CHAPTER 2

Frequency Distribution–Measures of Central Values and Dispersion

INTRODUCTION

Variability is the most common characteristic of the data with reference to agriculture, biological and physical science research. The need for statistical method arises from the variability.

Examples where variation is observed

1. Number of seedlings
2. Blood group in man and their frequency
3. Plaque morphology of Bacteriophages infecting different strains of *E. coli.*
4. Plant height
5. Number of insects with specific eye colour
6. Eye colour in *Drosophila.*
   Characteristics which show variation of this sort are called random variables or variates.

Continuous variable: A continuous variable is one that can take any value in a given range which may be finite or infinite. For example, the yield of a crop, height of a plant, animal height, human height are continuous variables.

Discrete variables: A discrete variable is one for which the possible values are not observed on a continuous scale.

For example, number of children in a family, number of fingers, number of plants bearing reel flowers, number of *Drosophila* with red eyes, number of wild type and mutant colonies of a microorganism, number of lethals, (dominant or recessive lethals), number of fragments in a cell observed during cell division.

POPULATION AND SAMPLE

A population is defined as the total set of actual or possible values of the variable. A population may be finite or infinite, and the variable, continuous or discrete.

The idea of infinite populations distributed in a frequency distribution in respect of one or more characters is fundamental to all statistical work (Fisher -1948). One of the principle objectives of statistics is to draw inferences with respect to populations by the study of groups of individuals forming part of the populations.

A sample is therefore any finite set of items drawn from a population. The purpose of drawing samples is to obtain information about the populations from which they are drawn. A random sample from a given population is a sample, chosen in such a manner that each possible sample has an equal chance of being drawn.

Quantities which characterise populations are known as parameters. Characters of sample are called statistics. A parameter is a fixed quantity, statistic is a variate. Generally, a statistic is sought which 'best' estimate the corresponding population parameter.

FREQUENCY DISTRIBUTIONS

The assemblage of x_i with their associated frequencies f is called a frequency distribution. A typical frequency distribution is presented in table.

Frequency distribution of variable X

Value of variable	Frequency
X_1	f_1
X_2	f_2
X_i	f_i
X_n	f_n
Total frequency	N

Example 2.1 *Prepare a frequency distribution for the following data of height of 20 children.*

133	136	120	138	133
131	127	141	127	143
130	131	125	144	128
134	135	137	133	129

Frequency distribution

Height(cm)	Tally	Frequency
119.5–124.5	|	1
124.5–129.5	~~||||~~	5
129.5–134.5	|| ~~||||~~	7
134.5–139.5	||||	4
139.5–144.5	|||	3
	Total	20

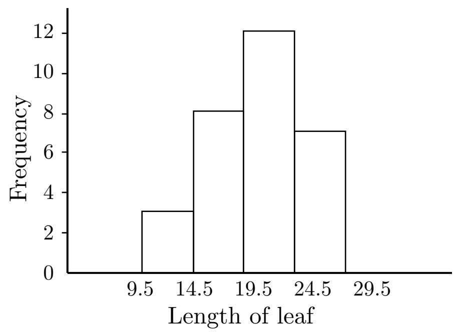
HISTOGRAM

Grouped data can be displayed in a histogram.

In a histogram rectangles are drawn so that the area of each rectangle is proportional to the frequency in the range covered by it.

Example 2.2 *The lengths of 30 plant leaves of species A were measured and the information grouped as shown. Measurements were taken correct to the nearest cm. Draw a histogram to illustrate the data.*

Length of leaf (cm)	9.5–14.5	14.5–19.5	19.5–24.5	24.5–29.5
Frequency	3	8	12	7



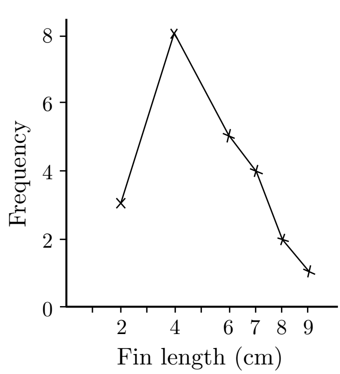
FREQUENCY POLYGONS

A frequency distribution may be displayed as a frequency polygon.

A frequency polygon may be superimposed on a histogram by joining the midpoints of the tops of the rectangles. This is in grouped data.

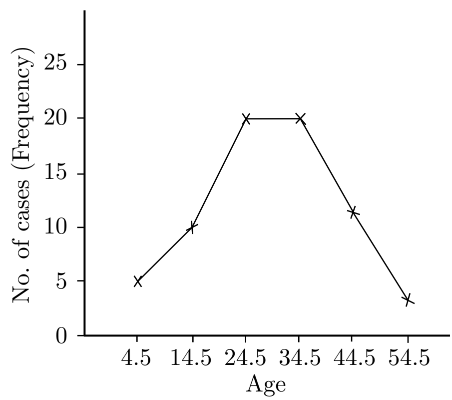
(a) *Ungrouped data:* The fin length in (cm) of particular type of fish is given. Draw a frequency polygon to illustrate this information.

Fin length (cm)	2	4	6	7	8	9
Frequency	3	8	5	4	2	1



(b) *Grouped data:* The following table shows the age distribution of cases of a certain disease reported during a year in a particular state. Prepare a frequency polygon.

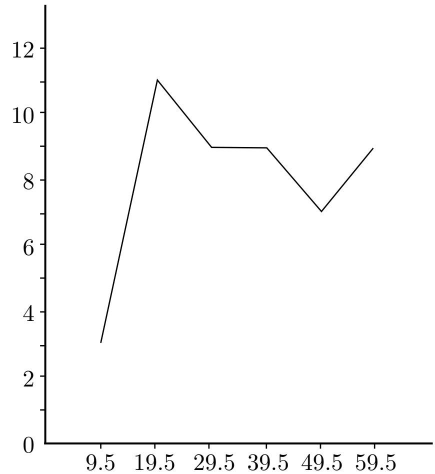
Age	**Number of cases**
5–14	5
15–24	10
25–34	22
35–44	22
45–54	13
55–64	5



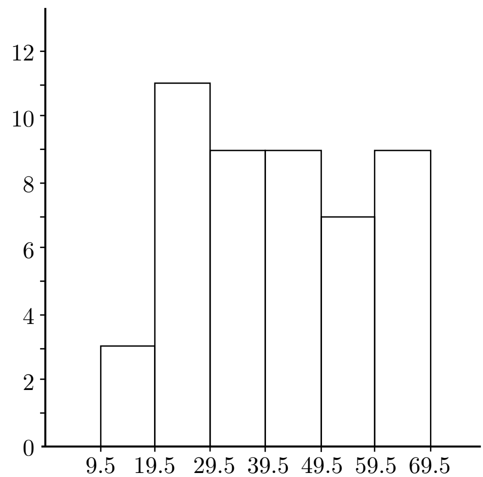
Example 2.3 *Pollen grain length in microns is given below. Prepare frequency table. Represent cumulative frequency, draw histogram, frequency polygon.*

