Measures of Central Tendency

SESSION 6

Structure

6.1 INTRODUCTION

In Lab Session 2, you have learnt how to classify raw data and obtain discrete and continuous frequency distributions in order to extract maximum information from the data. In Unit 1 of MST-002 (Descriptive Statistics), we have explained various measures of central tendency to represent the entire data in a single average value.

In this lab session, you will learn how to compute different measures of central tendency, such as arithmetic mean, geometric mean, harmonic mean, mode and median for raw data as well as discrete and continuous frequency distributions using MS Excel 2007. You will also learn how to compute partition values, viz. quartiles, deciles and percentiles for raw data as well as discrete and continuous frequency distributions. These measures give a value around which the data are concentrated. It also helps us in the assessment of a distribution as well as in comparing it with other distributions.

Prerequisite

- Lab Sessions 1 and 2 of MSTL-001 (Basic Statistics Lab).
- Unit 1 of MST-002 (Descriptive Statistics).

In the next lab session, we shall explain various measures of dispersion, which give us an idea about the scatter or spread of data.

Objectives

After performing the activities of this session, you should be able to:

- prepare the spreadsheet in MS Excel 2007;
- determine the arithmetic mean, geometric mean, harmonic mean and mode; and
- compute the median, quartiles, deciles and percentiles.

6.2 PROBLEM DESCRIPTION

In this lab session, we state five problems to illustrate how to compute various measures of central tendency.

For this lab session, we consider

1. The data of Problem 2 given in Lab Session 2 to compute the arithmetic mean, median, quartiles, 7^{th} decile, 15^{th} percentile for a discrete frequency distribution.
2. The data of Problem 3 given in Lab Session 2 to compute the arithmetic mean, median, quartiles, 7^{th} decile, 15^{th} percentile for raw data and continuous frequency distribution.
3. A researcher is studying the growth of a particular strain of bacteria. The daily percentage growth of the bacteria is recorded for 60 days. The data are given in Table 1.

Table 1: Growth of bacteria for 60 days

Day	Growth of Bacteria (%)	Day	Growth of Bacteria (%)	Day	Growth of Bacteria (%)
1	1.7	21	20.1	41	55.3
2	1.9	22	27.4	42	55.8
3	2.4	23	28.3	43	56.8
4	2.6	24	28.7	44	57.6
5	3.5	25	33.2	45	59.4
6	3.8	26	33.4	46	61.2
7	4.3	27	34.6	47	62.2
8	4.3	28	37.4	48	62.3
9	4.8	29	38.8	49	62.4
10	9.1	30	39.7	50	64.7
11	9.6	31	40.3	51	65.1
12	10.9	32	40.9	52	65.3
13	11.5	33	43.7	53	65.9
14	11.8	34	43.2	54	66.2
15	14.3	35	44.6	55	66.9
16	14.7	36	45.5	56	67.2
17	16.3	37	49.6	57	67.9
18	16.7	38	50.4	58	68.5
19	17.8	39	53.3	59	68.6
20	19.2	40	55.1	60	68.9

For this data,

- compute the average percentage growth of bacteria,
- construct the frequency distribution, and
- compute the average percentage growth of bacteria for the frequency distribution.

4. A person travels from city A to city B by car. The average speed of the car per kilometre distance travelled from A to B is recorded in Table 2.

Table 2: Average speed of the car per km from cities A to B

S. No.	Average Speed per km (km/h)	S. No.	Average Speed per km (km/h)	S. No.	Average Speed per km (km/h)
1	45	18	50	35	60
2	25	19	60	36	55
3	30	20	45	37	50
4	30	21	55	38	30
5	40	22	35	39	45
6	45	23	30	40	55
7	50	24	25	41	40
8	50	25	25	42	50
9	60	26	35	43	30
10	65	27	40	44	35
11	80	28	45	45	35
12	80	29	45	46	40
13	65	30	55	47	30
14	65	31	60	48	35
15	60	32	75	49	40
16	55	33	75	50	40
17	35	34	60		

For this data,

- compute the average speed of the car during the entire journey from city A to city B,
- construct the discrete and continuous frequency distributions, and
- compute the average speed of the car during the entire journey for both frequency distributions.

5. A shirt manufacturing company conducted a survey among 100 men to know their shirt size. The data are recorded in Table 3.

Table 3: Data for shirt size

S. No.	Size of Shirt	S. No.	Size of Shirt	S. No.	Size of Shirt
1	38	35	42	69	42
2	39	36	40	70	36
3	36	37	41	71	42
4	41	38	43	72	37
5	38	39	39	73	36
6	40	40	43	74	45
7	38	41	37	75	44
8	40	42	42	76	39
9	39	43	34	77	41
10	41	44	40	78	40
11	46	45	41	79	37
12	42	46	42	80	37
13	40	47	41	81	41
14	36	48	39	82	39
15	40	49	40	83	42
16	35	50	35	84	37
17	41	51	45	85	38
18	43	52	34	86	40

S. No.	Size of Shirt	S. No.	Size of Shirt	S. No.	Size of Shirt
19	39	53	40	87	39
20	38	54	38	88	37
21	43	55	37	89	40
22	42	56	42	90	35
23	43	57	40	91	42
24	40	58	43	92	39
25	39	59	41	93	40
26	41	60	40	94	39
27	41	61	38	95	40
28	40	62	36	96	38
29	43	63	35	97	40
30	44	64	39	98	39
31	36	65	44	99	41
32	38	66	38	100	38
33	42	67	37		
34	44	68	41		

For this data,

- compute the most preferable average size of the shirt,
- construct the discrete and continuous frequency distributions, and
- compute the most preferable average shirt size for frequency distributions.

6.3 ARITHMETIC MEAN

In Unit 1 of MST-002, you have learnt how to compute the arithmetic mean for ungrouped (raw) and grouped data. Arithmetic mean is the most commonly used measure of central tendency and often simply called the average or mean. It is calculated taking all observations of the data into account. We use arithmetic mean when data do not show wide fluctuation, i.e., do not have extreme values. We briefly mention the main points and formulae as follows:

The arithmetic mean is defined as the sum of all observations divided by the number of observations.

- **For Ungrouped (Raw) Data**

  The arithmetic mean of n observations, $x_1, x_2, ..., x_n$ is given by

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{... (1)}$$

- **For Discrete Frequency Distribution**

  If the observations $x_1, x_2, ..., x_k$ occur with frequencies $f_1, f_2, ..., f_k$, respectively, the arithmetic mean is given by

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{N} \quad \text{where } N = \sum_{i=1}^{k} f_i. \quad \text{... (2)}$$

- **For Continuous Frequency Distribution**

  If the observations are arranged in the form of class intervals, we take x_i in equation (2) as the mid value of the i^{th} class interval. The formula for computation of the arithmetic mean is the same as given in equation (2).

Let us now describe the computation of arithmetic mean in Excel 2007 for raw data, discrete and continuous frequency distributions, one at a time.

6.3.1 Arithmetic Mean for Raw Data

We first consider Problem 2 and explain how to compute arithmetic mean for the raw data using MS Excel 2007. In order to compute the average life of the electric bulb for the raw data, we follow the steps given below :

Step 1: We enter the data given in Table 3 of Lab Session 2, in Excel 2007 spreadsheet as shown in Fig. 6.1.

	A	B	C
1	S.No.	Life of Bulb (in hours)	
2	1	1087	
3	2	1289	
4	3	876	
5	4	725	
6	5	900	
7	6	1080	
8	7	952	
9	8	741	
10	9	1000	
11	10	900	

Mean_Raw Data

Fig. 6.1: Partial screenshot of the spreadsheet for the given data.

Step 2: We now compute the arithmetic mean, i.e., the average life of the bulbs. For the given raw data, we can use ***Average*** function as shown in Fig. 6.2. For this, we

1. select Cell B152,
2. click on the ***Formulas*** tab, and
3. click on ***More Functions*** → ***Statistical*** → ***Average***.

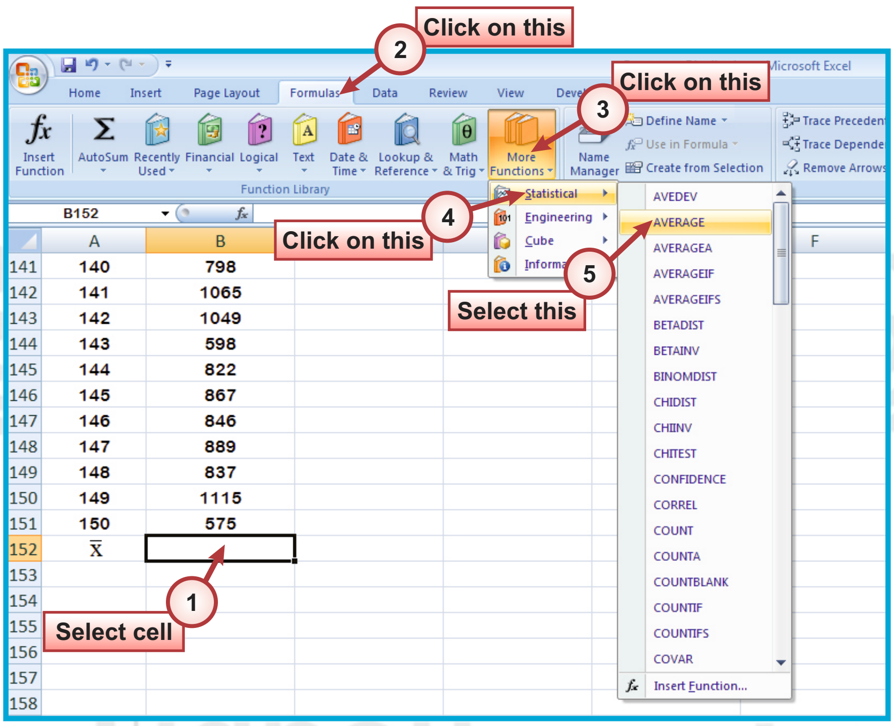


Fig. 6.2

Step 3: A new dialog box opens. We select Cells B2:B151 and click on ***OK*** as shown in Fig. 6.3.

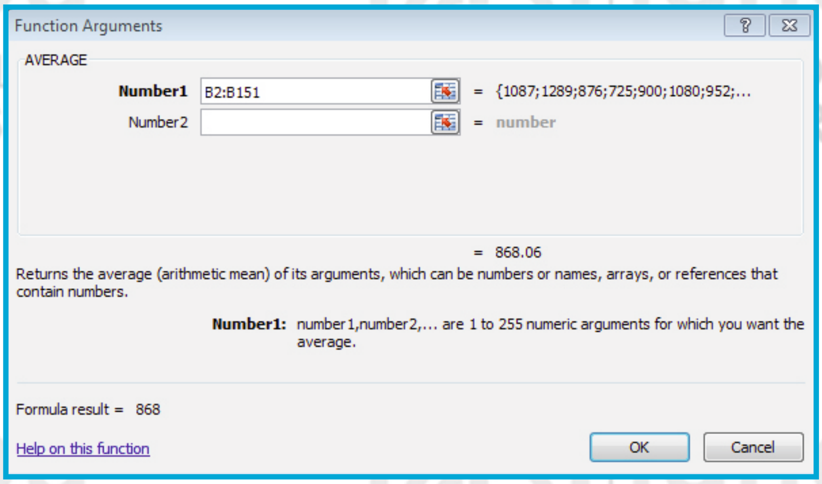


Fig. 6.3

Step 4: We obtain the value of arithmetic mean in Cell B152 as shown in Fig. 6.4.

B152 f_x =AVERAGE(B2:B151)

	A	B	C
151	150	575	
152	$\bar{x}$	868.0600	
153			

Fig. 6.4

You can also try the short-cut method for computing arithmetic mean as explained in Unit 1 of MST-002.

From Fig. 6.4, the arithmetic mean is given by

$$AM = \bar{x} = 868.0600 \text{ hrs}$$

We can say that the average life of the bulb is 868.06 hrs for the given data.

We now explain how to compute the arithmetic mean for discrete frequency distribution using Excel 2007. For this, we consider Problem 1. Note that Excel does not have any built-in function to compute the arithmetic mean for frequency distributions.

6.3.2 Arithmetic Mean for Discrete Frequency Distribution

In order to compute the average number of mobile phones per family for the discrete frequency distribution for Problem 1, we follow the steps given below:

Step 1: For the data given in Table 2 of Lab Session 2, we directly consider the discrete frequency distribution as obtained in Sec. 2.4 of Lab Session 2 (Fig. 6.5).

