \frac{370}{7} = 52.8$$

Ex.5 Marks of 5 students of a tutorial group are 8, 12, 13, 15, 22 then find the variance

Sol. $\bar{x} = \frac{8+12+13+15+22}{5} = 14$

CALCULATION OF VARIANCE

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
8	− 6	36
12	− 2	4
13	− 1	1
15	1	1
22	8	64
		$\sum(x_i - \bar{x})^2 = 106$

$\because$ n = 5, $\sum(x_i - \bar{x})^2 = 106$

$\therefore$ var (x) = $\frac{1}{n}\sum(x_i - \bar{x})^2 = \frac{106}{5} = 21.2$

Ex.6 Calculate the mean and standard deviation for the following data :

Class Interval	20–30	30–40	40–50	50–60	60–70	70–80	80–90
f	3	51	122	141	130	51	2

Sol. Take A = 55 and h = 10

Class Interval	Mid-value (x)	f	$d = \frac{x-55}{h=10}$	fd	fd^2
20–30	25	3	−3	−9	27
30–40	35	51	−2	−102	204
40–50	45	122	−1	−122	122
50–60	55	141	0	0	0
60–70	65	130	1	130	130
70–80	75	51	2	102	204
80–90	85	2	3	6	18

$$\bar{x} = A + \left(\frac{\sum fd}{N}\right) \times h = 55 + \frac{5}{500} \times 10$$

$$= 55 + .1 = 55.1$$

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

$$= \left[\frac{500 \times 705 - (5)^2}{(550)^2}\right] = \frac{14099}{100} = 140.99$$

So, $\sigma = \sqrt{140.99} = 11.87$

Ex.7 Given the $\bar{x}$ is the mean and σ^2 is the variance of n observation $x_1, x_2, x_3, \ldots,$ x_n. Prove that the mean and variance of observations $ax_1, ax_2, ax_3, \ldots, ax_n$ are $a\bar{x}$ and $a^2\sigma^2$ respectively $(a \neq 0)$.

Sol. $\bar{x} = \frac{\Sigma x}{n} = \frac{x_1 + x_2 + x_3 + \ldots x_n}{n}$ .....(1)

Also $\frac{\Sigma x^2}{n} = \frac{x_1^2 + x_2^2 + x_3^2 + \ldots x_n^2}{n}$ .....(2)

For the new data y = $ax_1, ax_2, ax_3, \ldots, ax_n$.

$\Rightarrow$ $\Sigma y = ax_1 + ax_2 + ax_3 + \ldots\ ax_n$.

$\therefore$ Mean, $\bar{y} = \frac{ax_1 + ax_2 + ax_3 + \ldots + ax_n}{n}$

$$= \frac{a(x_1 + x_2 + x_3 + ... + x_n)}{n} = a\bar{x}$$

[By using (1)]

Hence, the new mean, $\bar{y} = a\bar{x}$.

Variance of the first data, σ^2

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

$$= \frac{n(x_1^2 + x_2^2 + x_2^3 + ... + x_n^2) - (x_1 + x_2 + x_3 + ... + x_n)^2}{n^2}$$

.....(3)

Variance of the new data, $\sigma_1^2 = \frac{n\Sigma y^2 - (\Sigma y)^2}{n^2}$

$$= \frac{n(a^2x_1^2 + a^2x_2^2 + a^2x_3^2 + ... + a^2x_n^2)}{n^2}$$

$$= \frac{-(ax_1 - ax_2 + ax_3 + ... + ax_n)^2}{n^2}$$

$$= a^2\left[\frac{n(x_1^2 + x_2^2 + x_2^3 + ... + x_n^2) - (x_1 + x_2 + x_3 + ... + x_n)^2}{n^2}\right]$$

$= a^2\sigma^2$ [By using (3)]

Hence, the new variance $\sigma_1^2 = a^2\sigma^2$

8. The following are the runs scored by the two batsman A and B in ten innings.

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

Who is more consistent bastman ?

Sol. Calculation of coefficient of variation

Run scored x	Batsman A $dx = x - \bar{x}$ $= x - 39$	dx^2	Runs scored y	Batsman B $dy = y - \bar{y}$ $= y - 43$	dy^2
101	62	3844	97	54	2916
27	−12	144	12	−31	961
0	−39	1521	40	−3	9
36	−3	9	96	53	2809
82	43	1849	13	−30	1225
45	6	36	8	−35	1225
7	−32	1024	85	42	1764
13	−26	676	8	−35	1225
65	26	676	56	13	169
14	−25	625	15	−28	784
$\Sigma x = 390$	10404	$\Sigma y = 430$			12762