63,	64,	64,	67,	62,	60,	69,	68
19,	16,	10,	20,	22,	20,	20,	21
21,	23,	26,	25,	24,	27,	32,	30
31,	33,	34,	38,	39,	36,	43,	46
44,	47,	50,	55,	59,	58,	58,	41
42,	48,	52,	42,	42,	64,	30,	54

Class interval	**Frequency**	**CF**
10–19	||| = 3	3
20–29	𝍸 𝍸	= 11	3 + 11 = 14
30–39	𝍸 |||| = 9	14 + 9 = 23
40–49	𝍸 |||| = 9	23 + 9 = 32
50–59	𝍸 || = 7	32 + 7 = 39
60–69	𝍸 |||| = 9	39 + 9 = 48
	Total 48	



Frequency polygon



Histogram

MEASURES OF CENTRAL TENDENCY

One of the most important objectives of statistical analysis is to get one single value that describes the characteristic of the entire mass of data.

There are three main statistical measures which attempt to locate a 'typical' value. These are

1. Arithmetic mean (A.M.)
2. Median
3. Mode

Other measures of central tendency are

1. Geometric Mean (G.M.)
2. Harmonic Mean (H.M.)

ARITHMETIC MEAN

A numerical value which indicates the centre of the distribution is called the Arithmetic Mean.

For ungrouped data

$$\bar{X} = \sum_{i=1}^{n} \frac{x_i}{n}$$

Grouped data

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

where $n = \sum f$; $\sum(Sigma)$ = Summation; $\bar{X}(X\text{-}Bar)$ = Mean

Example 2.4 *Find the mean of set of numbers 63, 65, 67, 68, 69, 70, 71, 72, 74, 75.*

$$n = 10$$

$$\sum x = 63 + 65 + 67 + 68 + 69 + 70 + 71 + 72 + 74 + 75 = 694$$

$$\bar{X} = \frac{\sum X}{n} = \frac{694}{10} = 69.4$$

Frequency Distribution

No. of flowers *(X)*	1	2	3	4	5
No. of plants *(f)*	11	10	5	3	1

X	***f***	***fX***
1	11	11
2	10	20
3	5	15
4	3	12
5	1	5
	$\sum f = 30$	$\sum fX = 63$

$$\bar{X} = \frac{\sum fX}{n} = \frac{63}{30} = 2.1$$

where $n = \sum f$

Grouped Frequency Distribution

The lengths of 32 leaves were measured correct to the nearest mm. Find the mean length.

Length (mm)	20–22	23–25	26–28	29–31	32–34
Frequency	3	6	12	9	2

Length (mm)	**Midpoint *(X)***	***f***	***fX***
20–22	21	3	63
23–25	24	6	144
26–28	27	12	324
29–31	30	9	270
32–34	33	2	66
		$\sum f = 32$	$\sum fX = 867$

$$\bar{X} = \frac{\sum fX}{n} = \frac{867}{32} = 27.1 \text{ m}$$

where $n = \sum f$

Merits and demerits of arithmetic mean

1. It is the simplest to understand and the easiest to compute.
2. It is affected by the value of every item in series.

Demerit

1. Arithmetic mean is not always a good measure of central tendency, as for instance in extremely symmetrical distribution.

MEDIAN

The median refers to the middle value in a distribution. The median is called a positional average. The median is the middle value of a set of numbers arranged in order of magnitude. The median is the $\frac{1}{2}(n+1)$th value. The median is that value of the variate for which 50% of the observations, when arranged in order of magnitude, lie on each side. Median is the value of the variate which divides the total frequency in the whole range into two equal parts. Median is not affected by extreme values.

If the total frequency is even, the median is the arithmetic mean of the two middle values. Compared with the arithmetic mean, the median places less emphasis on the minimum or maximum value in the sample or population.

Example 2.5 *Find the median of each of the sets.*

(a) $7, 7, 2, 3, 4, 2, 7, 9, 31$

(b) $36, 41, 27, 32, 29, 38, 39, 43$

Solution:

(a) 2, 2, 3, 4, $\boxed{7,}$ 7, 7, 9, 31

$n = 9$, and the median is the $\frac{1}{2}(9+1)$ the value, i.e., the 5th value.

Median = 7

(b) Arranging in order of magnitude

27, 29, 32, $\boxed{36,\ 38,}$ 39,41,43

$n = 8$ and the median is average of the $\frac{1}{2}(8+1)$ the value, i.e., the $4\frac{1}{2}$ value. This does not exist, so we consider the 4th and 5th values.

$$\text{Median} = \frac{1}{2}(36+38) = 37$$

In general, if n is odd then there is a middle value, and this is the median. If n is even and the two middle value are c and d, then the median is $\frac{1}{2}(c+d)$.

Calculation of median

Grouped data

$$\text{Median} = L + \frac{(n/2) - m}{f} \times C$$

where L = Lower limit of median class; m = Cumulative frequency above median class; f = Frequency of median class; C = Class interval.

Example 2.6 *Find median of the following distribution*

Class interval	15–25	25–35	35–45	45–55	55–65	65–75	75–85
Frequency	3	61	132	153	136	51	6

Class interval (CI)	**Frequency (F)**	**Cumulative frequency (CF)**
15–25	3	3
25–35	61	64
35–45	132	196 m
45–55	153 F	349
55–65	136	485
65–75	51	536
75–85	6	542 $\leftarrow N$

$$\frac{N}{2} = \frac{542}{2} = 271$$

(This is nearer to 349 under CF) $\therefore$ Median class = 45-55.

$$\text{Median} = L + \frac{(n/2) - m}{f} \times C = 45 + \frac{271 - 196}{153} \times 10 = 49.9$$

MODE

Mode or modal value of a given data is the one which has maximum frequency or any value of the data which occurs repeatedly.

Example 2.7 *Find the mode of the following 2, 1, 0, 3, 5, 2, 3, 1, 0, 2, 2, 2, 4*

Value	0	1	2	3	4	5
Frequency	2	2	5	2	1	1

2 is the mode (unimodal) as this is having high frequency.

Example 2.8 *Find the mode of the following 1, 1, 2, 3, 0, 2, 4.*

Value	0	1	2	3	4
Frequency	1	2	2	1	1

These are two modal values in the data they are 1 and 2 (bimodal) having high frequency.

