Measures of Dispersion, Moments, Skewness and Kurtosis

SESSION 7

Structure

7.1 Introduction
Objectives
7.2 Problem Description
7.3 Range and Coefficient of Range
For Raw Data
7.4 Quartile Deviation and Coefficient of Quartile Deviation
For Raw Data
7.5 Mean Deviation
For Raw Data
For Discrete Frequency Distribution
7.6 Variance and Standard Deviation
For Raw Data
For Discrete Frequency Distribution
7.7 Moments, Skewness and Kurtosis
For Raw Data
For Discrete Frequency Distribution
7.8 Descriptive Statistics using Data Analysis ToolPak

7.1 INTRODUCTION

In Lab Session 6, you have learnt 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 have also learnt computation of the partition values, viz. quartiles, deciles and percentiles for raw data as well as discrete and continuous frequency distributions. These measures of central tendency provide a value around which the data is concentrated. But it does not give us any idea about the nature of scatter or spread. In Unit 2 of MST-002 (Descriptive Statistics), we have explained that the degree to which numerical data tend to spread about an average value is called the variation or dispersion of data. We have also described in Units 3 and 4, the concepts of moments, skewness and kurtosis which are useful for getting an idea about the shape of the distribution.

Prerequisite

- Lab Sessions 2 and 6 of MSTL-001 (Basic Statistics Lab).
- Units 2, 3 and 4 of MST-002 (Descriptive Statistics).

In this lab session, you will learn how to compute different measures of dispersion, such as, range, quartile deviation, and its coefficients for raw data as

well as discrete and continuous frequency distributions using MS Excel 2007. You will also learn the method of computation of mean deviation, variance, standard deviation, coefficient of variation, moments, skewness and kurtosis for raw data as well as discrete and continuous frequency distributions using MS Excel 2007.

In the next lab session, we shall explain correlation analysis for studying the linear relationship between two or more variables using Excel 2007.

Objectives

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

- prepare the spreadsheet in MS Excel 2007;
- determine the range, quartile deviation and its coefficients;
- compute the mean deviation, variance, standard deviation and coefficient of variation; and
- calculate the moments, skewness and kurtosis.

7.2 PROBLEM DESCRIPTION

In this lab session, we state two problems to illustrate the measures of dispersion, moments, skewness and kurtosis of the given data:

1. We consider the discrete frequency distribution constructed in Sec. 2.4 of Lab Session 2 and compute
   - i) the range, quartile deviation, and its coefficients,
   - ii) the mean deviation, variance, standard deviation and coefficient of variation, and
   - iii) the moments, skewness and kurtosis.
2. We consider the raw data of Table 3 of Lab Session 2 as well as continuous frequency distribution constructed in Sec. 2.5 of Lab Session 2 and compute
   - i) the range, quartile deviation, and its coefficients,
   - ii) the mean deviation, variance, standard deviation and coefficient of variation, and
   - iii) the moments, skewness and kurtosis.

7.3 RANGE

You have learnt in Unit 2 of MST-002 that range is the absolute measure of dispersion and coefficient of range is a relative measure of dispersion. In this section, we briefly mention the main points for calculating range as follows:

- We determine the minimum and maximum value of the variable for the given data.
- Range is the difference between the maximum and the minimum value of the variable. It can be defined as

$$R = X_{Max} - X_{Min} \qquad \ldots (1)$$

where X_{Max} – maximum value of the variable, and

X_{Min} – minimum value of the variable.

- The coefficient of range is given by

$$\text{Cofficient of range} = \frac{X_{Max} - X_{Min}}{X_{Max} + X_{Min}} \quad \ldots (2)$$

Steps in Excel

In order to compute the range and the coefficient of range for the raw data in MS Excel 2007, we consider Problem 2 and follow the steps given below.

7.3.1 For Raw Data

Step 1: We enter the data given in Table 3 of Lab Session 2, in Excel 2007 spreadsheet as shown in Fig. 7.1.

	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	
12	11	957	
13	12	745	

Fig. 7.1: Partial screenshot of the spreadsheet for the given data.

Step 2: We compute the maximum value of the given data in Cell B152 using the ***Max*** function as explained in Sec. 2.5 of Lab Session 2 (Fig. 7.2).