	A	B	C	D
1	S.No.	Number of Mobile Phones (x)	No. of Families (f)	
2	1	1	6	
3	2	2	16	
4	3	3	22	
5	4	4	24	
6	5	5	22	
7	6	6	16	
8	7	7	9	
9	8	8	5	

Fig. 6.5

Step 2: We type "fx" in Cell D1. We now multiply the first value of the number of mobile phones (x), i.e., 1 given in Cell B2 with the corresponding number of families (f), i.e., 6 given in Cell C2 to obtain the first value of fx in Cell D2. We type "=C2*B2" in Cell D2 as shown in Fig. 6.6a and press ***Enter*** to obtain the output (Fig. 6.6b).

AVERAGE | =C2*B2

	A	B	C	D
1	S.No.	Number of Mobile Phones (x)	No. of Families (f)	fx
2	1	1	6	=C2*B2
3	2	2	16	

(a)

ENTER

	C	D	E
1	No. of Families (f)	fx	
2	6	6	
3	16		

(b)

Fig. 6.6

Step 3: We drag Cell D2 down up to Cell D9 to obtain the remaining values of the product of the number of mobile phones (x) and the number of families (f). The result is as shown in Fig. 6.7.

	A	B	C	D	E
1	S.No.	Number of Mobile Phones (x)	No. of Families (f)	fx	
2	1	1	6	6	
3	2	2	16	32	
4	3	3	22	66	
5	4	4	24	96	
6	5	5	22	110	
7	6	6	16	96	
8	7	7	9	63	
9	8	8	5	40	
10					

Fig. 6.7

Step 4: In Cell C10, we compute the total number of families ($\sum f$) using "=Sum(C2:C9)" function as explained in Step 7 of Sec. 2.3, Lab Session 2 and press ***Enter*** (see Fig. 6.8).

AVERAGE | =SUM(C2:C9)

	A	B	C
1	S.No.	Number of Mobile Phones (x)	No. of Families (f)
2	1	1	6
3	2	2	16
4	3	3	22
5	4	4	24
6	5	5	22
7	6	6	16
8	7	7	9
9	8	8	5
10		Total	=SUM(C2:C9)
11			SUM(number1, [nur

(a)

ENTER

C10 | =SUM(C2:C9)

	A	B	C
1	S.No.	Number of Mobile Phones (x)	No. of Families (f)
2	1	1	6
3	2	2	16
4	3	3	22
5	4	4	24
6	5	5	22
7	6	6	16
8	7	7	9
9	8	8	5
10		Total	120
11			

(b)

Fig. 6.8

Step 5: In the same way, we obtain $\sum fx$ in Cell D10 as shown in Fig. 6.9.

D10 | f_x =SUM(D2:D9)

	B	C	D	E
9	8	5	40	
10	Total	120	509	
11				

Fig. 6.9

Step 6: From equation (2), we compute the arithmetic mean ($\bar{x}$) by typing "=D10/C10" in Cell B11 (Fig. 6.10a) and pressing ***Enter.*** We get the result shown in Fig. 6.10b.

MINVERSE | f_x =D10/C10

	A	B	C	D
9	8	8	5	40
10		Total	120	509
11	$\bar{x}$	=D10/C10		
12				

(a)

B11 | f_x =D10/C10

	A	B	C
9	8	8	5
10		Total	120
11	$\bar{x}$	4.2417	
12			

(b)

Fig. 6.10

Step 7: Since the average number of mobile phones cannot be a fraction, we round off its value. The rounded off average number of mobile phones is 4 (Cell C11, Fig. 6.11).

	A	B	C	D
10		Total	120	509
11	$\bar{x}$	4.2417	≈ 4	
12				

Fig. 6.11

From Fig. 6.11, the arithmetic mean is given by

$$AM = \bar{x} = 4$$

We can say that the average number of mobile phones per family is 4.

6.3.3 Arithmetic Mean for Continuous Frequency Distribution

We now consider Problem 2 and explain how to compute the arithmetic mean of a continuous frequency distribution using MS Excel 2007. In order to compute the average life of bulbs for the continuous frequency distribution, we follow the steps given below:

Step 1: For the data given in Table 3 of Lab Session 2, we take the continuous frequency distribution formed in Sec. 2.5 of Lab Session 2 (Fig. 6.12).

	A	B	C
1	Class Interval (x)	No. of Bulbs (f)	
2	500-600	11	
3	600-700	20	
4	700-800	25	
5	800-900	32	
6	900-1000	27	
7	1000-1100	19	
8	1100-1200	11	
9	1200-1300	5	
10			

Fig. 6.12

Step 2: We compute the mid values (Fig. 6.13) as explained in Steps 4 to 6 of Sec. 4.6 of Lab Session 4.

	A	B	C	D	E
1	Class Interval	No. of Bulbs (f)	Lower Limit	Upper Limit	Mid Value (x)
2	500-600	11	500	600	550
3	600-700	20	600	700	650
4	700-800	25	700	800	750
5	800-900	32	800	900	850
6	900-1000	27	900	1000	950
7	1000-1100	19	1000	1100	1050
8	1100-1200	11	1100	1200	1150
9	1200-1300	5	1200	1300	1250

Fig. 6.13

Step 3: To multiply the mid value (x) and frequency (f) of the first class, we type "=B2*E2" in Cell F2 as shown in Fig. 6.14a and press ***Enter*** to obtain the output (Fig. 6.14b).

AVERAGE fx =B2*E2

	B	C	D	E	F	G
1	No. of Bulbs (f)	Lower Limit	Upper Limit	Mid Value (x)	fx	
2	11	500	600	550	=B2*E2	
3	20	600	700	650		
4	25	700	800	750		

(a)

F2 fx =B2*E2

	B	C	D	E	F	G
1	No. of Bulbs (f)	Lower Limit	Upper Limit	Mid Value (x)	fx	
2	11	500	600	550	6050	
3	20	600	700	650		
4	25	700	800	750		

(b)

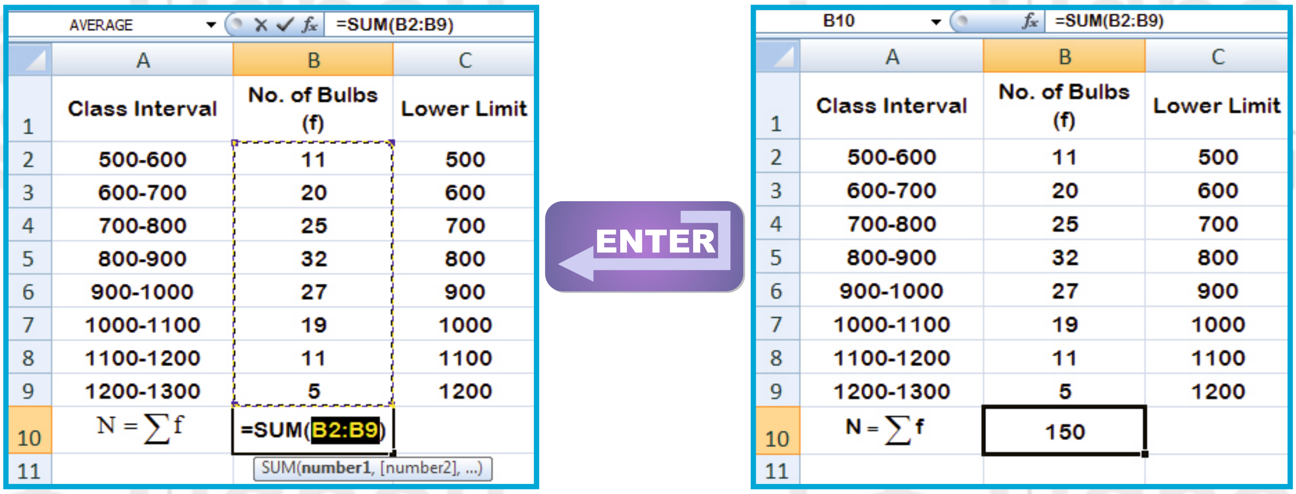
Fig. 6.14

Step 4: We drag Cell F2 down up to Cell F9 to obtain the product of the mid value and frequency for the remaining classes as shown in Fig. 6.15.

	B	C	D	E	F	G
1	No. of Bulbs (f)	Lower Limit	Upper Limit	Mid Value (x)	fx	
2	11	500	600	550	6050	
3	20	600	700	650	13000	
4	25	700	800	750	18750	
5	32	800	900	850	27200	
6	27	900	1000	950	25650	
7	19	1000	1100	1050	19950	
8	11	1100	1200	1150	12650	
9	5	1200	1300	1250	6250	
10						

Fig. 6.15

Step 5: We compute the total number of bulbs, i.e., the sum of the frequency ($\sum f$) in Cell B10 as shown in Fig. 6.16.



(a) (b)

Fig. 6.16

Step 6: In the same way, we calculate $\sum fx$ in Cell F10 as shown in Fig. 6.17.

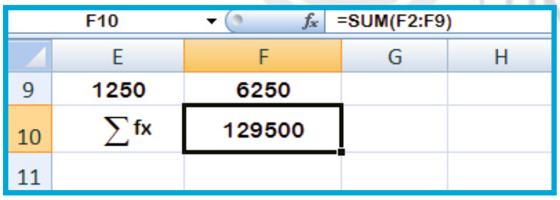


Fig. 6.17

Step 7: To calculate the arithmetic mean, we type "=F10/B10" in Cell F11 (Fig. 6.18a) and press ***Enter*** to obtain the output shown in Fig. 6.18b.

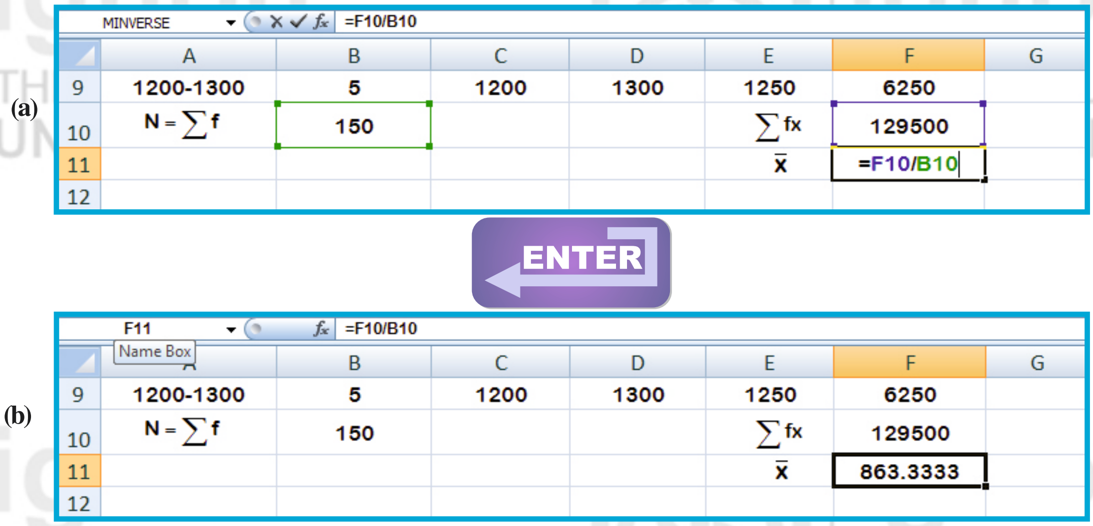


Fig. 6.18

From Fig. 6.18b, the arithmetic mean is given by

$$AM = \bar{x} = 863.3333 \text{ hrs}$$

We can say that the average life of the bulb is 863.33 hrs for the given data.

6.4 GEOMETRIC MEAN

You have learnt in Unit 1 of MST-002 that the geometric mean is commonly used in the computation of average rate or percentage growth. We also use geometric mean where we assign larger weights to small items and smaller weights to large items. We briefly mention the main points for calculating geometric mean as follows:

- **For Ungrouped (Raw) Data**

  If $x_1, x_2, ..., x_n$ are n observations of a variable X, then their geometric mean is given by

$$GM = \sqrt[n]{x_1.x_2...x_n} = (x_1.x_2...x_n)^{\frac{1}{n}}$$

or $$GM = \text{Antilog}\left(\frac{1}{n}\sum_{i=1}^{n}\log x_i\right) \qquad ... (3)$$

- **For Discrete Frequency Distribution**

  If $x_1, x_2, ..., x_k$ are k values of a variable X with frequencies

  $f_1, f_2, ..., f_k$, respectively, then

$$GM = \left(x_1^{f_1}.x_2^{f_2}....x_k^{f_k}\right)^{\frac{1}{N}}$$

where $N = f_1 + f_2 + ... + f_k$

or $$GM = \text{Antilog}\left(\frac{1}{N}\sum_{i=1}^{k}f_i\log x_i\right) \qquad ... (4)$$

- **For Continuous Frequency Distribution**

  If the observations are arranged in the form of class intervals, we consider x_i as the mid values of the i^{th} class interval and use equation (4) to compute the geometric mean.