Mode = 3 Median − 2 Mean

Example 2.9 *Find the Mode of the following*

Class interval (CI)	**Frequency (F)**
15–25	3
25–35	61
35–45	132 f_1
45–55	153 f_0
55–65	136 f_2
65–75	51
75–85	6

$$\text{Mode}=L+\frac{\Delta_1}{\Delta_1+\Delta_2}\times C$$

By inspection the highest frequency is (153) and the modal class is 45–55 where l = lower limit of the modal class

$$\Delta_1 = f_0 - f_1(153 - 132 = 21)$$
$$\Delta_2 = f_0 - f_2(153 - 136 = 17)$$
$$C = \text{length of class interval} = 10$$
$$\text{Mode} = 45 + \frac{21}{21+17}\times 10 = 45 + \frac{210}{38} = 50.5$$

HARMONIC MEAN FOR GROUPED DATA AND UNGROUPED DATA

Example 2.10 *Calculate Harmonic mean for the following grouped data,*

Class interval (CI)	0–10	10–20	20–30	30–40
Frequency (f)	4	5	5	7

CI	x	f	**1/x**	f**(1/x)**
0–10	5	4	(1/5)	$(1/5) \times 4 = (4/5)$
10–20	15	5	(1/15)	$(1/15) \times 5 = (1/3)$
20–30	25	5	(1/25)	$(1/25) \times 5 = (1/5)$
30–40	35	7	(1/35)	$(1/35) \times 7 = (1/5)$

$$\sum f.\frac{1}{x} = \frac{23}{15} = 1.53$$

$$H.M. = \frac{N}{\sum f \times \frac{1}{x}} = \frac{21}{1.53} = 13.72$$

where $N = \sum f$

Harmonic mean for ungrouped data

Harmonic mean is calculated by the following formula:

$$H.M. = \frac{N}{\left(\frac{1}{x_1} + \frac{1}{x_2} \ldots \frac{1}{x_n}\right)}$$

$x_1,\ x_2, \ldots, x_n$ are variables.

Example 2.11 *Calculate the harmonic mean of the following*
1, 0.5, 10, 45.0, 175, 0.01, 4.0, 11.2 ***(B.Com., Mysore, 1967)***

x	**l/x**
1	1.0000
0.5	2.0000
10	0.1000
45.0	0.0222
175	0.0057
0.01	100.00
4.0	0.2500
11.2	0.0893
	$\sum 1/x = 103.4672$

$$H.M. = \frac{N}{\sum \frac{1}{x}} = \frac{8}{103.467} = 0.0777$$

Hence,

$$\text{Mean deviation} = \frac{\left|X_1 - \bar{X}\right| + \left|X_2 - \bar{X}\right| + \ldots \left|X_n - \bar{X}\right|}{n}$$

$$\text{M.D.} = \text{Modulus } \frac{\sum \left|X_i - \bar{X}\right|}{n}$$

(i.e., absolute values are taken e.g.: 3-8 or (8-3) is written as 5)

Example 2.14 *Find the mean deviation of the scores 3, 5, 7, 9, 11 and 13 from the arithmetic mean.*

$$\text{Arithmetic mean } \bar{X} = \frac{3+5+7+9+11+13}{6} = \frac{48}{6} = 8$$

$$\text{Mean deviation} = \frac{\sum \left|X - \bar{X}\right|}{n} \quad \text{or} \quad \frac{\sum |d|}{N}$$

$$= \frac{|3-8| + |5-8| + |7-8| + |9-8| + |11-8| + |13-8|}{6}$$

$$= \frac{5+3+1+1+3+5}{6} = 3$$

Example 2.15 *Find the mean deviation from the A.M. for the following distribution*

Class interval	10–20	20–30	30–40	40–50	50–60
Frequency	8	15	25	9	3

Class interval	**Mid value X**	**Frequency f**	fX	$\lvert d\rvert = x - 32.33$ $\lvert x - \bar{x}\rvert$	$f.\lvert d\rvert$
10–20	15	8	120	17.33	138.64
20–30	25	15	375	7.33	109.95
30–40	35	25	875	2.67	66.75
40–50	45	9	405	12.67	114.03
50–60	55	3	165	22.67	68.01
		60	1940		497.38

$$\text{Mean} = \frac{\sum fX}{\sum f} = \frac{1940}{60}$$

$$\text{Mean Deviation} = \frac{\sum f\,|d|}{N} = \frac{497.38}{60} = 8.3$$

QUARTILE DEVIATION

Calculate the quartile deviation for the following frequency distribution.

Class interval	5–6	6–7	7–8	8–9	9–10	10–11
Frequency	40	56	60	96	84	68

Age	Frequency	Cumulative frequency
5–6	40	40
6–7	56	$96(m_1)$
7–8	60 (f_1)	156 Q_1, Class
8–9	96	$252(m_3)$
9–10	84 (f_3)	336 Q_3, Class
10–11	68	$404 = N$

Here $N = 404$; 25% of $N = N/4 = 101$ and 75% of $N = 3N/4 = 303$ (where $N = 404$) Q_1 class $= 7 - 8$

$$l_1 = 7, m_1 = 96,\ f_1 = 60 \text{ and } C = 1$$

$$Q_1 = l_1 + \left[\frac{n/4 - m_1}{f_1}\right] \times C$$

$$Q_1 = 7 + \left[\frac{101 - 96}{60}\right] \times 1 = 7 + 0.08 = 7.08$$

$$Q_3 = l_3 + \left[\frac{3n/4 - m_3}{f_3}\right] \times C_3$$

$$Q_3 \text{ Class} = 9 - 10$$

$$Q_3 = \frac{9 + (303 - 252)}{84} \times 1 = 9 + 0.607$$

$$Q_3 = 9.607$$

$$\therefore \text{ Quartile deviation } Q = \frac{Q_3 - Q_1}{2} = \frac{9.607 - 7.08}{2}$$

$$\text{Q.D.} = \frac{2.527}{2} = 1.263$$

VARIANCE AND STANDARD DEVIATION

One of the important measures of variation is that of variance, which indicates how the different values of the data are scattered away from the centre of the distribution. It is usually denoted by the symbol σ^2. The positive square root of variance is called the standard deviation and is denoted by σ.

$$\text{Variance} = \sigma^2 = \frac{\sum (x - \bar{x})^2}{n - 1} \quad \text{or} \quad \frac{\sum d^2}{n - 1}$$

$$\text{Standard deviation} = \sqrt{\text{Variance}} = \sqrt{\frac{\sum d^2}{n - 1}}$$

Example 2.16 *Find the standard deviation of the following scores 3, 5, 7, 9, 11 and 13.*

$$\text{Mean} = \frac{3 + 5 + 7 + 9 + 11 + 13}{6} = 8$$

$$\sum (x - \bar{x})^2 = (3 - 8)^2 + (5 - 8)^2 + (9 - 8)^2 + (11 - 8)^2 + (13 - 8)^2$$

different units. For example, we may wish to know, for a certain population, whether serum cholesterol levels, measured in mg per 100 ml, are more variable than body weight, measured in pounds.

