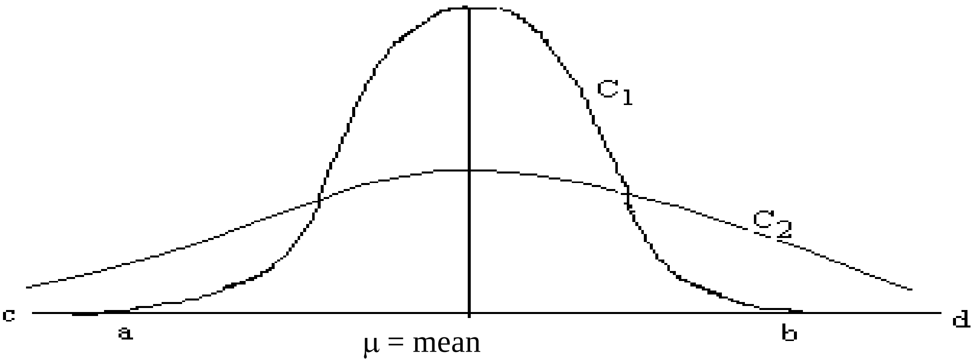
MAT2001-Statistics for Engineers

MEASURES OF VARIATION

INTRODUCTION

The measures of central tendency are descriptive statistics which attempt to summarize the data. Two sets of data may possess a common arithmetic mean, but differ in terms of their measurements. The observations of a data set have a tendency to cluster around their central value. The deviations of observations from their central value pot ray the spread of the observations. Larger deviations imply greater variation (or spread) of the measurements.

Measures of central tendency or averages describe the data, but such description is incomplete unless a measure of variation (or spread or closeness) of observations is augmented.



In the above diagram two symmetric curves are found with common mean. Each curve represents a frequency distribution. By visual comparison we notice the spread of the measurements of C_1 (fall between a and b) is less than that of C_2(fall between c and d). However, the degree of variation of measurements of a given data set can be measured, for the purpose of which we use two sets of measures, (a) measures of absolute variation and (b) measures of relative variation. A measure of absolute variation is expressed in terms of the units of measurement of the observations. For example, if units of measurement are in inches the corresponding measure of variation should also be expressed in inches.

MEASURES OF ABSOLUTE VARIATION:

The various measures of absolute variation are (i) Range (ii) Quartile Deviation (iii) Mean Deviation and (iv) Standard Deviation.

MEASURES OF RELATIVE VARIATION:

These are the measures that are free from units of measurement. A measure of relative variation is expressed in percentage or ratio. Such measures are (i) co-efficient of range (ii) co-efficient of quartile deviation (iii) co-efficient of mean deviation and (iv) co-efficient of variation.

Range:

The simplest but a crude measure of variation is range. it is defined as the difference between the highest and lowest values in the series (data).

Case A: Raw data:

Suppose a raw data set contains 'n' observations. Then the Range is the difference between the largest and smallest values in the data.

Range = Largest value – Smallest value

Case B: Discrete Frequency Distribution:

In this case, the Range is the difference between the maximum and minimum; values of the given variable.

Case C: Continuous Frequency Distribution:

In a continuous frequency distribution, Range is defined as the difference between the upper bound of the final class and the lower bound of the initial class, provided that the bounds appear in increasing magnitude.

Uses of Range:

Range is an unsatisfactory measure of dispersion. However, it is used in Industrial engineering and statistical quality control. It is extremely useful when measures of dispersion are to be calculated repeatedly, for a large number of times.

A relative measure of dispersion based on range is the coefficient of range which is useful to compare the variations in two or more series of data.

$$\text{Coefficient of range} = \left[\frac{\text{Maximum value-Minmamum value}}{\text{Maximum value+minimum value}}\right]$$

Solved problems on Range:

Problem: Find the range of the following series

80, 90, 60, 68, 100, 75, 89

Solution : Largest observation = 100

Smallest observation = 60

Range = 100 – 60 = 40

QUARTILE DEVIATION (Q.D):

It is a better measure of variation than range. It is based on the first and third quartiles, namely Q_1 and Q_3 It is defined as,

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

The computation of quartile deviation is based on the computations of Q_3 and Q_1.