Here we describe how to compute the geometric mean using Excel 2007 for raw data and continuous frequency distribution, one at a time.

6.4.1 Geometric Mean for Raw Data

We consider Problem 3 to explain the computation of geometric mean for raw data. In order to compute the average growth of bacteria for the given data, we apply the geometric mean (not the arithmetic mean) and follow the steps given below:

Step 1: We enter the data given in Table 1 in an Excel sheet as shown in Fig. 6.19 and name it "**GM**".

	A	B	C
1	Day	Bacteria Growth %	
2	1	1.7	
3	2	1.9	
4	3	2.4	
5	4	2.6	
6	5	3.5	
7	6	3.8	
8	7	4.3	
9	8	4.3	
10	9	4.8	
11	10	9.1	

GM

Fig. 6.19

Step 2: Since the given data shows the percentage growth of bacteria, we consider the initial growth as 100. We compute the growth of the first day by typing "=100+B2" in Cell C2 (Fig. 6.20a) and dragging it down up to Cell C61 to compute the growth of the remaining days as shown in Fig. 6.20b.

(a)

(b)

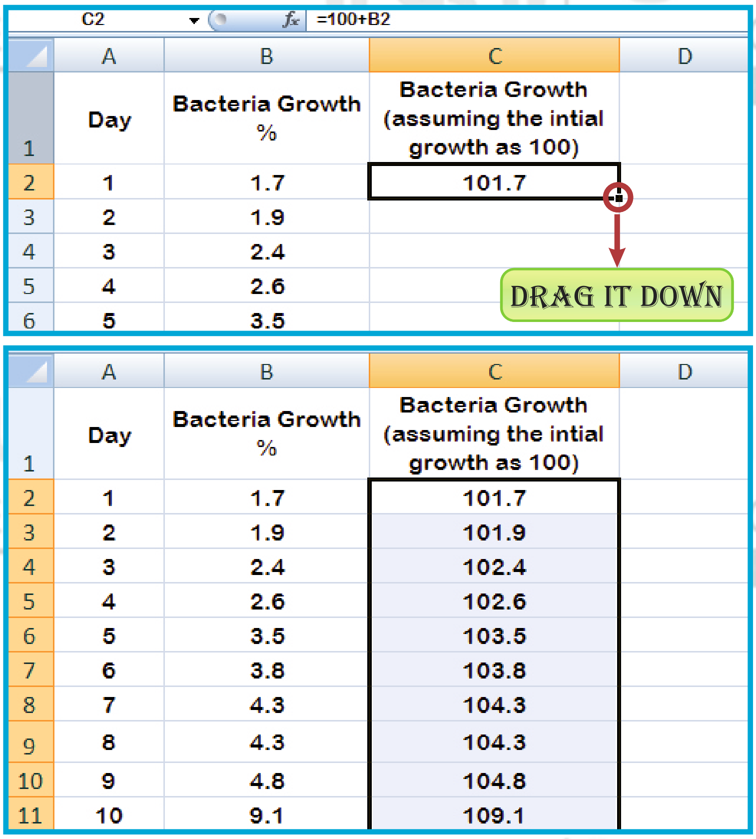


Fig. 6.20

Step 3: To compute the geometric mean for the given raw data, we refer to Fig. 6.21 and

1. select Cell C62,
2. click on the ***Formulas*** tab, and
3. click on ***More Functions*** → ***Statistical*** → ***Geomean***.

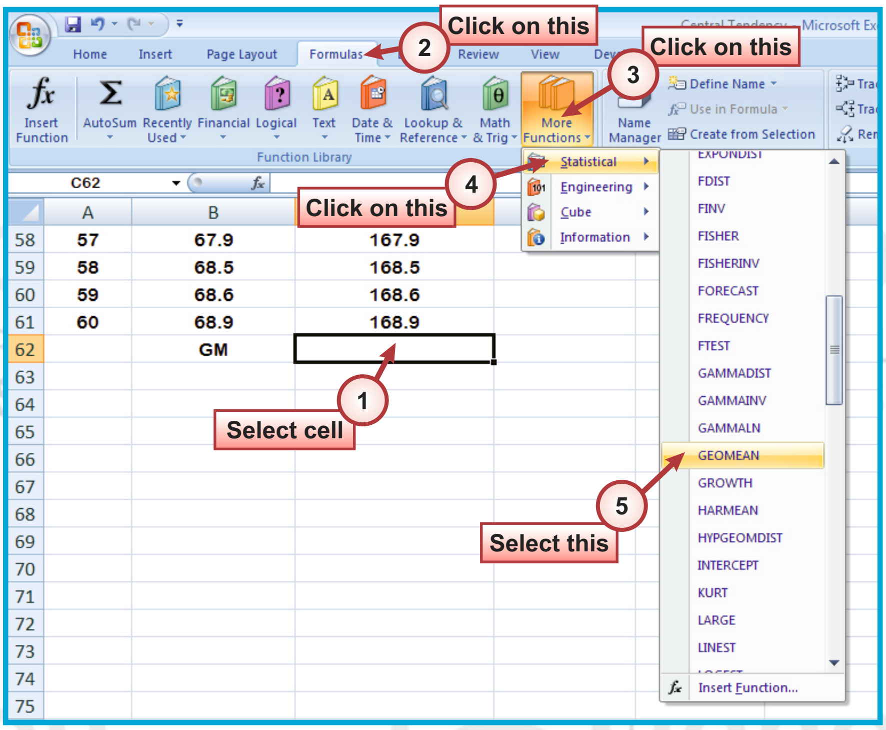


Fig. 6.21

Step 4: The dialog box shown in Fig. 6.22 appears. We select Cell C2:C61 as shown in Fig. 6.22 and click on ***OK***.

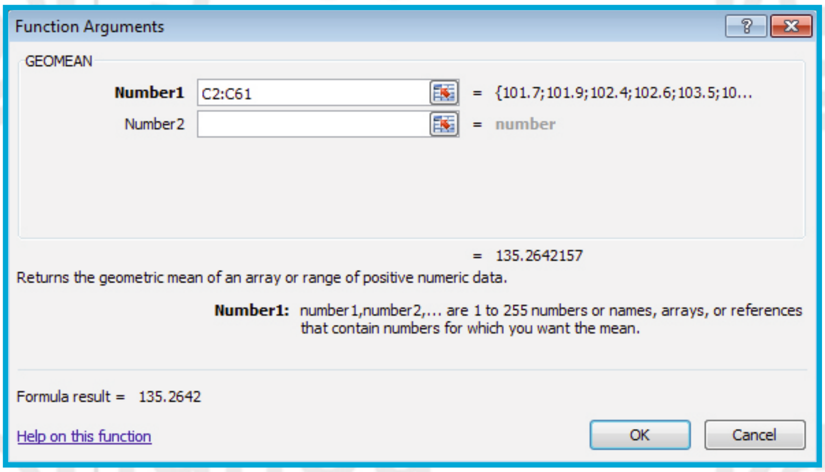


Fig. 6.22

Step 5: The value of the geometric mean is given in Cell C62 (Fig. 6.23).

C62 =GEOMEAN(C2:C61)

	A	B	C	D
61	60	68.9	168.9	
62		GM	135.2642	
63				

Fig. 6.23

From Fig. 6.23, the geometric mean is 135.2642. Therefore, the average percentage growth from day 1 to 60 = 135.2642 – 100 = 35.2642 or approximately 35%.

To understand the computation of the geometric mean for the grouped frequency distribution, we shall construct the frequency distribution for the data given in Table 1. Since this data has distinct values, we cannot form the discrete frequency distribution. So we shall form the continuous frequency distribution. Here we are continuing with the same Excel sheet. Note that Excel does not have any built-in function to compute the geometric mean for frequency distributions.

6.4.2 Geometric Mean for Continuous Frequency Distribution

In order to compute the average bacteria growth for the continuous frequency distribution, we follow the steps given below:

Step 1: We construct the continuous frequency distribution as explained in Sec. 2.5 of Lab Session 2 (Fig. 6.24).

	D	E	F	G
1		Class Interval	Bin	No. of Days (Frequency) (f)
2	Min. =	100-110	109.99	11
3	101.7	110-120	119.99	9
4	Max. =	120-130	129.99	4
5	168.9	130-140	139.99	6
6	Class Width	140-150	149.99	7
7	9.7292	150-160	159.99	8
8		160-170	169.99	15
9				

Fig. 6.24

Step 2: We compute the mid values as explained in Steps 4 to 6 of Sec. 4.6 of Lab Session 4. The output is shown in Fig. 6.25.

	E	F	G	H	I	J
1	Class Interval	Bin	No. of Days (Frequency) (f)	Lower Limit	Upper Limit	Mid Value (x)
2	100-110	109.99	11	100	110	105
3	110-120	119.99	9	110	120	115
4	120-130	129.99	4	120	130	125
5	130-140	139.99	6	130	140	135
6	140-150	149.99	7	140	150	145
7	150-160	159.99	8	150	160	155
8	160-170	169.99	15	160	170	165
9						

Fig. 6.25

Step 3: To obtain the logarithms with base 10 of the mid values (x), i.e., $\log_{10}(x)$, we refer to Fig. 6.26 and

1. select Cell K2,
2. click on the ***Formulas*** tab, and
3. click on ***Math & Trig*** → ***Log10***

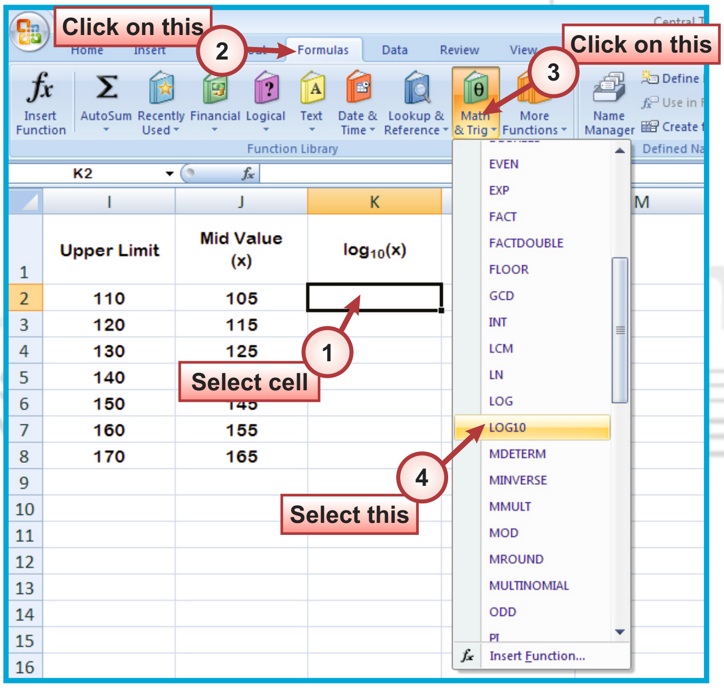


Fig. 6.26

Step 4: A new dialog box appears. We select Cell J2 as shown in Fig. 6.27 and click on ***OK***.

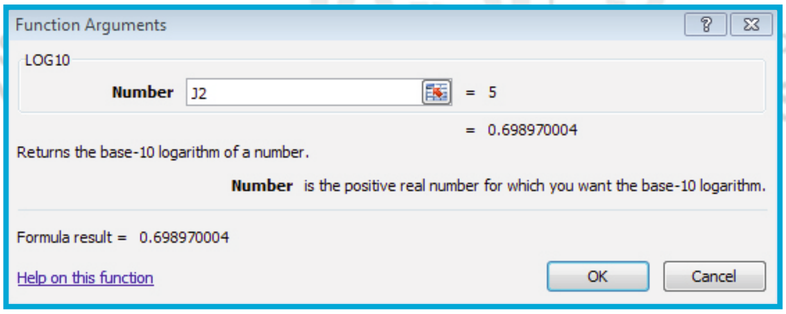
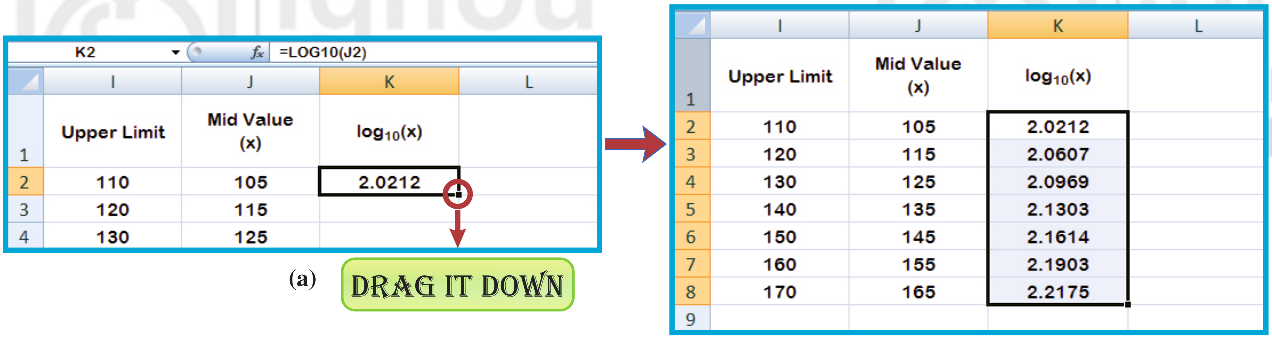


Fig. 6.27

Step 5: The value of $\log_{10}(105)$ is obtained in Cell K2 as shown in Fig. 6.28a. We drag Cell K2 down up to Cell K8 to compute the logarithm of the remaining values as shown in Fig. 6.28b.