What is needed in situations like these is a measure of relative variation rather than absolute variation. Such a measure is found in the coefficient of variation which expresses the standard deviation as a percentage of the mean. The formula is given by

$$\text{C.V.} = \frac{\text{S.D}}{\bar{x}} \times 100$$

We see that, since the mean and standard deviation are expressed in the same unit of measurement, the unit of measurement cancels out in computing the coefficient of variation. What we have, then, is a measure that is independent of the unit of measurement.

	Sample 1	Sample 2
Age	25 years	11 years
Mean Weight	145 pounds	80 pounds
Standard Deviation	10 pounds	10 pounds

A comparison of the standard deviations might lead one to conclude that the two samples possess equal variability. If we compute the coefficients of variation, however, we have for the twenty-five years old.

$$\text{C.V.} = \frac{10}{145} \times 100 = 6.9$$

and for the eleven-year old

$$\text{C.V.} = \frac{10}{80} \times 100 = 12.5$$

If we compare these results, we get quite a different impression.

Example 2.19 *Calculate Mean, Median, Mode of the following data*

CI	0–10	10–20	20–30	30–40	40–50	50–60
f	12	14	16	28	10	8

CI	**Midpoint** x	**Frequency** (f)	**CF**	fx
0–10	5	12	12	60
10–20	15	14	26	210
20–30	25	$16 - f_1$	$42m$	400
30–40	35	$28 - f_0$	70	980
40–50	45	$10 - f_2$	80	450
50–60	55	8	$88n$	440
		88		2540

$$\text{Mean} = \frac{\sum fx}{N} = \frac{2540}{88} = 28.86 \quad (\text{where } N = \sum f)$$

$$\text{Median} = L + \frac{(n/2) - m}{f} \times c \quad (n/2 = 88/2 = 44)$$

Check under 'cf' which is same as $n/2$ or just above, i.e., = 70; against that is the median class = 30–40.

$$l = 30, \quad m = 42, \quad C = 10, \quad \frac{n}{2} = \frac{88}{2} \quad f = 28$$

$$= 30 + \frac{(88/2) - 42}{28} \times 10$$

$$= 30 + \frac{44 - 42}{28} \times 10$$

$$= 30 + \frac{2}{28} \times 10$$

$$= 30 + 0.07142 \times 10$$

$$= 30 + 0.7142 = 30.7142$$

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$$

L = Lower limit of modal class

$$\Delta_1 = f_0 - f_1 = 28 - 16 = 12$$

$$\Delta_2 = f_0 - f_2 = 28 - 10 = 18$$

C = (Class interval)=10. The highest among frequency=28.

Against that modal class = 30 − 40

$$= 30 + \frac{12}{12 + 18} \times 10$$

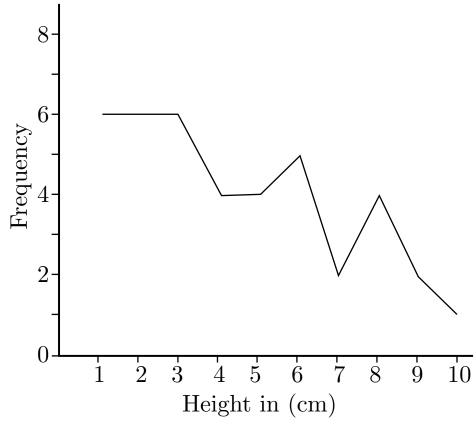
$$= 30 + \frac{12}{30} \times 10$$

$$= 30 + 4 = 34$$

Construct Frequency Distribution Table

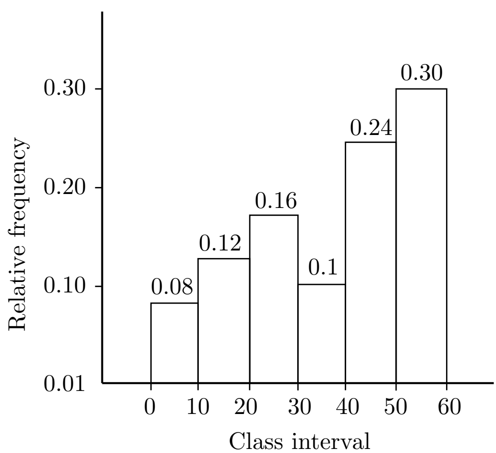
Example 2.20 *Draw Histogram*

58	61	80	65	76	64	82	78	82	74	74	80	72	72
47	88	60	95	56	80	72	86	86	74	88	70	74	90
70	87	76	74	94	84	76	74	74	60	92	92	70	61
94	68	77	62	74	60	60	76	76	75	74	74	66	78
84	75	64	68	84	54	70	50	50	90	56	96	90	80



Example 2.22 *Prepare relative frequency histogram*

Class interval	**Frequency**	**Relative frequency**
0-10	4	$4/50 = 0.08$
10-20	6	$6/50 = 0.12$
20-30	8	$8/50 = 0.16$
30-40	5	$5/50 = 0.10$
40-50	12	$12/50 = 0.24$
50-60	15	$15/50 = 0.30$



Example 2.23 *The following table gives the number of yeast cells in 100 squares of a haemocytometer. Calculate standard deviation, variance, coefficient of variation (C.V.) and standard error.*

Class interval	0-2	3-5	6-8	9-11	12-14	15-17
Frequency	6	6	12	40	30	6

Class interval	**Frequency (*f*)**	**Midpoint *X***	***fX***	**Deviation from Mean $f(x-\bar{X})$**	$(x-\bar{x})^2$	$f(x-\bar{X})^2$
0–2	6	1	6	-9	81	486
3–5	6	4	24	-6	36	216
6–8	12	7	84	-3	9	108
9–11	40	10	400	0	0	0
12–14	30	13	390	3	9	270
15–17	6	16	96	6	36	216
	100		1000			1296

$$\bar{X} = \frac{\sum fX}{n} = \frac{1000}{100} = 10 \quad (\text{where } n = \sum f)$$

$$\text{Variance} = S^2 = \frac{\sum f(x-\bar{x})^2}{n-1} \quad \text{or} \quad \frac{\sum fd^2}{n-1} = \frac{1296}{99} = 13.1$$

$$\text{S.D.} = \sqrt{13.1} = 3.6$$

$$\text{S.E.} = \frac{\text{S.D.}}{\sqrt{n}} = \frac{3.6}{\sqrt{100}} = \frac{3.6}{10} = 0.36$$

$$\text{Coefficient of variation} = \frac{\text{S.D.}}{\bar{X}} \times 100$$

$$= \frac{3.6}{10} \times 100 = 36$$

Example 2.24 *Calculate mean for white blood counts: (X 100) 5, 6, 4, 5, 4, 4, 8, 4*

$$\sum X_i = 5+6+4+5+4+4+8+4 = 40$$

$$n = 8; \quad \bar{X} = 40/8 = 5$$

Example 2.25 *Weight of 11 tables removed from box for quality control purpose is given below. Calculate median*