USES OF QUARTILE DEVIATION:

It is superior and more reliable measure than the range as it makes use of 50% of the data. When a frequency distribution contains open end classes, Q.D. is the appropriate measure of dispersion.

Remarks: Q.D. is a measure of absolute dispersion. A measure of relative dispersion based on Q.D. is known as "Coefficient of Quartile Deviation" which is defined as coefficient of $Q.D = \left[\frac{Q_3 - Q_1}{Q_3 + Q_1}\right]$. It is useful to compare the variations in two or more series of data.

Now, we shall explain the concept of 'Partition values' which include the concept of quartiles.

Mean Deviation (M.D.)

It is defined as the arithmetic mean of absolute deviations (obtained by ignoring the sign) of observations taken from an average (Mean, Median or Mode). Thus we have

(a) mean deviation about arithmetic mean

(b) mean deviation about median and

(c) mean deviation about mode. Some times we may choose any arbitrary value 'A' in the place of an average.

Case A Raw data:

Suppose a data set contains n observations say $x_1, x_2, \ldots\ldots x_n$. Let A be any arbitrary value (A may be A.M. or Median or Mode or any arbitrary value), then Mean deviation about 'A' is defined, by

M.D. = $\frac{\sum|x-A|}{n}$ where |x-A| is an absolute deviation taken from A.

Case B : In this case M.D is defined as M.D = $\frac{\sum f\,|x-A|}{N}$

Where $\sum f\,|x-A|$ is the sum of products of absolute deviations of observations from 'A' and their corresponding frequencies. N is total frequency.

Here x's are variable values and

f's are corresponding frequencies

A is any arbitrary value

Case C : Continuous frequency distribution : Here also the M.D. is defined same as in the case of discrete frequency distribution. but, midvalues of classes may be taken instead of variable values.

M.D. = $\frac{\sum f|x-A|}{N}$

where x's are mid values of classes

f's are frequencies

N is total frequency

A is any arbitrary mid value or an average value.

Uses of Mean Deviation :

It is used as an appropriate measure of variation in many economic studies. It is a better measure of dispersion than range and quartile deviation as it properly reflects the amount of variation in the data. It has a vital role in forecasting business cycles.

Remarks :

1. Mean deviation is minimum when it is calculated about median.
2. M.D is measure of absolute dispersion. A relative measure of dispersion based on M.D is known as "coefficient of M.D" which is defined as,

$$\text{Coefficient of M.D} = \frac{\text{M.D about an average}}{\text{Average from which it is calculated}} = \frac{\text{M.D}}{\text{A.M}} \text{ or } \frac{\text{M.D}}{\text{Median}} \text{ or } \frac{\text{M.D}}{\text{Mode}} \text{ etc.}$$

SOLVED PROBLEMS ON MEAN DEVIATION

Problem: Find range and Mean Deviation about Mean for the following data :

Marks	5-15	15-25	25-35	35-45	45-55
No. of students	1	5	14	3	2

Also find their coefficients.

Solution : Range

Upper boundary of the final class = 55

Lower boundary of the initial class = 5

Range = 55 – 5 = 50

$$\text{Coefficient of Range} = \left[\frac{55-5}{55+5}\right] = \left[\frac{50}{55}\right] = 0.9091$$

Mean Deviation about Mean :

Marks	No. of students	Mid value X	fx	$\|x-\bar{x}\|$	$f\|x-\bar{x}\|$
5-15	1	10	10	20	20
15-25	5	20	100	10	50
25-35	14	30	420	0	0
35-45	3	40	120	10	30
45-55	2	50	100	20	40
Total	25	-	750	-	140

From table we have N = 25, $\sum fx = 750$

$$\text{A.M.} = \bar{x} = \frac{\sum fx}{N} = \frac{750}{25} = 30$$

$$\text{M.D. about A.M.} = \frac{\sum f\,|x-\bar{x}|}{N} = \frac{140}{25} = 5.6$$

$$\text{Coefficient of M.D.} = \frac{M.D}{Mean} = \frac{5.6}{30} = 0.1867$$

MERITS AND DEMERITS OF RANGE, QUARTILE DEVIATION AND MEAN DEVIATION

We can compare various measures of dispersion according to the desirable characteristics of an ideal measure of dispersion, proposed by Prof. Yule. These characteristics are the same as those for an ideal measure of central tendency.