K2 =LOG10(J2)

	I	J	K	L
1	Upper Limit	Mid Value (x)	$\log_{10}(x)$	
2	110	105	2.0212	
3	120	115		
4	130	125		

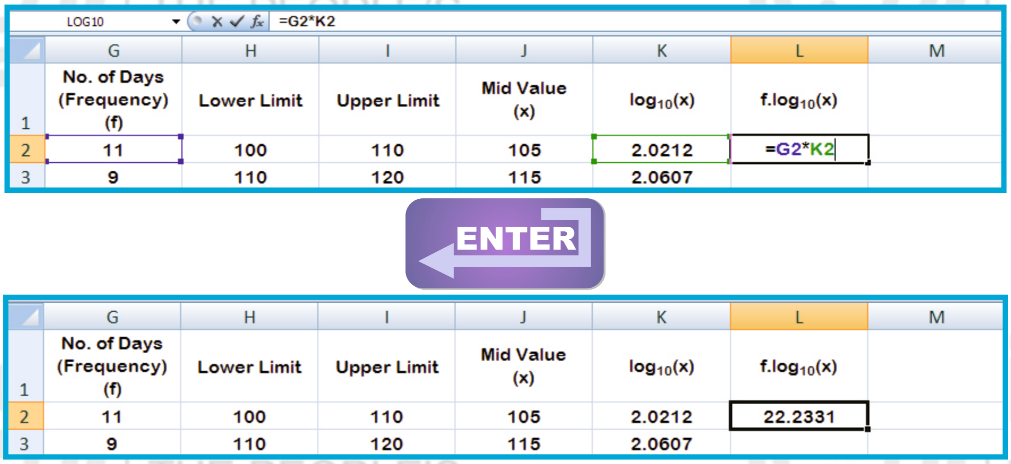
(a)

	I	J	K	L
1	Upper Limit	Mid Value (x)	$\log_{10}(x)$	
2	110	105	2.0212	
3	120	115	2.0607	
4	130	125	2.0969	
5	140	135	2.1303	
6	150	145	2.1614	
7	160	155	2.1903	
8	170	165	2.2175	
9				

(b)

Fig. 6.28

Step 6: To multiply $\log_{10}(x)$ by the corresponding frequency (f), we type "=G2*K2" in Cell L2 as shown in Fig. 6.29a and press ***Enter*** to obtain the result (Fig. 6.29b).



LOG10 =G2*K2

	G	H	I	J	K	L	M
1	No. of Days (Frequency) (f)	Lower Limit	Upper Limit	Mid Value (x)	$\log_{10}(x)$	$f.\log_{10}(x)$	
2	11	100	110	105	2.0212	=G2*K2	
3	9	110	120	115	2.0607		

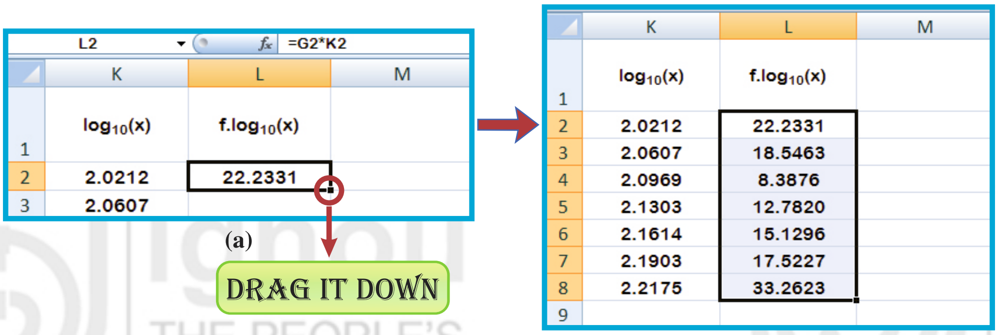
(a)

	G	H	I	J	K	L	M
1	No. of Days (Frequency) (f)	Lower Limit	Upper Limit	Mid Value (x)	$\log_{10}(x)$	$f.\log_{10}(x)$	
2	11	100	110	105	2.0212	22.2331	
3	9	110	120	115	2.0607		

(b)

Fig. 6.29

Step 7: We drag down Cell L2 up to Cell L8 to obtain the remaining values as shown in Fig. 6.30.



L2 =G2*K2

	K	L	M
1	$\log_{10}(x)$	$f.\log_{10}(x)$	
2	2.0212	22.2331	
3	2.0607		

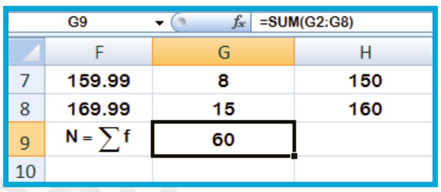
(a)

	K	L	M
1	$\log_{10}(x)$	$f.\log_{10}(x)$	
2	2.0212	22.2331	
3	2.0607	18.5463	
4	2.0969	8.3876	
5	2.1303	12.7820	
6	2.1614	15.1296	
7	2.1903	17.5227	
8	2.2175	33.2623	
9			

Fig. 6.30

(b)

Step 8: To compute the total frequency, i.e., $\sum f$ in Cell G9, we use ***AutoSum*** function under the ***Home*** tab and press ***Enter***. The output is shown in Fig. 6.31.



G9 =SUM(G2:G8)

	F	G	H
7	159.99	8	150
8	169.99	15	160
9	$N = \sum f$	60	
10			

Fig. 6.31

Step 9: In the same way, we compute the sum of the product, i.e., $\sum f \log_{10}(x)$ in Cell L9 as shown in Fig. 6.32.

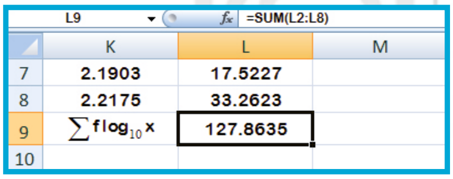
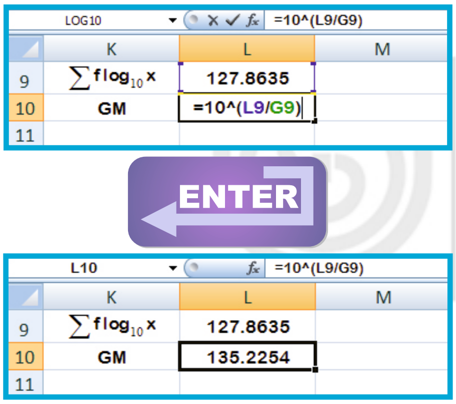


Fig. 6.32

Step 10: We use equation (4) to compute the geometric mean. For this, we type "=10^(L9/G9)" in Cell L10 and press ***Enter*** as shown in Fig. 6.33a. The result shown in Fig. 6.33b.



(a)

(b)

Fig. 6.33

From Fig. 6.33b, the value of geometric mean is given by

$$GM = 135.2254$$

Therefore, the average percentage growth of bacteria from day 1 to 60 = 135.2254 – 100 = 35.2254% or approximately 35%.

6.5 HARMONIC MEAN

You have learnt in Unit 1 of MST-002 that the harmonic mean is commonly used in the computation of average speed. We briefly mention the main points as follows:

- **For Ungrouped (Raw) Data**

  If $x_1, x_2, ..., x_n$ are n observations of a variable X, then their harmonic mean is given by

$$HM = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \qquad \text{... (5)}$$

- **For Discrete Frequency Distribution**

  If $x_1, x_2, ..., x_k$ are k values of a variable X with their corresponding frequencies $f_1, f_2, ..., f_k$, respectively, then

$$HM = \frac{N}{\sum_{i=1}^{k} \frac{f_i}{x_i}} \qquad \text{... (6)}$$

  where $N = \sum_{i=1}^{k} f_i$.

Note: HM cannot be calculated if any observation has value zero.

- **For Continuous Frequency Distribution**

  If the observations are arranged in the form of class intervals, we take x_i as the mid value of the i^{th} class interval and use equation (6).

Here we explain how to compute harmonic mean using Excel 2007 for raw data, discrete and continuous frequency distribution, one at a time for Problem 4.

6.5.1 Harmonic Mean for Raw Data

In order to compute the average speed of the car for the given data, we apply the harmonic mean and follow the steps given below:

Step 1: We enter the data of Table 2 in an Excel sheet as shown in Fig. 6.34 and name it "**HM**".

	A	B	C
1	S. No.	Speed (km/h)	
2	1	45	
3	2	25	
4	3	30	
5	4	30	
6	5	40	
7	6	45	
8	7	50	
9	8	50	
10	9	60	
11	10	65	

HM

Fig. 6.34

Step 2: To compute the harmonic mean, we refer to Fig. 6.35 and

1. select Cell B52,
2. click on the ***Formulas*** tab, and
3. click on ***More Functions*** → ***Statistical*** → ***Harmean***.

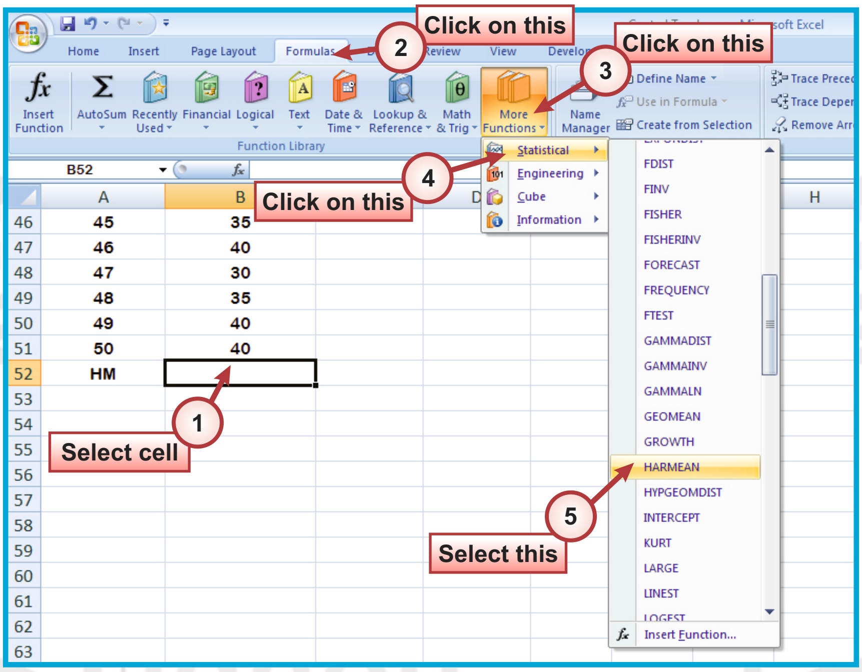


Fig. 6.35

Step 3: A new dialog box appears. We select Cells B2:B51 as shown in Fig. 6.36 and click on ***OK***.

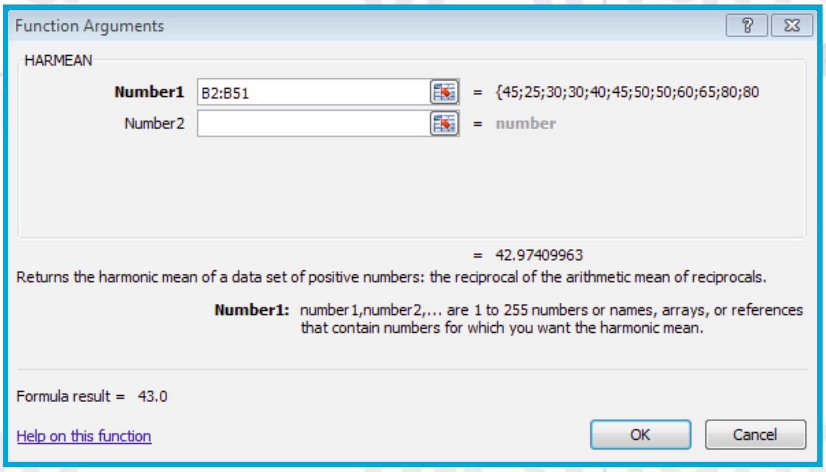


Fig. 6.36

Step 4: The value of the harmonic mean is shown in Cell B52 (Fig. 6.37).