251, 231, 245, 250, 250, 251, 255, 260, 265, 275, 300

(Hint: Arrange data in ascending order)

$$\text{Median} = n + 1/2$$
$$= 11 + 1/2 = 12/2$$
$$= 251 \text{ (this is the 6th position value)}$$

Example 2.26 *Concentration of a therapeutic agent in vials of commercially available products given below is mg/ml. Calculate mode.*

Vial no.	1	2	3	4	5	6	7	8	9	10
Cone. (mg/ml)	200	205	205	201	199	199	202	205	205	206

The most common value in the above data is 205 mg/ml. Therefore, the mode is 205.

Example 2.27 *Find mode 20, 22, 23, 25, 25, 25, 28, 29, 30, 31, 31, 31, 31, 31, 33.*

Mode is 25, 31. Two modes exist, 31 is major mode and 25 is minor mode.

Example 2.28 *Calculate mode for the following data:*

Class interval	0–l0	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100
Frequency	4	8	10	15	25	20	18	16	18	10

$$\begin{aligned}\text{Mode} &= L + \frac{\blacktriangle 1}{\blacktriangle 1 + \blacktriangle 2} \times C \quad \text{Mode class} = 40 - 50\\ &= 40 + 10/10 + 5 \times 10\\ &= 40 + 20/3 = 46.66\end{aligned}$$

Example 2.29 *Calculate mean, median and mode.*

CI	***X***	***f***	***fX***	**Cumulative frequency**
0–5	2.5	4	10	4
5–10	7.5	6	45	10
10–15	12.5	8	100	18
15–20	17.5	12	210	30
20–25	22.5	14	315	44 m
25–30	27.5	16	440	60
30–35	32.5	10	325	70
35–40	37.5	9	337.5	79
40–45	42.5	8	340	87
45–50	47.5	7	332.5	94
50–55	52.5	6	315	100
55–60	57.5	10	575	110
60–65	62.5	5	312.5	115
65–70	67.5	4	270	119 N
			$\sum fX = 3927.5$	

$$\text{Mean} = \sum fX/n = 3927.5/119 = 33$$

$$\begin{aligned}\text{Median} &= L + (n/2 - m)/f \times c\\ &= 25 + 59.5 - 44/16 \times 5\\ &= 29.84\end{aligned}$$

$$\begin{aligned}\text{Mode} &= L + \blacktriangle 1/\blacktriangle 1 + \blacktriangle 2 \times C\\ &= 25 + 2/8 \times 5\\ &= 25 + 1.25\\ &= 26.25\end{aligned}$$

Example 2.30 *Calculate harmonic mean*

Class interval	0–4	4–8	8–12	12–16	16–20	20–24
Frequency	3	4	6	5	5	8

CI	f	x	$1/x$	$f(1/x)$
0–4	3	2	0.5	1.5
4–8	4	6	0.16	0.64
8–12	6	10	0.1	0.6
12–16	5	14	0.07	0.35
16–20	5	18	0.05	0.25
20–24	8	22	0.04	0.32
	31			$\sum f(l/x) = 3.66$

$$\begin{aligned}\text{H.M.} &= N/f(1/x)\\ &= 31/3.66\\ &= 8.47\end{aligned}$$

Example 2.31 *Calculate mean and mode for the following data*

Class interval	X	f	fX
0–5	2.5	4	10.0
5–10	7.5	3	22.5
10–15	12.5	2	25.0
15–20	17.5	8	140.0
20–25	22.5	6	135.0
25–30	27.5	12	330.0
30–35	32.5	4	130.0
35–40	37.5	6	225.0
40–45	42.5	8	340.0
45–50	47.5	6	285.0
50–55	52.5	8	420.0
55–60	57.5	2	115.0
		$\sum f = 69$	$\sum fX = 2177.5$

$$\text{Mean} = \sum fx/n = 2177.5/69 = 31.56$$

$$\begin{aligned}\text{Mode} &= L + \blacktriangle 1/\blacktriangle 1 + \blacktriangle 2 \times C\\ &= 25 + 6/6 + 8 \times 5\\ &= 25 + 6/14 \times 5\\ &= 25 + 15/7\\ &= 25 + 2.14\\ &= 27.14\end{aligned}$$

Mode class = 25–30

Example 2.36 *Calculate the mean deviation from the following data.*

CI	X	f	fX	|d|	f |d|
0–10	5	5	25	37	185
10–20	15	8	120	27	216
20–30	25	12	300	17	204
30–40	35	15	525	07	105
40–50	45	20	900	03	60
50–60	55	14	770	13	182
60–70	65	12	780	23	276
70–80	75	6	450	33	198
		$\sum f = 92$	$\sum fX = 3870$		$\sum f\|d\| = 1426$

$$\begin{aligned} \text{Mean} &= 3870/92 = 42 \\ \text{Mean deviation} &= \Sigma f|d|/(n-1) \\ &= 1426/(92-1) \\ &= 1426/91 = 15.67 \end{aligned}$$

Example 2.37 *Calculate the median, mode and variance from the following data.*

CI	X	f	fX	CF	d	d^2	fd^2
0–10	5	5	25	5	37	1369	6845
10–20	15	8	120	13	27	729	5832
20–30	25	12	300	25	17	289	3468
30–40	35	15	525	40m	07	49	735
40–50	45	20	900	60	03	9	180
50–60	55	14	770	74	13	169	2366
60–70	65	12	780	86	23	529	6348
70–80	75	6	450	92N	33	1089	6534
		$\sum f = 92$	$\sum fX = 3870$				$\sum fd^2 = 32308$

$$\begin{aligned} \text{Mean} &= 3870/92 = 42 \\ \text{Median} &= L + n/2 - m/f \times c \\ &= 40 + 46 - 40/20 \times 10 \\ &= 40 + 6/20 \times 10 = 43 \end{aligned}$$

$$\begin{aligned} \text{Variance} &= \sum fd^2/n - 1 \\ &= 32308/92 - 1 \\ &= 32308/91 = 355.03 \end{aligned}$$

$$\begin{aligned} \text{Mode} &= 40 + 5/(5+6) \times 10 \\ &= 40 + 0.45 \times 10 \\ &= 40 + 4.5 = 44.5 \end{aligned}$$

$$\text{Median} = 140 + 140/2 = 140$$
$$\text{Mean deviation} = \Sigma\,|d|\,/N = 300/10 = 30$$
$$\begin{aligned}\text{Standard deviation} &= \sqrt{26600/10 - 1}\\ &= \sqrt{26600/9}\\ &= \sqrt{2955.5} = 54.36\end{aligned}$$

Mode is 140.