Measure of Dispersion	Merits	Demerits
(a) Range	(i) It is the simplest measure of dispersion (ii) It is easy to calculate (iii) It is easy to understand.	(i) It is not rigidly defined (ii) It is not based on all the observations. (iii) It is not suitable for further mathematical manipulations. (iv) It is very much affected by fluctuations of sampling. (v) It is affected by extreme values. (vi) It can not be calculated for distribution with open ended classes.
(b) Quartile deviation	(i) It is rigidly defined. (ii) It is easy to calculate. (iii) It is easy to understand (iv) It is not affected by extreme values. (v) It is the most appropriate measure of dispersion when a frequency distribution has open end classes.	(i) It is not based on all the observations. (ii) It is not suitable for further mathematical manipulations (iii) It is affected by sampling fluctuations
(c) Mean deviation	(i) It is rigidly defined. (ii) It is based on all the observations.	(i) It is not easy to calculate. (ii) It is not easy to understand. (iii) It is not suitable for further mathematical manipulations. (iv) It is affected considerably by sampling fluctuations.

Standard Deviation (S.D):

It is defined as the positive square root of the arithmetic mean of the squares of the deviation of the observations taken from their arithmetic mean. It is usually denoted by the Greek letter (small sigma) σ. The square of standard deviation is known as the variance of data (σ^2).

Case A : Raw data :

Suppose raw data contain 'n' observations, say, $x_1, x_2, \ldots, x_n$. The standard deviation is defined as .

(i) Direct formula : $\sigma = \sqrt{\dfrac{\sum(x-\bar{x})^2}{n}}$ where $\bar{x} = \dfrac{\sum x}{n} = A.M.$

(ii) Simplified Formula : $\sigma = \sqrt{\dfrac{\sum X^2}{n} - (\bar{X})^2}$ where $\bar{X} = \dfrac{\sum x}{n}$

(iii) Short cut formula.

$$\sigma = \sqrt{\frac{\sum d^2}{n} - (\bar{d})^2} \text{ where } \bar{d} = \frac{\sum d}{n}$$

d = x-A, A is an origin parameter chosen suitabley

Remarks : 1. Other equivalent presentations S.D.

(a) $\sigma = \sqrt{\dfrac{1}{n}\left(\sum x^2 - \dfrac{(\sum x)^2}{n}\right)}$ where $\bar{x} = \dfrac{\sum x}{n}$

(b) For computational purpose, we generally use either simplified formula or short cut method formula for S.D.

Case (B): Discrete frequency distribution:

Suppose $x_1, x_2, \ldots, x_k$ be variable values and $f_1, f_2, \ldots, f_k$ be their respective frequencies. The standard deviation is given by.

(i) Direct formula : $\sigma = \sqrt{\dfrac{\sum f(x-\bar{X})^2}{N}}$, where $\bar{X} = \dfrac{\sum fx}{N}$, N = Total frequency

(ii) Simplified formula : $\sigma = \sqrt{\dfrac{\sum fx^2}{N} - (\bar{X})^2}$, where $\bar{X} = \dfrac{\sum fx}{N}$

(iii) Short cut formula :

$\sigma = \sqrt{\frac{\sum fd^2}{N} - (\bar{d})^2}$, where $\bar{d} = \frac{\sum fd}{N}$ d = [x - A], 'A' is any arbitrary value.

suitably chosen.

In this case, the A.M. of the data is given by

A.M. = $\bar{X} = A + \bar{d}$

(c)Continuous frequency distribution :

In the case of continuous frequency distribution, the formula is the same as in the discrete case. But the difference is that x's are midvalues of class intervals instead of variable values.