B52 =HARMEAN(B2:B51)

	A	B	C
51	50	40	
52	HM	42.9741	
53			

Fig. 6.37

From Fig. 6.37, the harmonic mean is 42.9741. Therefore, the average speed of the car is approximately 42.97 km per hour.

To explain the computation of the harmonic mean for the grouped frequency distribution, we shall first construct the discrete frequency distribution for the data given in Table 2. Here we are continuing with the same Excel sheet. Note that Excel does not have any built-in function for computing the harmonic mean for the frequency distributions.

6.5.2 Harmonic Mean for Discrete Frequency Distribution

In order to compute the average speed of the car for the discrete frequency distribution, we follow the steps given below:

Step 1: We construct the discrete frequency distribution for the data of Table 2 as explained in Sec. 2.4 of Lab Session 2. The output is shown in Fig. 6.38.

	D	E
1	Speed (km/h) (x)	Distance (km) (f)
2	25	3
3	30	6
4	35	6
5	40	6
6	45	6
7	50	5
8	55	5
9	60	6
10	65	3
11	75	2
12	80	2

Fig. 6.38

Step 2: To compute f/x for the first value, we type "=E2/D2" in Cell F2 as shown in Fig. 6.39a and press ***Enter***. We obtain the result shown in Fig. 6.39b.

F2 | f_x =E2/D2

	D	E	F
1	Speed (km/h) (x)	Distance (km) (f)	f / x
2	25	3	0.1200
3	30	6	

(a)

ENTER

HARMEAN | × ✓ f_x =E2/D2

	D	E	F
1	Speed (km/h) (x)	Distance (km) (f)	f / x
2	25	3	=E2/D2
3	30	6	

(b)

Fig. 6.39

Step 3: We drag down Cell F2 up to Cell F12 to compute (f/x) for the remaining values as shown in Fig. 6.40.

F2 | f_x =E2/D2

	D	E	F	G
1	Speed (km/h) (x)	Distance (km) (f)	f / x	
2	25	3	0.1200	
3	30	6		
4	35	6		

(a)

DRAG IT DOWN

	D	E	F	G
1	Speed (km/h) (x)	Distance (km) (f)	f / x	
2	25	3	0.1200	
3	30	6	0.2000	
4	35	6	0.1714	
5	40	6	0.1500	
6	45	6	0.1333	
7	50	5	0.1000	
8	55	5	0.0909	
9	60	6	0.1000	
10	65	3	0.0462	
11	75	2	0.0267	
12	80	2	0.0250	

(b)

Fig. 6.40

Step 4: We determine the values of $\sum f$ and $\sum f/x$ in Cells E13 and F13, using the ***Sum*** function as shown in Figs. 6.41a and b, respectively.

E13 | f_x =SUM(E2:E12)

	D	E	F
12	80	2	0.0250
13	Total	50	
14			

(a)

F13 | f_x =SUM(F2:F12)

	D	E	F
12	80	2	0.0250
13	Total	50	1.1635
14			

(b)

Fig. 6.41

Step 5: We type "=E13/F13" in Cell F14 to compute the harmonic mean as shown in Fig. 6.42a and press ***Enter***. We get the result shown in Fig. 6.42b.

HARMEAN =E13/F13

	D	E	F
12	80	2	0.0250
13	Total	50	1.1635
14		HM	=E13/F13
15			

(a)

ENTER

F14 =E13/F13

	D	E	F
12	80	2	0.0250
13	Total	50	1.1635
14		HM	42.9741
15			

(b)

Fig. 6.42

From Fig. 6.42b, the value for harmonic mean is given by

$$\text{HM} = 42.9741$$

Thus, the average speed of the car is 42.9741 km/hr for the entire journey on the basis of the discrete frequency distribution.

6.5.3 Harmonic Mean for Continuous Frequency Distribution

We now explain how to compute the harmonic mean for a continuous frequency distribution.

Step 1: We construct the continuous frequency distribution as explained in Sec. 2.5 of Lab Session 2 and also compute the mid values as explained in Steps 4 to 6 of Sec. 4.6 of Lab Session 4. The output is shown in Fig. 6.43.

	H	I	J	K	L	M	N
1		Class Interval	Bin	Distance (km) (Frequency) (f)	Lower Limit	Upper Limit	Mid Value (x)
2	Min. =	20-30	29	3	20	30	25
3	25	30-40	39	12	30	40	35
4	Max. =	40-50	49	12	40	50	45
5	80	50-60	59	10	50	60	55
6	Class Width	60-70	69	9	60	70	65
7	8.2782	70-80	79	2	70	80	75

Fig. 6.43

Step 2: We type "=K2/N2" in Cell O2 and drag it down up to Cell O7 to compute f/x for all classes as shown in Fig. 6.44a and b.

(a)

O2 =K2/N2

	K	L	M	N	O	P
1	Distance (km) (Frequency) (f)	Lower Limit	Upper Limit	Mid Value (x)	f / x	
2	3	20	30	25	0.1200	
3	12	30	40	35		
4	12	40	50	45		

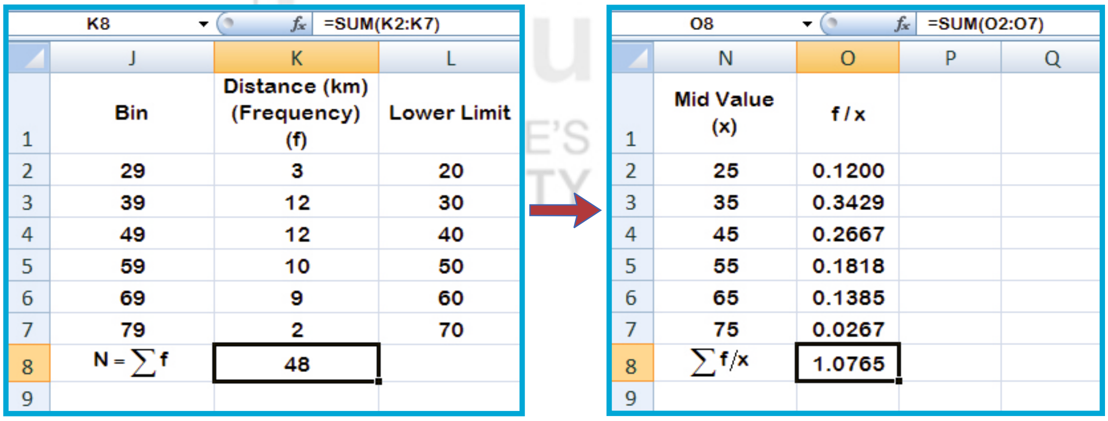
DRAG IT DOWN

(b)

	K	L	M	N	O	P
1	Distance (km) (Frequency) (f)	Lower Limit	Upper Limit	Mid Value (x)	f / x	
2	3	20	30	25	0.1200	
3	12	30	40	35	0.3429	
4	12	40	50	45	0.2667	
5	10	50	60	55	0.1818	
6	9	60	70	65	0.1385	
7	2	70	80	75	0.0267	

Fig. 6.44

Step 3: We compute the values of $\sum f$ and $\sum f/x$ in Cells K8 and O8, respectively, using the ***Sum*** function as shown in Fig. 6.45.



(a) (b)

Fig. 6.45

Step 4: To determine the harmonic mean using equation (6), we type "=K8/O8" in Cell O9 as shown in Fig. 6.46a and press ***Enter***. We get the result shown in Fig. 6.46b.

(a) (b)

Fig. 6.46

From Fig. 6.46b, the harmonic mean is given by

HM = 44.5902

Thus, the average speed is 44.5902 km/hr for the entire journey on the basis of the continuous frequency distribution.

6.6 MODE

You have learnt in Unit 1 of MST-002 that the mode is defined as the value that is repeated most often in the data. Mode is widely used in production, i.e., to determine shirt size, shoe size, etc. Here we briefly mention the important points as follows:

- **Mode for Ungrouped (Raw) Data**

  Let $x_1, x_2, \ldots, x_n$ be n observations. If some observations are repeated in the data, we determine the observation which is repeated the highest number of times. This value is known as mode.

- **Mode for Discrete Frequency Distribution**

  If observations $x_1, x_2, \ldots, x_k$ occur with frequencies $f_1, f_2, \ldots, f_k$, respectively, the mode will be the value corresponding to the highest frequency.

- **Mode for Continuous Frequency Distribution**

  For continuous frequency distribution, we first determine the modal class. Then we use the formula given below to compute the mode:

$$M_0 = L + \frac{f_1 - f_0}{(2f_1 - f_0 - f_2)} \times h \qquad \ldots (7)$$

where L = lower class limit of the modal class,

f_1 = frequency of the modal class,

f_0	=	frequency of the pre-modal class,
f_2	=	frequency of the post-modal class, and
h	=	width of the modal class.

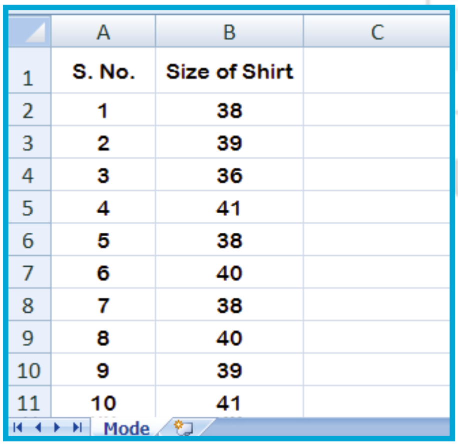
Note that modal class is the class that has the maximum frequency.

Here we explain how to compute mode using Excel 2007 for raw data, discrete and continuous frequency distributions, one at a time for Problem 5.

6.6.1 Mode for Raw Data

In order to compute mode for raw data, we follow the steps given below:

Step 1: We enter the data of Table 3 in Excel sheet as shown in Fig. 6.47 and name it "**Mode**".



	A	B	C
1	S. No.	Size of Shirt	
2	1	38	
3	2	39	
4	3	36	
5	4	41	
6	5	38	
7	6	40	
8	7	38	
9	8	40	
10	9	39	
11	10	41	

Mode

Fig. 6.47 : Partial screenshot of the spreadsheet for the given data.

Step 2: We now compute the mode for the given data. For this, we refer to Fig. 6.48 and

1. select Cell B102,
2. click on the ***Formulas*** tab, and
3. click on ***More Functions*** $\rightarrow$ ***Statistical*** $\rightarrow$ ***Mode***.

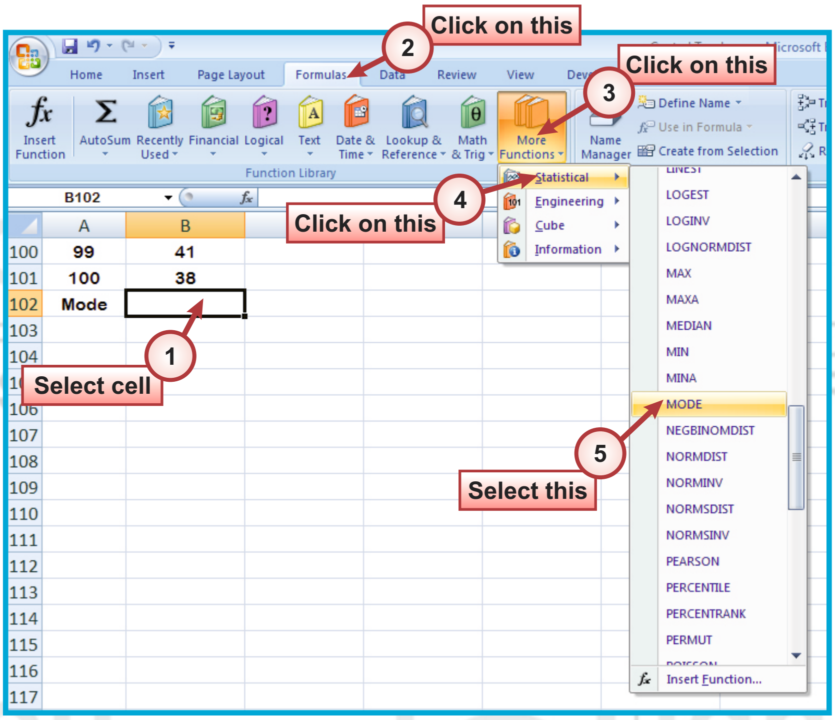


Fig. 6.48

Step 3: A new dialog box appears and we select Cell B2:B101 and click on ***OK*** as shown in Fig. 6.49.