Batsman A

$$\bar{x} = \frac{\Sigma x}{n} = \frac{390}{10} = 39$$

$$\sigma = \sqrt{\frac{\Sigma dx^2}{n}} = \sqrt{\frac{10404}{10}} = \sqrt{1040.4} = 32.22$$

$$C.V. = \frac{S.D.}{\bar{x}} \times 100 = \frac{32.22}{39} \times 100 = 82.62$$

Batsman B

$$\bar{y} = \frac{\Sigma y}{n} = \frac{430}{10} = 43$$

$$\sigma = \sqrt{\frac{\Sigma dy^2}{n}} = \sqrt{\frac{12762}{10}} = \sqrt{1276.2} = 35.72$$

$$C.V. = \frac{S.D.}{\bar{y}} \times 100 = \frac{35.72}{43} \times 100 = 83.07$$

EXERCISE - I UNSOLVED PROBLEMS

Q.1 Find the mean deviation about the mean for the following data: 7, 8, 4 13, 9, 5, 16, 18

Q.2 Find the mean deviation about the median for the following data : 12, 5, 14, 6, 11, 13, 17, 8, 10

Q.3 Find the mean deviation about the mean for the following data :

x_i	6	12	18	24	30	36
f_i	5	4	11	6	4	6

Q.4 Find the mean deviation about the median for the following data :

x_i	5	7	9	11	13	15	17
f_i	2	4	6	8	10	12	8

Q.5 Find the mean deviation about the mean for the following data :

Mark	0-10	10-20	20-30	30-40	40-50	50-60
No.	6	8	14	16	4	2

Q.6 Find the mean deviation about the median for the following data:

Class	0-10	10-20	20-30	30-40	40-50	50-60
Freq.	6	7	15	16	4	2

Q.7 Find the mean, variance and standard deviation for the numbers 4, 6, 10, 12, 7, 8, 13, 12.

Q.8 Find the mean, variance and standard deviation for first six odd natural numbers.

Q.9 Using cut method, find the mean, variance and standard deviation for the data:

x_i	4	8	11	17	20	24	32
f_i	3	5	9	5	4	3	1

Q.10 If the standard deviation of the numbers 2, 3, 2x, 11, is 3.5, calculate the possible values of x.

Q.11 The variance of 15 observations is 6. If each observation is increased by 8, find the variance of the resulting observations.

Q.12 The variance of 20 observations is 5. If each observation is multiplied by 2, find the variance of the resulting observations.

Q.13 The mean and variance of five observations are 6 and 4 respectively. If three of these are 5, 7 and 9, find the other two observations.

Q.14 The mean and variance of five observation are 4.4 and 8.24 respectively. If three of these are 1, 2, and 6, find the other two observations.

Q.15 The mean and standard deviation of 18 observations are found to be 7 and 4 respectively. On rechecking it was found that an observation 12 was misread as 21. Calculate the correct mean and standard deviation.

Q.16 For a group of 200 candidates, the mean and standard deviations of scores were found to be 40 and 15 respectively. Later on it was discovered that the score of 43 was misread as 34. Find the correct mean and standard deviation.

Q.17 The mean and standard deviations of a group of 100 observations were found to be 20 and 3 respectively. Later on it was found that three observations 21, 12 and 18 were incorrect. Find the mean and standard deviation if the incorrect observations were omitted.

Q.18 The following results show the number of workers and the wages paid to them in two factories A and B.

Factory	**A**	**B**
Number of workers	3600	3200
Mean wages	Rs 5300	Rs 5300
Variance	100	81

Which factory has more variation in wages ?

Q.19 Coefficient of variation of two distribution are 60 % and 80 % respectively, and their standard deviation are 21 and 16 respectively. Find their arithmetic means.

Q.20 The mean and variance of the height and weights of the students of a class are give below:

	Heights	Weights
Mean	63.2 inches	63.2 kg
SD	11.5 inches	5.6 kg

Which shows more variability, heights or weights?

Q.21 The following results show the number of workers and the wages paid to them in two factories A and B of the same industry.

Firms	**A**	**B**
Numbers of workers	560	650
Mean monthly wages	Rs 5460	Rs 5460
Variance of distribution of wages	100	121

(i) Which firm pays larger amount as monthly wages ?

(ii) Which firm shows greater variability in individual wages ?

Q.22 The sum and the sum of squares of length x (in cm) and weight y (in g) of 50 plant products are given below :

$$\sum_{i=1}^{50} x_i = 212, \quad \sum_{i=1}^{50} x_i^2 = 902.8, \quad \sum_{i=1}^{50} y_i = 261$$

and $\sum_{i=1}^{50} y_i^2 = 1457.6$

Which is more variable, the length or weight?

ANSWER KEY

1.	4.25	**2.**	3
3.	8	**4.**	2.72
5.	10.24	**6.**	10.16
7.	Mean = 9, Variance = 9.25 and SD = 3.04		
8.	Mean = 6, Variance = 11.67 and SD = 3.41		
9.	Mean = 14, Variance = 45.8 and SD = 6.77		
10.	3 and 7/3	**11.**	6
12.	20	**13.**	3 and 6
14.	4 and 9	**15.**	6.5, 2.5
16.	40.045, 14.995	**17.**	20.06, 14.4
18.	A	**19.**	35, 20
20.	Heights	**21.**	(i) A (ii) B
22.	Weight		