(i) Direct formula : $\sigma = \sqrt{\frac{\sum f(x-\bar{X})^2}{N}}$, where $\bar{X} = \frac{\sum fx}{N}$,

Here, x's are midvalues of class

(ii) Simplified method : $\sigma = \sqrt{\frac{\sum fx^2}{N} - (\bar{X})^2}$, where $\bar{X} = \frac{\sum fx}{N}$

Step Deviation Method for Standard Deviation

If the bounds of class intervals are sufficiently large in magnitude, and class intervals are of equal width, to reduce computational labour we use the step deviation method.

By this method, we have

$$\bar{X} = A + \left(\frac{\sum fd}{N}\right)C$$

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

where d = $\frac{X - A}{C}$

x's are mid values of classes

'A' is origin parameter, chosen as mid value of middle most class interval.

C is the common length of class intervals

Also variance of the data is given by

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

Properties of Standard Deviation :

The following are some of the important properties of standard deviation.

1. Standard Deviation is invariant (unchanged) under change of origin (if d=x- A).
2. It is invariant under change of scale (if d = $\frac{x-A}{C}$)
3. The value of S.D. is always positive.
4. The value of S.D. is always less than or equal to Range. i.e. S.D. ≤ Range

Combined Standard Deviation

If n_1, n_2 are the sizes , $\bar{x}_1, \bar{x}_2$ arithmetic means; and σ_1, σ_2 the standard deviations of two series, then the standard deviation of combined series is given by,

$$\sigma = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$$

where $d_1 = \bar{x}_1 - \bar{\bar{X}}$ and $d_2 = \bar{x}_2 - \bar{\bar{X}}$, $\bar{\bar{X}} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

Uses of Standard Deviation: Standard Deviation is one of the most important and commonly used measure of dispersion. It is often used in advanced statistical analysis, particularly in testing of hypotheses and statistical quality control. By comparing the standard deviations of two or more series, we can compare the degree of variability or consistency. It is used to compute some advanced statistical measures like

(1) Coefficient of Skewness
(2) Coefficient of Kurtosis
(3) Coefficient of Correlation
(4) Standard Error etc.

Coefficient of Variation (C.V.):

it is a relative measure of dispersion. This concept was first suggested by Prof. Karl Pearson in 1895. For comparing the variability of two or more series, we calculate the coefficient of variations for each series. The C.V. is defined as 100 times the coefficient of dispersion based upon standard deviation.

$$\text{i.e., C.V.} = \left(\frac{\sigma}{\bar{x}}\right)100$$

The coefficient of variation as defined above is a measure of relative variation. It is independent of the units of measurement of the observations of the data set. The data with lesser C.V. are known as consistent data or homogeneous data with less variation. The data with larger C.V. is known as inconsistent data or heterogeneous data or the Data with more variation.

SOLVED PROBLEMS ON STANDARD DEVIATION

Problem: Calculate the Mean and S.D. for the following data:

Daily Wages (in Rs.)	40-50	50-60	60-70	70-80	80-90	90-100
No. of workers	10	15	25	35	8	7

Solution : Direct Method :

Daily wages (in Rs. class)	No. of workers f	Mid value x	Fx	fx^2
40-50	10	45	450	20250
50-60	15	55	825	45375
60-70	25	65	1625	105625
70-80	35	75	2625	196875
80-90	8	85	680	57800
90-100	7	95	665	63175
Total	100	-	6870	489100

From the above table we have N = 100, $\sum fx = 6870$, $\sum fx^2 = 489100$

$$A.M. = \bar{X}\left[\frac{\sum fx}{N}\right] = \frac{6870}{100} = 68.7$$

$$S.D = \sigma = \sqrt{\frac{\sum fx^2}{N} - (\bar{X})^2}$$

$$=\sqrt{\frac{489100}{100}-(68.7)^2} = \sqrt{4891-4719.69} = \sqrt{171.31} = 13.0$$

Hence, mean daily wages of workers is given by Rs. 68.70 and standard deviation of daily wages is given by Rs. 13.0.