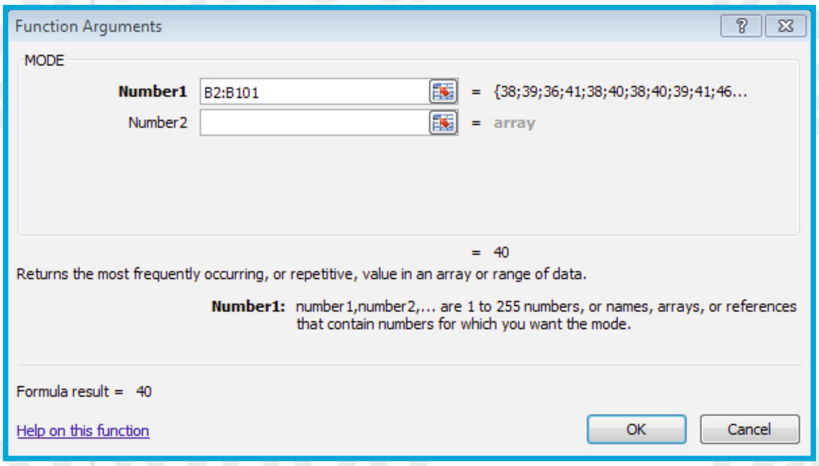


Fig. 6.49

Step 4: The value of the mode is obtained in Cell B102 as shown in Fig. 6.50.

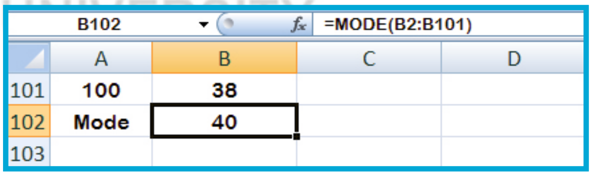


Fig. 6.50

From Fig. 6.50, the mode is given by

Mode = 40.

Therefore, the modal size of the shirt is 40.

To explain the computation of mode for the grouped frequency distribution, we shall first construct the discrete frequency distribution for the data given in Table 3. Here we are continuing with the same Excel sheet. Note that Excel does not have any built-in function to compute the mode for frequency distributions.

6.6.2 Mode for Discrete Frequency Distribution

In order to compute the average size of the shirt for the discrete frequency distribution, we follow the steps given below:

Step 1: We now construct the discrete frequency distribution for the data of Table 3 as explained in Sec. 2.4 of Lab Session 2. The output is shown in Fig. 6.51.

	D	E	F
1	Size of Shirt	Number of Person (Frequency)	
2	34	2	
3	35	4	
4	36	6	
5	37	8	
6	38	11	
7	39	13	
8	40	18	
9	41	13	
10	42	11	
11	43	7	
12	44	4	
13	45	2	
14	46	1	

Fig. 6.51

Step 2: Since there is no built-in function in Excel to compute the mode for the discrete frequency distribution, we visually identify the mode after forming the discrete frequency distribution. Notice from Fig. 6.52 that the highest frequency, i.e., 18, corresponds to the shirt size 40. We have highlighted the corresponding cells with purple colour as shown in Fig. 6.52.

	D	E	F
1	Size of Shirt	Number of Person (Frequency)	
2	34	2	
3	35	4	
4	36	6	
5	37	8	
6	38	11	
7	39	13	
8	40	18	
9	41	13	
10	42	11	
11	43	7	
12	44	4	
13	45	2	
14	46	1	
15	Mode	40	

Fig. 6.52

Therefore, the modal size of the shirt is 40.

6.6.3 Mode for Continuous Frequency Distribution

We now explain how to compute mode for a continuous frequency distribution in the steps given below:

Step 1: We construct the continuous frequency distribution for the data of Table 3 as explained in Sec. 2.5 of Lab Session 2 (see Fig. 6.53).

	G	H	I	J	K
1		Class Interval	Bin	Number of Person (Frequency)	
2	Min. =	34-36	35	6	
3	34.0	36-38	37	14	
4	Max. =	38-40	39	24	
5	46.0	40-42	41	31	
6	Class Width	42-44	43	18	
7	1.5699	44-46	45	6	
8		46-48	47	1	

Fig. 6.53

Step 2: The highest frequency, i.e., 31, corresponds to the class 40-42. We highlight the highest frequency and its corresponding class interval with purple colour as shown in Fig. 6.54.

	H	I	J	K
1	Class Interval	Bin	Number of Person (Frequency)	
2	34-36	35	6	
3	36-38	37	14	
4	38-40	39	24	
5	40-42	41	31	
6	42-44	43	18	
7	44-46	45	6	
8	46-48	47	1	

Fig. 6.54

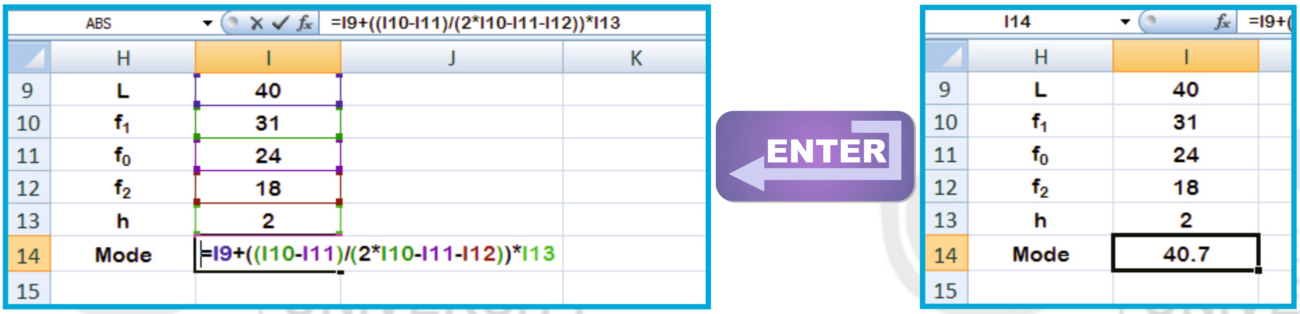
Step 3: We type the values of L, f_1, f_0, f_2 and h in Cells I9, I10, I11, I12 and I13, respectively, as shown in Fig. 6.55.

	H	I	J
1	Class Interval	Bin	Number of Person (Frequency)
2	34-36	35	6
3	36-38	37	14
4	38-40	39	24
5	40-42	41	31
6	42-44	43	18
7	44-46	45	6
8	46-48	47	1
9	L	40	
10	f_1	31	
11	f_0	24	
12	f_2	18	
13	h	2	

Fig. 6.55

Step 4: To compute mode for the continuous frequency distribution, we use equation (7). The values of L, f_1, f_0, f_2 and h are given in Cells I9, I10, I11, I12 and I13, respectively. We type "= I9+((I10-I11)/(2*I10-I11-I12))*I13" in Cell I14 as shown in Fig. 6.56a and press ***Enter***. The output is shown in Fig. 6.56b.

In the same way as for discrete frequency distribution, we can compute the weighted average by considering weights (w) instead of frequency (f).



(a) **(b)**

Fig. 6.56

From Fig. 6.56b, the mode is given by

$$\text{Mode} = 40.7 \approx 41$$

Thus, the modal size of the shirt is approximately 41.

6.7 PARTITION VALUES

In Unit 1 of MST-002, you have learnt that the partition values are the measures which divide the data into several equal parts. The median divides the data into 2 equal parts, quartiles divide the data into 4 equal parts, deciles divide data into 10 equal parts, and percentiles divide data into 100 equal parts. Here we briefly mention the important points and formulae used to compute these partition values as follows:

Median is the middle most value of the variable which divides the entire data into two equal parts.

For Ungrouped (Raw) Data

If $x_1, x_2, \ldots, x_n$ are n observations, we arrange the given data in ascending or descending order and compute the partition values as follows:

- The **median** is computed from

$$M_d = \left(\frac{n+1}{2}\right)^{th} \text{observation} \quad \text{(when n is odd)} \qquad \ldots (8)$$

$$M_d = \frac{\left(\frac{n}{2}\right)^{th} \text{observation} + \left(\frac{n}{2}+1\right)^{th} \text{observation}}{2} \quad \text{(when n is even)} \; \ldots (9)$$

- The **k^th quartile** is

$$Q_k = k\left(\frac{n+1}{4}\right)^{th} \text{observation} \qquad (k = 1, 2, 3) \qquad \ldots (10)$$

- The **k^th decile** is

$$D_k = k\left(\frac{n+1}{10}\right)^{th} \text{observation} \qquad (k = 1, 2, \ldots, 9) \qquad \ldots (11)$$

- The **k^th percentile** is

$$P_k = k\left(\frac{n+1}{100}\right)^{th} \text{observation} \qquad (k = 1, 2, \ldots, 99) \qquad \ldots (12)$$

For Discrete Frequency Distribution

- If observations $x_1, x_2, ..., x_k$ occur with frequencies $f_1, f_2, ..., f_k$, respectively, we compute the less than type cumulative frequencies.
- If $N = \sum_{i=1}^{k} f_i$ is the total frequency, the formula used for discrete frequency distribution can be written in the same way as equations (8), (10), (11) and (12) by considering N instead of n.

  Note that if N/2 is not the exact cumulative frequency, we consider the observation corresponding to the next cumulative frequency as the required partition value.

For Continuous Frequency Distribution

We compute the less than type cumulative frequency for the given continuous frequency distribution and then calculate the partition values as follows:

- The k^{th} quartile is

$$Q_k = L + \frac{k\left(\frac{N}{4}\right) - C}{f} \times h \; ; \qquad (k = 1, 2, 3) \qquad \ldots (13)$$

- The k^{th} d1ecile is

$$D_k = L + \frac{k\left(\frac{N}{10}\right) - C}{f} \times h \; ; \qquad (k = 1, 2, \ldots, 9) \qquad \ldots (14)$$

- The k^{th} percentile is

$$P_k = L + \frac{k\left(\frac{N}{100}\right) - C}{f} \times h \; ; \qquad (k = 1, 2, \ldots, 99) \qquad \ldots (15)$$

where L – lower class limit of the corresponding partition value class,

h – width of the partition value class,

N – total frequency,

C – cumulative frequency of the class preceding the partition value class, and

f – frequency of the partition value class.

Here we describe the computation of the partition values in Excel 2007 for the raw data, discrete and continuous frequency distribution, one at a time for Problem 2.

6.7.1 Partition Values for Raw Data

In order to compute the median, quartiles, deciles and percentiles for the given data, we follow the steps given below:

Median is defined as the middle or central value of the data.

Step 1: We enter the data of Table 3 of Lab Session 2 in an Excel sheet as shown in Fig. 6.57.

	A	B	C
1	S.No.	Life of Bulb (in hours)	
2	1	1087	
3	2	1289	
4	3	876	
5	4	725	
6	5	900	
7	6	1080	
8	7	952	

Partition Values

Fig. 6.57: Partial screenshot of the spreadsheet for the given data.

Step 2: To compute the median for the raw data, Excel has a built-in function ***Median***. For this, we refer to Fig. 6.58 and

1. select Cell B152,
2. click on the ***Formulas*** tab, and
3. click on ***More Functions*** → ***Statistical*** → ***Median***.

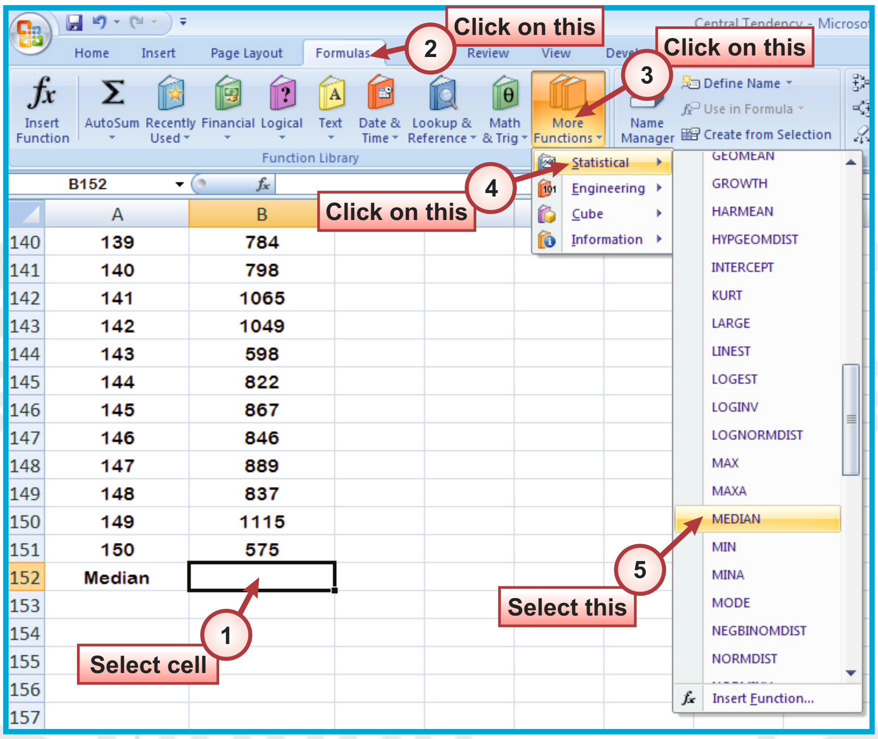


Fig. 6.58

Step 3: A new dialog box appears in which we select Cells B2:B151 and click on ***OK*** as shown in Fig. 6.59a. The value of the median is shown in Fig. 6.59b.