Example 2.40 *Following is the information on the distribution obtained for the character plant height (cm) in castor in the control data. Calculate mean, mode.*

CI	X	f	fX
30–45	37.5	1	37.5
45–60	52.5	8	420.0
60–75	67.5	26	1755.0
75–90	82.5	31	2557.5
90–105	97.5	22	2145.0
105–120	112.5	17	1912.5
120–135	127.5	14	1785.0
135–150	142.5	10	1425.0
150–165	157.5	7	1102.5
165–180	172.5	5	862.5
180–195	187.5	5	937.5
195–210	202.5	4	810.0
		$\sum f = 150$	$\sum fX = 15750$

$$\begin{aligned}\text{Mean} &= 15750/150 = 105\\ \text{Mode} &= 75 + 5/5 + 9 \times 15\\ &= 75 + 5/14 \times 15\\ &= 75 + 0.35 \times 15\\ &= 75 + 5.25 = 80.25\end{aligned}$$

Example 2.41 *Following is the information on the distribution obtained for the character no. of capsules per main raceme in castor in the control data. Calculate C.V.*

CI	X	f	fX	d	d^2	fd^2
0–4	2	6	12	−9.5	90.25	541.50
5–9	7	49	343	−4.5	20.25	992.25
10–14	12	92	1104	0.5	0.25	23.00
15–19	17	22	374	5.5	30.25	665.50
20–24	22	10	220	10.5	110.25	1102.50
25–29	27	1	27	15.5	240.25	240.25
		$\sum f = 180$	$\sum fX = 2080$			$\sum fd^2 = 3565$

$$\begin{aligned}\text{Mean} &= 2080/180 = 11.5\\ \text{S.D.} &= \sqrt{3565/179}\\ &= \sqrt{19.91}\\ &= 4.4627\end{aligned}$$

$$\begin{aligned}\text{C.V.} &= 4.4627/11.5 \times 100\\ &= 0.3880 \times 100\\ &= 38.80\end{aligned}$$

Example 2.42 *Calculate mean, variance, S.D., S.E. for the following data.*

No. of days	X	f	fX	d	d^2	fd^2
130–132	131	6	786	−10	100	600
133–135	134	15	2010	−7	49	735
136–138	137	35	4795	−4	16	560
139–141	140	27	3780	−1	1	27
142–144	143	25	3575	2	4	100
145–147	146	24	3504	5	25	600
148–150	149	9	1341	8	64	576
151–153	152	6	912	11	121	726
154–156	155	3	465	14	196	588
		$\sum f = 150$	$\sum fX = 21168$			$\sum fd^2 = 4512$

$$\begin{aligned}\text{Mean} &= 21168/150 = 141 \text{ (Approx.)}\\ \text{Variance} &= \Sigma fd^2/n - 1\\ &= 4512/149\\ &= 30.28\end{aligned}$$

$$\begin{aligned}\text{S.D.} &= \sqrt{30.28}\\ &= 5.5\\ \text{S.E.} &= \text{S.D}/\sqrt{n}\\ &= 5.5/\sqrt{150} = 0.45\end{aligned}$$

Example 2.43 *Calculate mean, variance, S.D., S.E.*

No. of days	X	f	fX	d	d^2	fd^2
0–5	2.5	10	25	9.52	90.25	902.5
5–10	7.5	51	382.51	4.52	20.43	1041.9
10–15	12.5	48	600	0.48	0.23	11.0
15–20	17.5	26	455	5.48	30.03	780.7
20–25	22.5	9	202.5	10.48	109.83	988.4
25–30	27.5	3	82.5	15.48	239.63	718.8
30–35	32.5	1	32.5	20.48	419.43	419.4
35–40	37.5	0	0	25.48	647.23	0
40–45	42.5	0	0	30.48	929.03	0
45–50	47.5	0	0	35.48	1258.83	0
		$\sum f = 148$	$\sum fX = 1780$			$\sum fd^2$ =4862.7

I/b	1.76	1.68	1.27	1.60	1.33	1.75	1.77	1.71	1.60	1.53

Mean $I/b = 16/10 = 1.6$

Example 2.46 *Find the harmonic mean of 4,8,10,12,14,16, and 18.*

X	***I/X***
4	0.25
8	0.125
10	0.1
12	0.0833
14	0.0714
16	0.0625
18	0.055
	$\sum I/X = 0.7469$

$$\begin{aligned} \text{H.M.} &= N/\sum I/X \\ &= 7/0.7469 \\ &= 9.37 \end{aligned}$$

Example 2.47 *AIDS is due to Retroviridae. The size in (nm) is given below. Calculate G.M.*

CI	x	f	**log** x	f **log** x
80–85	82.5	4	1.91	7.64
85–90	87.5	6	1.94	11.64
90–95	92.5	10	1.96	19.6
95–100	97.5	4	1.98	7.92
				$\sum f \log x = 46.8$

$$\begin{aligned} \text{G.M.} &= \text{Antilog } 46.8/24 = 1.95 \\ &= \text{Antilog } 1.95 = 90.15 \end{aligned}$$

Example 2.48 *Root length (cm) in safflower is given below. Calculate mean, S.E.*

CI	***X***	***f***	***fX***	***d***	d^2	fd^2
5.6–5.8	5.7	8	45.6	0.41	0.16	1.34
5.8–6.0	5.9	12	70.8	0.21	0.04	0.52
6.0–6.2	6.1	8	48.8	0.01	0.0001	0.0008
6.2–6.4	6.3	20	126	0.19	0.03	0.72
6.4–6.6	6.5	6	39	0.39	0.15	0.91
		$\sum f = 54$	$\sum fX = 330.2$			$\sum fd^2 = 3.4908$

$$\text{Mean} = 330.2/54 = 6.11$$

$$\text{S.D.} = \sqrt{3.49/53} = \sqrt{0.06} = 0.244$$

$$\text{S.E.} = 0.244/\sqrt{54} = 0.244/7.34 = 0.034$$

Example 2.49 *Calculate the mean temperature and mean absorbance for the following.*