II Method: Step Deviation Method:

Class	F	Mid value x	$d=\frac{x-65}{10}$	fd	fd^2
40-50	10	45	-2	-20	40
50-60	15	55	-1	-15	15
60-70	25	65=A	0	0	0
70-80	35	75	1	35	35
80-90	8	85	2	16	32
90-100	7	95	3	21	63
Total	100	-	-	37	185

From the table, we have N = 100, C = 10, $\sum$fd = 37 and $\sum fd^2$ = 185.

$$A.M = \bar{X} = A + \left[\frac{\sum fd}{N}\right]C$$

$$= 65 + \left[\frac{37}{100}\right]10 = 65 + 3.7 = 68.7$$

$$S.D = \sigma = \left\{\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}\right\}C$$

$$= \left\{\sqrt{\frac{185}{100} - \left(\frac{37}{100}\right)^2}\right\}10 = \left\{\sqrt{1.85 - 0.1369}\right\}10 = \left\{\sqrt{1.7131}\right\}10$$

$$= (1.30885)10 = 13.0885$$

Hence A.M = Rs 65.70, SD = Rs. 13.09

Problem: Find the Mean and S.D for the following data.

Daily wages (Rs.)	15-25	25-30	30-50	50-60	60-100
No. of Employees	7	15	20	12	8

Solution: Since, we are given unequal class intervals, we can use simplified method to compute S.D.

Daily wages class	No. of employees (f)	Mid value (x)	fx	fx^2

15-25	7	20	140	2800
25-30	15	27.5	412.5	11343.75
30-50	28	40	1120	44800
50-60	12	55	660	36300
60-100	8	80	640	51200
Total	70	-	2972.5	146443.75

From the computation table we have, N = 70

$\sum fx = 2972.5$, $\sum fx^2 = 146443.75$

A.M. = $\bar{x} = \left[\frac{\sum fx}{N}\right]$

$$= \frac{2972.5}{70} = 42.4643$$

$$S.D = \sigma = \sqrt{\frac{\sum fx^2}{N} - (\bar{X})^2}$$

$$= \sqrt{\frac{146443.75}{70} - (42.4643)^2} \quad = \sqrt{2092.0536 - 1803.2968} \quad = \sqrt{288.8368}$$

$$\Rightarrow \sigma = 16.9952$$

Hence, A.M. and SD of daily wages are respectively Rs. 42.46 and Rs. 17.00.

Problem: the means of two series of 50 and 100 observations respectively are 54 and 50 and the standard deviations are 8 and 6. Obtain the standard deviation of the combined series of 150 observations.

Solution: Given that $n_1 = 50$, $n_2 = 100$, $\bar{X}_1 = 54, \bar{X}_2 = 50, \sigma_1 = 8, \sigma_2 = 6$

Combined Arithmetic mean = $\bar{X} = \left[\frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}\right]$

$$= \frac{50(54) + 100(50)}{150}$$

$$= \frac{2700 + 500}{150}$$

$$= \frac{7700}{150} = 51.3333$$

Consider $d_1 = [(\bar{X}_1 - \bar{X})] = 54 - 51.3333 = 2.6667$

$d_2 = [(\bar{X}_2 - \bar{X})] = 50 - 51.3333 = 1.3333$

Combined standard deviation $\sigma = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$

$$= \sqrt{\frac{(50)(8)^2 + (100)(6)^2 + (50)(2.6667)^2 + (100)(1.3333)^2}{50 + 100}}$$

$$= \sqrt{\frac{7333.335}{150}} = \sqrt{48.8889} = 6.9921$$

Combined AM and S.D are respectively, 51.33 and 6.9921.

Problem: Two workers on the same job show the following results over a long period of time :

	Worker A	Worker B
Mean time of completing the job (in minutes)	30	25
S.D (in minutes)	6	4

(i)Which worker appears to be more consistent in time, to complete the job?