(a)

(b)

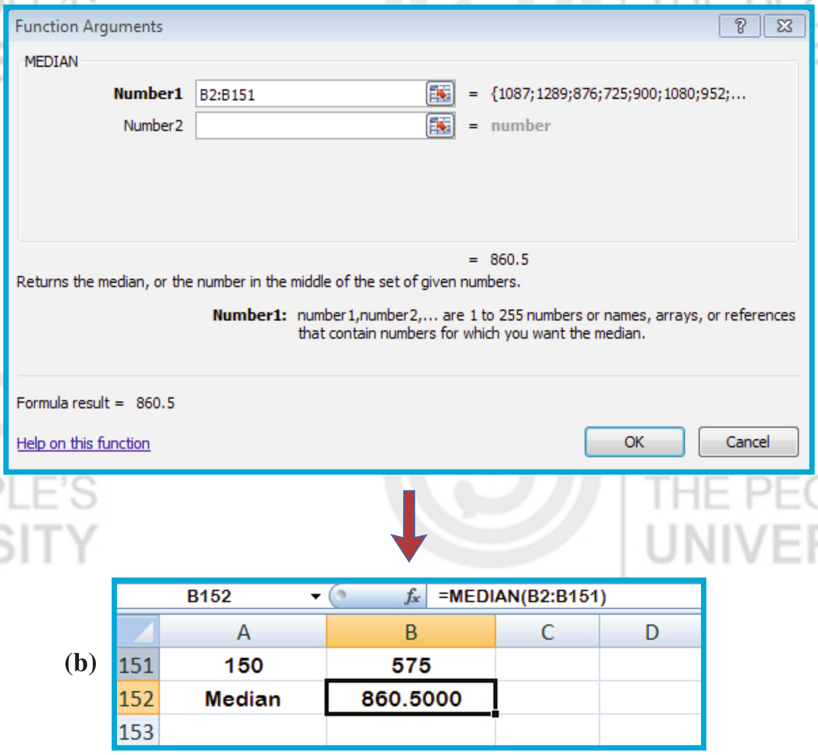


Fig. 6.59

Step 4: To compute the quartile for the raw data, Excel has the built-in function ***Quartile***. To use it, we refer to Fig. 6.60 and

1. select Cell B153,
2. click on the ***Formulas*** tab, and
3. click on ***More Functions*** → ***Statistical*** → ***Quartile***.

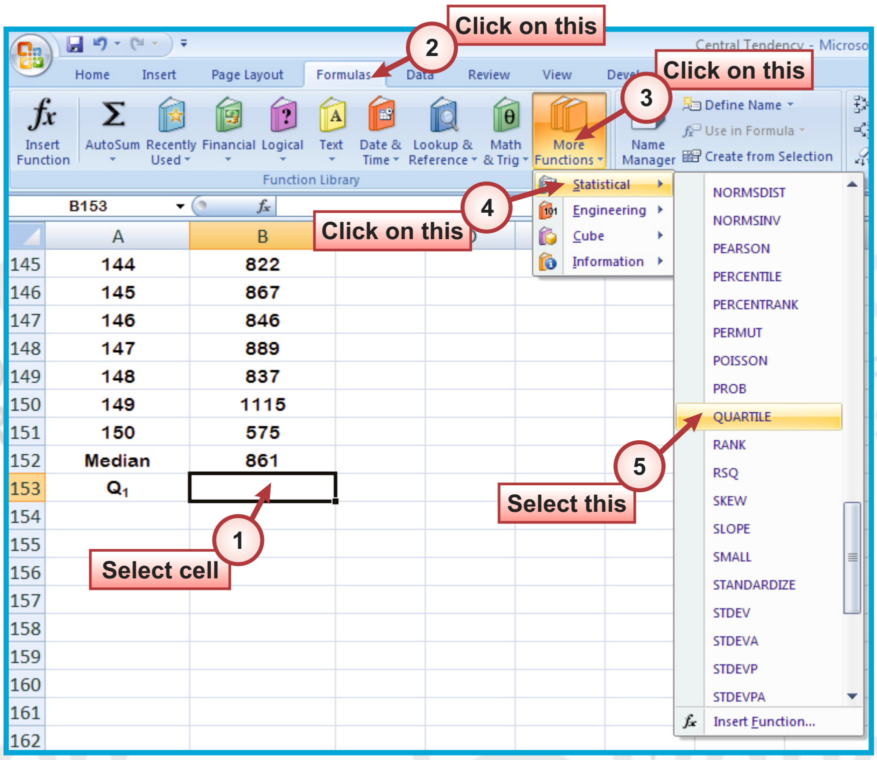
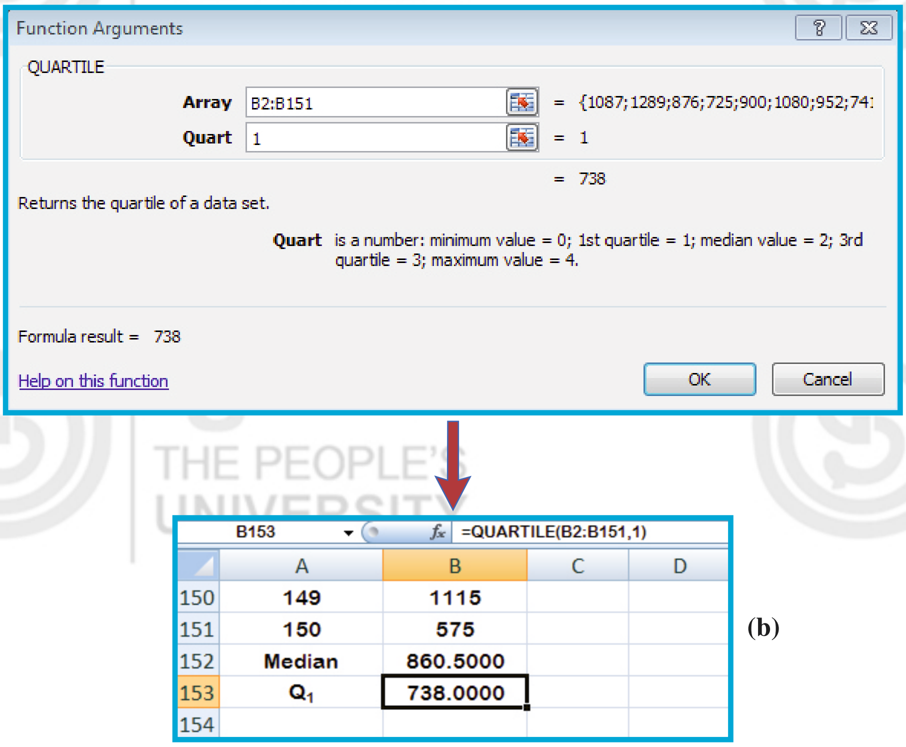


Fig. 6.60

Step 5: A new dialog box appears in which we select Cells B2:B151 in ***Array*** and type "1" in ***Quart*** to compute the first quartile and click on ***OK*** as shown in Fig. 6.61a. The value of the first quartile is shown in Fig. 6.61b.

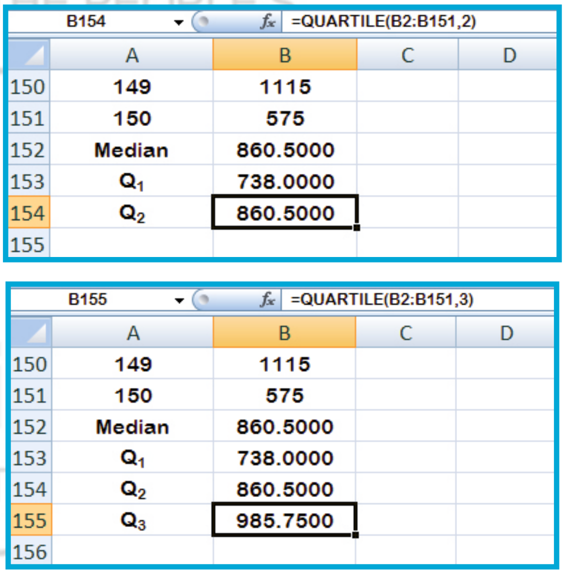


(a)

(b)

Fig. 6.61

Step 6: We can also compute the values of second and third quartiles in Cells B154 and B155 by typing "2" and "3" in ***Quart*** (Fig. 6.62), respectively, in the same way as explained in Steps 4 and 5.



(a)

(b)

Fig. 6.62

Step 7: Before computing deciles, we explain how to determine the percentile for the raw data. To compute the quartile for the raw data, Excel has a built-in function ***Percentile***. To use it, we refer to Fig. 6.63 and

1. select Cell B156,
2. click on the ***Formulas*** tab, and
3. click on ***More Functions*** → ***Statistical*** → ***Percentile***.

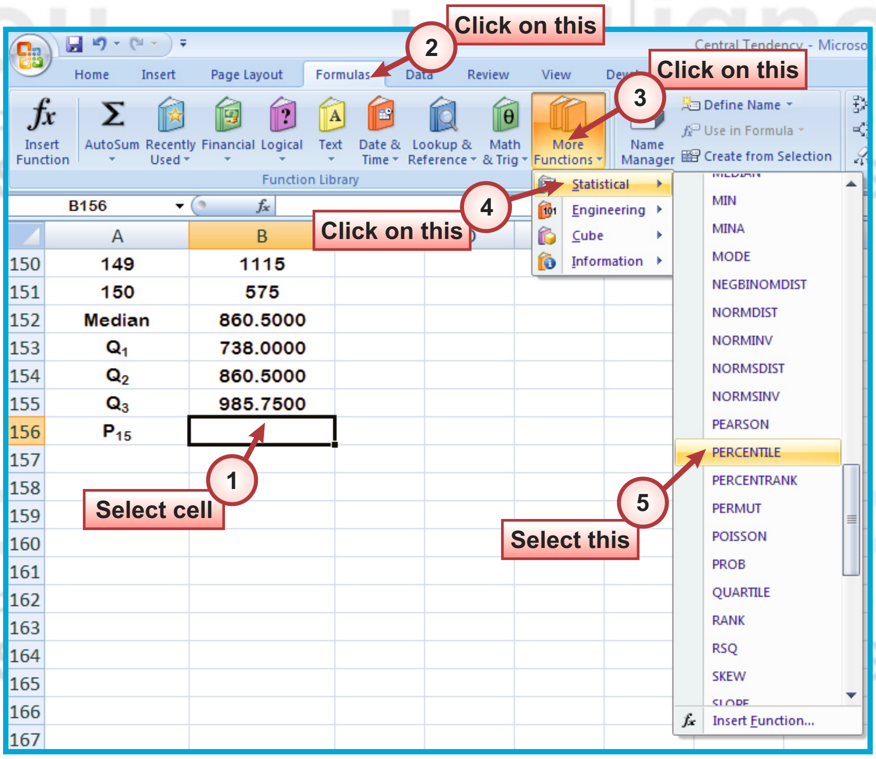
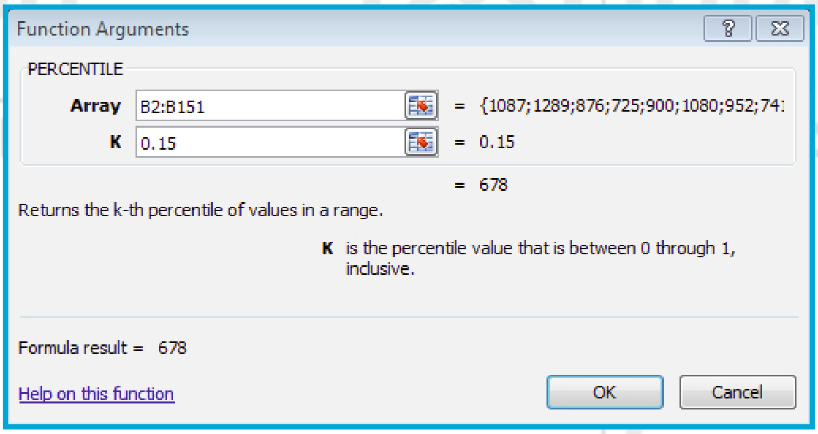


Fig. 6.63

> For calculating the k[th] percentile, we type the value (k/100) in ***K*** as shown in dialog box (Fig. 6.64a).