Temperature (C)	10	20	30	40	50	60
Absorbance (A 420)	0.2	0.3	0.3	0.35	0.37	0.39

Mean Temperature = 210/6 = 35
Mean Absorbance = 1.91/6 = 0.32

Example 2.50 *The no. of E. coli (X 104) observed every 40 min is given below 60, 80, 100,125,145. Calculate G.M.*

X	log X
60	1.778
80	1.903
100	2.000
125	2.096
145	2.161
	log X = 9.938

G.M = Antilog 9.938 = 97.185

EXERCISE

1. Calculate the Mean deviation from the following data.

(B.Com., Andhra Univ.; 1986)

Class	**Frequency**
0–10	5
10–20	8
20–30	12
30–40	15
40–50	20
50–60	14
60–70	12
70–80	6

Also calculate (a) Median (b) Mode (c) Variance.

2. Compute Quartile Deviation from the following data.

   (B.Com., Kakatiya Univ.; 1987)

X	10–20	20–30	30–40	40–50	50–60	60–70
f	12	19	5	10	9	6

3. From the results given below calculate Mean, Standard Deviation, Variance and Coefficient of Variation (C.V.)

No. of seeds per pod	**Frequency**
2	4
3	2
4	21
5	18
6	4
7	10
8	1

4. The following measurements of weight (in grams) have been recorded for a common strain of rats.
   (a) Choose appropriate class intervals and group the data into a frequency distribution;
   (b) Calculate the relative frequency of each class interval;
   (c) Plot the relative frequency histogram.

100	112	128	126	104	124
104	114	126	100	108	122
106	116	132	100	106	126
108	118	140	126	112	128
110	120	130	128	118	130

5. Calculate the Mean, Median, Variance, Standard Deviation and Range for each of the following sets.
   (a) 5, 10, 15, 20, 25
   (b) 2, 4, 2, 2, 6
   (c) 4, 6, 8, 10
   (d) −2, 1, −1, 0, 4, −2, −3

6. Blood cholesterol levels mg/ml were recorded in a survey.

150	160	140	185	135
180	170	145	195	155
190	130	150	165	125
200	140	155	175	120

   (a) Group the data into a frequency distribution.
   (b) Compute mean and standard deviation from the ungrouped data.
   (c) Compute the mean and standard deviation from the grouped data.

(c) Median
(d) Standard Deviation
(e) Mean Deviation

12. The table shows a distribution of bristle number in *Drosophoila*,

Bristle number	**No. of individuals**
1	2
2	3
3	8
4	30
5	50
6	16
7	6

Calculate Mean, Variance, S.D. of the distribution.

13. Following is the information on the distribution obtained for the character plant height(cm) in castor in the control data.
(a) Calculate Mean, Mode
(b) Calculate S.D., C.V., Variance, S.E.

Height in cm	30–45	45–60	60–75	75–90	90–105	105–120
Frequency	1	8	26	31	22	17

Height in cm	120–135	135–150	150–165	165–180	180–195	195–210
Frequency	14	10	7	5	5	4

14. Following is the information on the distribution obtained for the character number of capsules per main raceme in castor in the control data. Calculate C.V.

Number of capsules per main raceme	0–4	5–9	10–14	15–19	20–24	25–29
Frequency	6	49	62	22	10	1

15. Following is the information on the distribution obtained for the character number of days to maturity in castor in the control data.
Calculate Mean, Variance, S.D. and S.E.

Number of days to maturity	130–132	133–135	136–138	139–141	142–144
Frequency	6	15	35	27	25

Number of days to maturity	145–147	148–150	151–153	154–156
Frequency	24	9	6	3

16. Following is the information on the distribution obtained for the character number of capsules/main raceme in castor for different treatments. Calculate Mean, Variance, S.D. and S.E.

(a)

Class interval	0–5	5–10	10–15	15–20	20–25	25–30	30–35
Frequency 0.1% Hz	10	51	48	26	9	3	1

Class interval	35–40	40–45	45–50
Frequency 0.1% Hz	0	0	0

(b)

Class interval	0–5	5–10	10–15	15–20	20–25	25–30	30–35
Frequency 30 *Kr* + 0.1% Hz	7	65	46	21	3	6	1

Class interval	35–40	40–45	45–50	50–55
Frequency 30 *Kr* + 0.1% Hz	0	0	0	1

17. Following is the information on the distribution obtained for the character number of days to maturity in castor for different treatments. Calculate Mean, Median, Mode.

(a)

CI	125–130	130–135	135–140	140–145	145–150	150–155
Frequency 30 *Kr*	6	31	41	47	18	7

(b)

CI	125–130	130–135	135–140	140–145	145–150	150–155
Frequency 0.1% Hz	7	32	46	30	22	12

24. Effect of temperature on Chitinase activity was tested in TMV2 groundnut by incubating the enzymatic reaction mixture in a water bath ranging from 10 ℃–60 ℃ and measuring the absorbance.
Calculate Mean Absorbance for Mean Temperature.

Temperature (0 ℃)	10	20	30	40	50	60
Absorbance (A-420)	0.2	0.3	0.3	0.35	0.37	0.39

25. Distribution of persons as per Hb

Grams of Hb per 100 CC blood	**Number**	**Grams of Hb per 100 CC blood**	**Number**
8–8.5	2	10.0–10.5	12
8.5–9.0	1	10.5–11.0	14
9.0–9.5	3	11.0–11.5	10
9.5–10.0	5	11.5–12.0	1

Calculate Mean, Standard Deviation, S.E.

26. Draw frequency polygon

Total Serum Protein g/100 ml	**No. of Females**
6–7	2
7–8	4
8–9	12
9–10	10
10–11	2

27. The number of *E. coli* ($\times 10^4$) observed every 40 minutes is given below:

    60, 80, 100, 125, 145

    Calculate Geometric Mean.

28. Serum creatine kinase levels in the blood is provided

CI	**Data Frequency**	**CI**	**Data Frequency**
0-10	6	30-40	14
10-20	4	40-50	2
20-30	12	50-60	4

Calculate Standard Deviation and C.V.

29. Trigonella leaflet data is given

Length (mm)	**Breadth (mm)**
2.3	1.3
2.7	1.6
2.3	1.8
2.4	1.5
2.0	1.5
2.1	1.2
1.6	0.9
2.4	1.4
2.4	1.5
2.0	1.3

Calculate mean length, mean breadth, data and mean for l/b.