(ii)Which worker appears to be faster in completing the job?

Solution: We find C.V. for worker A and B separately.

Worker A: $\bar{X} = 30, \sigma = 6$

C.V. (for worker A) = $\left(\frac{\sigma}{\bar{X}}\right)100 = \left(\frac{6}{30}\right)100 = 20\%$

Worker B: $\bar{X} = 25, \sigma = 4$

C.V. (for worker B) = $\left(\frac{\sigma}{\bar{X}}\right)100 = \left(\frac{4}{25}\right)100 = 16\%$

Since C.V. is smaller for worker-B, he appears be more consistent in time to complete the job.

Worker–B had taken on an average 25 minutes as against 30 minutes taken by worker–A. to complete the job. Therefore worker – B appears to be faster in completing the job.

Problem: Mean of 100 items is 50 and their standard Deviation is 4. Find the sum and sums of squares of all the items.

Solution: Given that $\bar{X} = 50. n = 100. S.D. = \sigma = 4$

We have, $\bar{X} = \frac{\sum x}{n}$ or $\sum x = n\bar{X} = (100)(50) = 5000$

Also, $\sigma = \sqrt{\frac{\sum x^2}{n} - (\bar{X})^2}$ or $\sigma^2 = \frac{\sum x^2}{n} - (\bar{X})^2$

Substituting the values of n, $\bar{X}$ and σ, we get

$$16 = \frac{\sum x^2}{100} - (50)^2$$

$$i.e. \frac{\sum x^2}{100} = (50)^2 + 16 = 2516$$

$$\therefore \sum x^2 = (2516)(100) = 251600$$

Hence, sum of items $\sum x = 5{,}000$ and sum of squares of items $\sum x^2 = 251{,}600$

Problem: The mean and the Standard Deviation of a sample of 100 observations were calculated as 40 and 5.1 respectively by a student, who took by mistake 50 instead of 40 for one observation. Calculate the correct Mean and Standard Deviation.

Solution: Given that $\bar{X} = 40, \sigma = 5.1$ and $n = 100$

We have, $\bar{X} = \frac{\sum x}{n}$ or $\sum x = n\bar{X} = (100)(40) = 4000$

We have $\sigma^2 = \frac{\sum x^2}{n} - (\bar{X})^2$ or $\frac{\sum x^2}{n} = \sigma^2 + (\bar{X})^2$ or $\sum x^2 = n\left[\sigma^2 + (\bar{X})^2\right]$

$$or \sum x^2 = (100)\left[(5.1)^2 + (40)^2\right]$$

$$= 100\,(162.01) = 162601$$

While calculating sum, the student took 50 instead of 40. The difference between correct and wrong observation = 50 – 40 = 10

Correct sum = $\sum x$-Difference = 4000 - 10 = 3990

Correct A.M. = Correct sum/n =3990/100=39.9

The difference between the squares of correct and wrong observations.

$$=(50)^2-(40)^2=2500-1600=900$$

Correct sum of squares = $\sum x^2$ - Difference between the squares of correct and wrong observations.

= 162,601 – 900 = 161,701

$$\text{Correct S.D} =\sqrt{\frac{\text{correct}\sum x^2}{n}-(\text{Correct }\bar{x})^2}$$

$$=\sqrt{\frac{161701}{100}-(39.9)^2}$$

$$=\sqrt{1617.01-1592.01}$$

$$=\sqrt{25}=5$$

Correct $\bar{X}$ =39.9 and correct S.D. = 5.