Step 8: A new dialog box appears. To compute the 15[th] percentile, we select Cells B2:B151 in ***Array*** and type “0.15” in ***K*** and click on ***OK*** as shown in Fig. 6.64a. The value of the 15[th] percentile is shown in Fig. 6.64b.

(a)



(b)

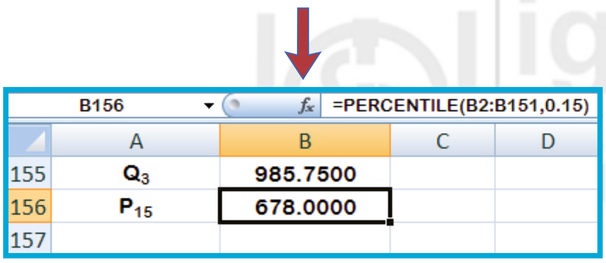
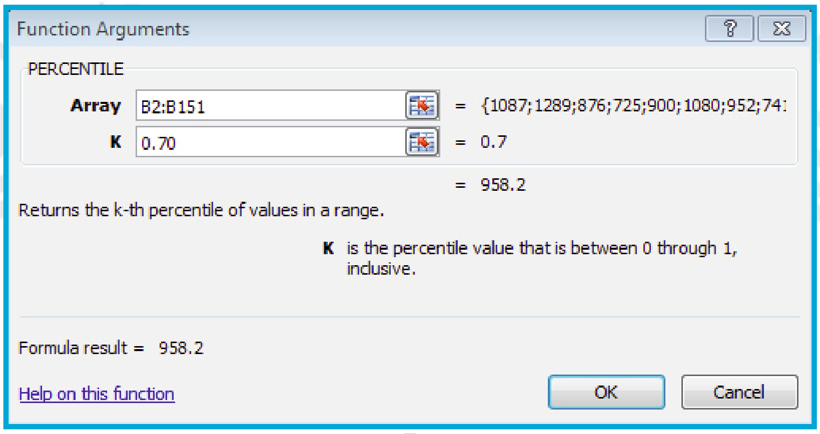


Fig. 6.64

Step 9: A built-in function to determine the deciles is not available in Excel 2007. So we calculate deciles using ***Percentile*** function. We repeat Step 7 by selecting Cell B157. To compute the 7[th] decile which corresponds to the 70[th] percentile, we select Cells B2:B151 in ***Array*** and type “0.70” in ***K*** and click on ***OK*** as shown in Fig. 6.65a. The value of the 7[th] decile is shown in Fig. 6.65b.



(a)

B157 f_x =PERCENTILE(B2:B151,0.7)

	A	B	C	D
155	Q_3	985.7500		
156	P_{15}	678.0000		
157	D_7	958.2000		
158				

(b)

Fig. 6.65

> The 1st, 2nd, …, 10th deciles correspond to 10th, 20th, …, 100th percentiles.

6.7.2 Partition Values for Discrete Frequency Distribution

Excel does not have any built-in function to compute the Partition values for the frequency distributions. In order to compute the median, quartiles, deciles and percentiles for the discrete frequency distribution, we consider Problem 1 and follow the steps given below:

Step 1: We consider the discrete frequency distribution determined in Sec. 2.4 of Lab Session 2 and compute the less than type cumulative frequencies as shown in Fig. 6.66.

	A	B	C
1	Number of Mobile Phones (x)	No. of Families (f)	No. of Bulbs (Less Than Cumulative Frequency)
2	1	6	6
3	2	16	22
4	3	22	44
5	4	24	68
6	5	22	90
7	6	16	106
8	7	9	115
9	8	5	120
10	$N = \sum f$	120	

Fig. 6.66

Step 2: We type "=B10/2" in Cell B12 to compute the value of N/2 as shown in Fig. 6.67. The value of N/2 = 60 corresponds to the cumulative frequency 68 (Cell C5). The value of the variable (number of mobile phones) corresponding to the cumulative frequency (68) is 4 (Cell A5). Thus, the median is 4.

B12 f_x =B10/2

	A	B	C
11	Median		
12	N/2	60	
13	M_d	4	

Fig. 6.67

Step 3: To compute the first quartile, we type "=B10/4" in Cell B16 and obtain the value of N/4 as 30, which corresponds to the cumulative frequency 44 (Cell C4). Thus, the number of mobile phones corresponding to the cumulative frequency (44) is 3 (Fig. 6.68). The value of the first quartile is 3 as shown in Fig. 6.68.

B16 fx =B10/4

	A	B	C
15	1^{st} Quartile		
16	N/4	30	
17	Q_1	3	
18			

Fig. 6.68

Step 4: In the same way as explained in Steps 2 and 3, we can compute the 2^{nd} quartile, 3^{rd} quartile, 7^{th} decile and 15^{th} percentile as shown in Figs. 6.69a to d.

(a)

B20 fx =2*B10/4

	A	B	C
19	2^{nd} Quartile		
20	2N/4	60	
21	Q_2	4	

(c)

B24 fx =3*B10/4

	A	B	C
23	3^{rd} Quartile		
24	3N/4	90	
25	Q_3	5	

(b)

B28 fx =7*B10/10

	A	B	C
27	7^{th} Decile		
28	7N/10	84	
29	D_7	5	

(d)

B32 fx =15*B10/100

	A	B	C
31	15^{th} Percentile		
32	15N/100	18	
33	P_{15}	2	

Fig. 6.69

6.7.3 Partition Values for Continuous Frequency Distribution

We now consider Problem 2 and explain how to compute the partition values, i.e., median, quartiles, deciles and percentiles for continuous frequency distribution in the steps given below:

Step 1: We consider the continuous frequency distribution determined in Sec. 2.5 of Lab Session 2. We compute the less than type cumulative frequencies and type the value of h in Cell B11 as shown in Fig. 6.70.

	A	B	C
1	Class Interval	No. of Bulbs (f)	No. of Bulbs (Less Than Cumulative Frequency)
2	500-600	11	11
3	600-700	20	31
4	700-800	25	56
5	800-900	32	88
6	900-1000	27	115
7	1000-1100	19	134
8	1100-1200	11	145
9	1200-1300	5	150
10	$N = \sum f$	150	
11	h	100	

Fig. 6.70

Step 2: To compute the median, we refer to Fig. 6.71 and

1. compute the value of N/2 by typing "=B10/2" in Cell B14 and highlight the corresponding median class with purple colour (Fig. 6.71a),
2. type the corresponding values of L, C and f in Cells B15, B16 and B17, respectively, and

3. type "=B15+((B14-B16)/B17)*B11" in Cell B18 and get the value of median as shown in Fig.6.71b.

	A	B	C
1	Class Interval	No. of Bulbs (f)	No. of Bulbs (Less Than Cumulative
2	500-600	11	11
3	600-700	20	31
4	700-800	25	56
5	800-900	32	88
6	900-1000	27	115
7	1000-1100	19	134
8	1100-1200	11	145
9	1200-1300	5	150

(a)

B18 f_x =B15+((B14-B16)/B17)*B11

	A	B	C
13	Median		
14	N/2	75	
15	L	800	
16	C	56	
17	f	32	
18	M_d	859.3750	
19			

(b)

Fig. 6.71

Step 3: To compute the 1st quartile, we refer to Fig. 6.72 and

1. compute the value of N/4 by typing "=B10/4" in Cell B21 and highlight the corresponding 1st quartile class with purple colour (Fig.6.72a),
2. type the corresponding values of L, C and f in Cells B22, B23 and B24, respectively, and
3. type "=B22+((B21-B23)/B24)*B11" in Cell B25 and get the value as shown in Fig.6.72b.

	A	B	C
1	Class Interval	No. of Bulbs (f)	No. of Bulbs (Less Than Cumulative
2	500-600	11	11
3	600-700	20	31
4	700-800	25	56
5	800-900	32	88
6	900-1000	27	115
7	1000-1100	19	134
8	1100-1200	11	145
9	1200-1300	5	150

(a)

B25 f_x =B22+((B21-B23)/B24)*B11

	A	B	C
20	1st Quartile		
21	N/4	37.5	
22	L	700	
23	C	31	
24	f	25	
25	Q_1	726.0000	
26			

(b)

Fig. 6.72

Step 4: In the same way as explained in Steps 2 and 3, we can compute the 2^{nd} quartile, 3^{rd} quartile, 7^{th} decile and 15^{th} percentile as shown in Figs. 6.73a to d.

(a)

B32 | fx =B29+((B28-B30)/B31)*B11

	A	B	C
27	2^{nd} Quartile		
28	2N/4	75	
29	L	800	
30	C	56	
31	f	32	
32	Q_2	859.3750	
33			

(c)

B39 | fx =B36+((B35-B37)/B38)*B11

	A	B	C
34	3^{rd} Quartile		
35	3N/4	112.5	
36	L	900	
37	C	88	
38	f	27	
39	Q_3	990.7407	
40			

(b)

B46 | fx =B43+((B42-B44)/B45)*B11

	A	B	C
41	7^{th} Decile		
42	7N/10	105	
43	L	900	
44	C	88	
45	f	27	
46	D_7	962.9630	
47			

(d)

B53 | fx =B50+((B49-B51)/B52)*B11

	A	B	C
48	15^{th} Percentile		
49	15N/100	22.5	
50	L	600	
51	C	11	
52	f	20	
53	P_{15}	657.5000	
54			

Fig. 6.73

You should now apply this method to other problems for practice.

Activity

Apply suitable measures of central tendency with the help of MS Excel 2007 to :

A1) Examples 1 to 15 given in Unit 1 of MST-002.

A2) Exercises E1 to E16 given in Unit 1 of MST-002.

Match the results with the manual computation of data carried out in Unit 1 of MST-002.

Continuous Assessment 6

1. Consider the data given in Table 5 of Lab Session 2 to compute the arithmetic mean, median, quartiles, 4^{th} decile, 37^{th} percentile for the discrete frequency distribution.

2. Consider the data given in Table 6 of Lab Session 2 to compute the arithmetic mean, median, quartiles, 4^{th} decile, 37^{th} percentile for raw data and the continuous frequency distribution.

3. The price of a particular product is noted for 20 years. The yearly rise in the price is recorded in Table 4.

Table 4: Yearly rise in the price of a product

Year	Rise in Price (%)	Year	Rise in Price (%)
1994	11.8	2004	48.9
1995	12.7	2005	50.4
1996	14.4	2006	54.7
1997	19.7	2007	59.7
1998	21.6	2008	65.2
1999	26.4	2009	67.7
2000	30.2	2010	72.4
2001	38.8	2011	80.1
2002	44.7	2012	90.2
2003	47.5	2013	98.2

Compute the average percentage increase in the price of the product.

4. A student travels a total distance of 35 km from home to college by a two-wheeler. The average speed of the two-wheeler for each kilometre of distance travelled is recorded in Table 5.

Table 5: Average speed of the two-wheeler from home to college

S. No.	Speed (km/h)	S. No.	Speed (km/h)	S. No.	Speed (km/h)
1	30	13	35	25	45
2	25	14	45	26	40
3	30	15	30	27	35
4	35	16	40	28	30
5	35	17	20	29	40
6	45	18	20	30	25
7	30	19	25	31	35
8	35	20	30	32	20
9	45	21	30	33	20
10	45	22	40	34	25
11	40	23	45	35	20
12	20	24	45		

Compute the average speed of the two-wheeler during the entire journey.

5. A shoe manufacturing company conducted a survey among 50 women to determine the size of their shoes. The data is recorded in Table 6.

Table 6: Size of shoes

S. No.	Size of Shoe	S. No.	Size of Shoe	S. No.	Size of Shoe
1	5	18	5	35	8
2	6	19	6	36	5
3	7	20	5	37	7
4	8	21	10	38	6
5	5	22	9	39	8
6	7	23	8	40	7
7	5	24	7	41	4
8	7	25	6	42	9
9	6	26	5	43	7
10	8	27	6	44	6
11	9	28	8	45	7
12	9	29	7	46	5
13	7	30	6	47	7
14	7	31	7	48	6
15	7	32	8	49	8
16	4	33	6	50	5
17	8	34	9		

For this data,

- compute the most preferable average size of the shoes.
- construct the discrete and continuous frequency distributions.
- compute the most preferable average size for the frequency distributions.

Home Work: Do It Yourself

1) Follow the steps explained in Secs. 6.3 to 6.7 to compute the various measures of the central tendency. Take the final screenshots and keep them in your record book.

2) Develop the spreadsheets for the exercises of "Continuous Assessment 6" as explained in this lab session. Take screenshots of the final spreadsheets.

3) **Do not forget** to keep all screenshots in your record book as these will contribute to your continuous assessment in the Laboratory.