30. Consider the following data:

Tumour size (cm)	**No. of patients**
1–2	5
2–3	6
3–4	8
4–5	4

Calculate Mean.

31. After ingestion of drug, the excretion % of the same in urine samples is given after certain length of time.

% excreted (mg)	**Frequency**
10–20	3
20–30	4
30–40	2
40–50	5
50–60	6

Calculate Mean excretion.

32. IQ score are given below:

Mid value	**Frequency**	**Mid value**	**Frequency**
72	4	114	14
76	6	116	10
80	8	118	8
88	30	120	4
92	35	122	2
98	40	126	1
102	10		

Calculate variance.

Height in inches (CI)	**Frequency**	**Height in inches (CI)**	**Frequency**
57–59	15	67–69	15
59–61	18	69–71	8
61–63	35	71–73	6
63–65	20	73–75	2
65–67	10		

Calculate S.D., Variance and Standard error.

39. Dysentery is caused by the virus family. The double stranded RNA mass (kbp = Kilobase pairs) are given below:

    18, 20, 22, 24, 26, 28, 30
    Calculate Harmonic mean.

40. AIDS is due to retroviridae. The size (nm) is given below. Calculate Geometric mean (G.M.)

CI(nm)	**Frequency**
80–85	4
85–90	6
90–95	10
95–100	4

41. Out of some of the main virus families infecting humans, the family adenoviridae causing common cold has the following size measurements (nm) collected among 20 samples. Prepare frequency distribution.

70.0	74.0	79.0	84.0	88.0	72.0	78.0	80.0	85.0	88.0
72.0	79.0	81.0	86.0	90.0	74.0	79.0	84.0	86.0	90.0

42. Hepatitis B is caused by the family of viruses. The genome as DNA shows following kilobase pairs (kbp) of double stranded DNA among 10 samples.

    1.7, 1.9, 2.0, 2.4, 2.4, 2.1, 2.2, 2.3, 2.6, 2.8

    Calculate Standard Deviation, Range, Variance, Standard error and C.V.

43. Influenza is caused by the family of virus whose frequency distribution of size (nm) is given below. Calculate Mean, Median and Mode.

CI(nm)	**Frequency**
90–95	2
95–100	2
100–105	12
105–110	3
110–115	2
115–120	1

Samples no.	Under stress	Under relaxed condition
1	142.8	142.0
2	147.1	145.2
3	134.5	132.0
4	129.6	128.0
5	133.0	131.0
6	134.0	132.0

49. Compute Mean and Mode for the given data:

Diameter (in mm)	0-4	4-8	8-12	12-16	16-20
No. of bearings	4	6	14	6	4

50. Compute Standard error for the following data:

    6, 9, 11, 8, 14, 3, 20, 12

51. Six bushes of flowering plant gave the following data with reference to no. of buds:

    60, 40, 80, 50, 30, 64

    Calculate Standard deviation 'S' by short cut method.

    $$Hint : S^2 = \left[\frac{\Sigma x_i^2 - (\Sigma x_i)^{2/n}}{n-1}\right]$$

52. Protein in gms/litre is giving below. Calculate Variance, Coefficient of Variation.

10.4	10.6	11.2	12.1	12.2	12.4
11.2	12.1	13.4	12.8	12.6	12.6
13.2	13.4	12.9	13.6	13.0	11.2
14.2	14.6	14.3	14.4	14.6	14.5
15.4	15.0	16.2	16.1	16.6	16.8

53. Haemoglobin (g/dl) data is given below:

    3.8, 2.6, 3.2, 4.8, 3.9, 2.8, 3.6, 6.2

    Find Mean deviation.

54. Albumin (g/dl) data is given below:

CI	Frequency
3.2-3.4	5
3.4-3.6	6
3.6-3.8	8
3.8-4.0	4
4.0-4.2	3
4.2-4.4	2

Calculate Mean deviation.

55. Creatinine levels mg/kg body weight are given below:

    23, 24, 25, 26, 28, 30, 32, 34, 36.

    Calculate Standard error.

56. Phosphatase levels (mg/litre) are given:

CI	**Frequency**	**CI**	**Frequency**
180–200	8	260–280	4
200–220	12	280–300	6
220–240	18	300–320	5
240–260	10		

    Draw histogram and frequency polygon.

57. Total aflatoxin levels are given below in (mg/l). On days 6, 9 and 12 after infection with Aspergillus in groundnut.

No. of days	**Amount**
6	8.2
9	9.4
12	10.6

    Draw Frequency polygon.

Review

I. The temperatures
In degrees Fahrenheit are given below.
Prepare frequency distribution—Draw histogram, frequency polygon.

68, 69, 68, 70, 72, 74, 74, 76, 75, 77, 80, 82, 84, 86, 88, 89, 88, 90, 91, 93, 94, 95, 97, 100, 106, 86, 103, 68, 103, 106, 81, 97

(Hint prepare Class intervals 68–72, 72–76, ...)

Consider the following data:

No. of stamens	**Frequency**
6–8	5
8–10	3
10–12	9
12–14	2
14–16	3
16–18	4

Calculate Standard error and C.V.

Calculate relative frequency and cumulative frequency.

CI	Frequency
3–5	4
5–7	6
7–9	15
9–11	14
11-13	12
13-15	10
15-17	9
17-19	14
19-21	16

II. Define:
Variable, Discrete variable, Continuous variable, Unimodal, Bimodal, Range, Frequency distribution, Histogram, Frequency polygon, Quartile deviation, Geometric mean, Harmonic mean, Standard error, Coefficient of variation (C.V.)

Skewness: Departure of a frequency distribution from symmetry.

Kurtosis: The shape of the vertex or peak of the curve.

III. (a) A graph of a cumulative frequency distribution is ....

(b) $\frac{S}{\sqrt{n}}$ is ....

(c) $S^2 =$ ....

(d) $\sum$ (Sigma) means ....

(e) $\sum_{i=1}^{n} \frac{x_i}{n}$ is ....

(f) $\sum_{i=1}^{N} \frac{x_i}{N}$ is ....

(g) $Q_3 - Q_1$ is ....

(h) $S =$ ....

(i) $\sqrt[n]{X_1 \times X_2 \times X_3 \times \ldots X_n} =$ ....

(j) 6, 8, 12, 14, 16 and .... gives the mean value 10.

(k) Standard error = 2, Standard deviation = 10
then $n =$ ....

(l) If frequency = 4, 6, 5, 2, 8, 7, give the cumulative frequency values

(m) The Median of the data 12, 6, 7, 9, 11

(n) The Median of the data 12, 4, 44, 120, 624, 60