Problem: find the mean and standard deviation of the first n natural numbers.

$$\sum x=1+2+3........+n=\frac{n(n+1)}{2}$$

$$\sum x^2=1^2+2^2+3^2+..................+n^2=\frac{n(n+1)(2n+1)}{6}$$

$$\text{Mean }(\bar{X})=\frac{\sum x}{n}=\frac{n(n+1)}{2n}=\frac{n+1}{2}$$

$$S.D.=\sigma=\sqrt{\left[\frac{\sum x^2}{n}-(\bar{X})^2\right]}$$

$$=\sqrt{\frac{n(n+1)(2n+1)}{6n}-\frac{(n+1)^2}{4}}$$

$$=\sqrt{\frac{2(n+1)(2n+1)-3(n+1)2}{12}}=\sqrt{\frac{(n+1)(4n+2-3n-3)}{12}}$$

$$=\sqrt{\frac{(n+1)(n-1)}{12}}=\sqrt{\frac{n^2-1}{12}}$$

Hence, Mean and S.D. of the first 'n' natural numbers are given by $\left(\frac{n+1}{2}\right)$ and $\sqrt{\frac{n^2-1}{12}}$ respectively.

Problem: The mean and standard deviation of 20 items is found to be 10 and 2 respectively. At the time of checking it was found that one item 8 was incorrect. Calculate the mean and standard deviation if

(i) the wrong item is omitted,

(ii) if it is replaced by 12

Solution: Given $n = 20, \overline{X} = 10$ *and* $\sigma = 2$ *or* $\sigma^2 = 4$

Uncorrected $\sum x = n\,\overline{X} = (20)(10) = 200;$

We have, $\sigma^2 = \frac{\sum x^2}{n} - \left(\overline{x}\right)^2$ *or* $\sum x^2 = n\left(\sigma^2 + \overline{X}^2\right)$

Uncorrected $\sum x^2 = n\left(\sigma^2 + \overline{X}^2\right) = 20(4+100) = 2080$

(i) Corrected $\sum x = 200 - 8 = 192$

Corrected $\sum x^2 = \text{Uncorrected}\sum x^2 - (\text{omitted observation})^2$

$= 2080 - (8)^2 = 2016$

Corrected $\overline{x} = \frac{\text{corrected}\sum x}{n_1} = \frac{192}{19} = 10.1053$

Corrected $\sigma^2 = \left[\frac{2016}{19} - (10.1053)^2\right] = [106.1053 - 102.1171] = 3.9882$

$\therefore \textit{Corrected S.D.} = \sqrt{3.9882} = 1.9970$

(ii) Replace wrong observation 8 by correct value 12

Corrected $\sum x = 200 - 8 + 12 = 204$

Corrected $\sum x^2 = 2080 - (8)^2 + (12)^2 = 2080 - 64 + 144 = 2160$

Now, we have, n = 20

$$\text{Corrected } \overline{X} = \frac{204}{20} = 10.2$$

$$\text{Corrected } \sigma^2 = \left[\frac{2160}{20} - (10.2)^2\right]$$

$$= [108 - 104 - 04] = 3.96$$

$$\text{Corrected } \sigma = \sqrt{3.96} = 1.9900$$

Problem: In the following data, two class frequencies are missing.

Class	Frequency
100 – 110	4
110 – 120	7
120 – 130	15
130 – 140	?
140 – 150	40
150 – 160	?
160 – 170	16
170 – 180	10
180 – 190	6
190 – 200	3

Given that the total frequency and median are respectively 150 and 146.25 (i) find the missing frequencies, (ii) hence find AM and SD (iii) without using direct formula find the value of mode.

Solution: Suppose the two missing frequencies be f_1 and f_2 respectively for the two classes 130 – 140 and

150 – 160

Class	Frequency (f)	Less than cumulative frequency
100 – 110	4	4
110 – 120	7	11
120 – 130	15	26
130 – 140	f_1	$26+f_1 = m$
140 – 150	$40 = f$	$66+f_1$
150 – 160	f_2	$66+f_1+f_2$
160 – 170	16	$82+f_1+f_2$
170 – 180	10	$92+f_1+f_2$
180 – 190	6	$98+f_1+f_2$
190 – 200	3	$101+f_1+f_2$
Total	$101+f_1+f_2$	-

Given N = 150 = 101+f_1+f_2

$\therefore$ f_1+f_2 = 150 – 101 = 49 … (1)

We are given median as 146.25. The median class is 140 – 150. The median formula,

$$\text{Median} = l + \left(\frac{\frac{N}{2} - m}{f} \right) C$$

$$\text{we get, } 146.25 = 140 + \left[\frac{\frac{150}{2} - (26 + f_1)}{40} \right] 10$$

$$6.25 = \frac{75.26 - f_1}{4}$$

$\therefore$ (6.25) (4) = 49-f_1

$\therefore$ f_1 = 49 – 25 = 24

From (1) 24+f_2 = 49

f_2 = 49-24 = 25 , By substituting f_1 = 24 and f_2 = 25, we have ,

Class	Mid Value X	Frequency f	$d = \frac{X-145}{10}$	fd	fd^2
100 – 110	105	4	-4	-16	64
110 – 120	115	7	-3	- 21	63
120 – 130	125	15	-2	- 30	60
130 – 140	135	24	-1	-24	24
140 – 150	145 = A	40	0	0	0
150 – 160	155	25	1	25	25
160 – 170	165	16	2	32	64
170 – 180	175	10	3	30	90
180 – 190	185	6	4	24	96
190 – 200	195	3	5	15	75
Total	-	N = 150	-	Σ fd = 35	Σ fd^2 = 561

$$= 145 + \left(\frac{35}{150} \right) 10 = 147.3333$$

$$S.D. = \left\{ \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N} \right)^2} \right\} C$$

$$= \left\{ \sqrt{\frac{561}{150} - \left(\frac{35}{150} \right)^2} \right\} 10 = \left\{ \sqrt{3.75 - 0.0544} \right\} 10$$

$$= \left\{ \sqrt{3.6856} \right\} 10 \quad = (109198)10 \quad =19.198$$

(iii) Without using the direct formula, mode can be obtained by using the empirical relationship.

MODE = 3 MEDIAN – 2 MEAN

= 3 x 146.25 – 2 x 147.3333

= 438.75 – 294.6667

= 144-0833.

Problem In an industrial establishment, the coefficients of variation of wages of male and female workers were 55% and 70% respectively. Standard Deviations were Rs. 22.00 and Rs. 15.40 respectively. Calculate the combined average wage of all the workers if 80% of the workers are male.

Solution: C.V. for Male workers = 55% and C.V. for female workers = 70%

Suppose $\bar{X}_1$ *and* $\bar{X}_2$ denote the mean wages (in Rs. and) σ_1, σ_2 denote the standard deviation of the wages (in Rs.) of male and female workers respectively in the industrial establishment.

$\sigma_1 = 22.00$, $\sigma_2 = 15.40$

Since, $C.V. = \left[\frac{\sigma}{\bar{X}} \right] 100, we\, have, \bar{X} \left[\frac{\sigma}{C.V.} \right] 100$

For Male workers :

$$\bar{X}_1 = \left[\frac{\sigma_1}{C.V. fFor\ Male} \right] 100 = \left[\frac{22}{55} \right] = Rs. 40/-$$

For female workers :

$$\therefore \bar{X}_2 = \left[\frac{\sigma_2}{C.V.\,for\ Females}\right]100 = \left[\frac{(15.40)}{70}\right]100 = 22$$

Let the total no. of workers in the establishment be 100. Since given that 80% of workers were males, we have, no. of female workers = 100 – 80 = 20

Hence, the combined average wage $\bar{X}$ of all the workers in the establishment is

$$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2} = \frac{(80)(40)+(20)(22)}{100} = \frac{3200+440}{100} = \frac{3640}{100} = 36.40$$

$\therefore$Combined average wage of all the workers is given by Rs 36-40

MERITS AND DEMERITS OF STANDARD DEVIATION :

MERITS	DEMERITS
1. It is well defined. 2. It is based on all the observations. 3. It is suitable for further mathematical manipulations. 4. It is least affected by sampling fluctuations. 5. It is one of the most stable measures of dispersion. It has a vital role in the theory of statistics.	1. It is not easy to calculate. 2. It is not easy to understand. 3. It is affected by extreme values. 4. It can not be computed for distributions with open end classes.