B152 =MAX(B2:B151)

	A	B	C
150	149	1115	
151	150	575	
152	Max	1298	
153			

Fig. 7.2

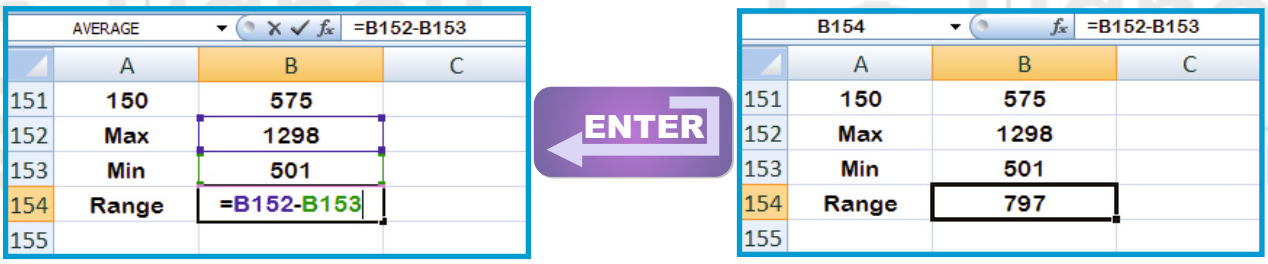
Step 3: We compute the minimum value of the given data in Cell B153 using the ***Min*** function as explained in Sec. 2.5 of Lab Session 2 (Fig. 7.3).

B153 =MIN(B2:B151)

	A	B	C
151	150	575	
152	Max	1298	
153	Min	501	
154			

Fig. 7.3

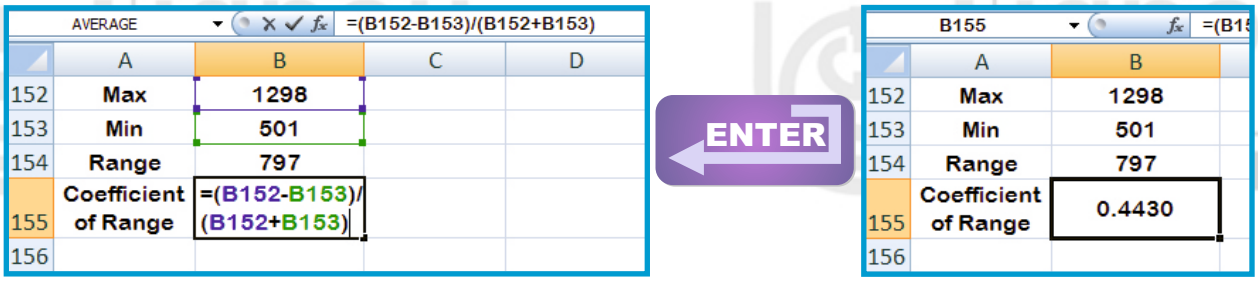
Step 4: To compute the range of the given data, we use equation (1), type "=B152-B153" in Cell B154 and press ***Enter*** as shown in Fig. 7.4.



(a) (b)

Fig. 7.4

Step 5: To compute the coefficient of range of the given data using equation (2), we type "=(B152-B153)/(B152+B153)" in Cell B155 and press ***Enter*** as shown in Fig. 7.5.



(a) (b)

Fig. 7.5

You have learnt how to compute the range and its coefficient for the raw data. You should try to compute the range and its coefficients for discrete as well as continuous frequency distributions before going to the next section.

Note: 1. For discrete frequency distribution, the minimum and maximum value will be the lowest and largest values, respectively, taken by the variable.

2. The minimum and maximum value for the continuous frequency distribution will be the lower limit of the first class interval and the upper limit of the last class interval, respectively.

Activity 1

Compute the range and coefficient of range for the discrete and continuous frequency distributions constructed in Secs. 2.4 and 2.5, respectively, of Lab Session 2 with the help of MS Excel 2007.

7.4 QUARTILE DEVIATION

In Unit 2 of MST-002, you have studied the absolute and relative measures of dispersion which are based on the first and third quartiles. In this section, we mention the main points as follows:

- If Q_1 and Q_3 are the first and third quartiles, respectively, the inter quartile range is calculated as

$$\text{Inter quartile range} = (Q_3 - Q_1) \qquad \ldots (3)$$

- The quartile deviation (QD) or the semi inter quartile range is an absolute measure of dispersion and is calculated as

$$\text{Quartile deviation (QD)} = (Q_3 - Q_1)/2 \qquad \ldots (4)$$

- The relative measure of QD is known as the coefficient of QD and is defined as

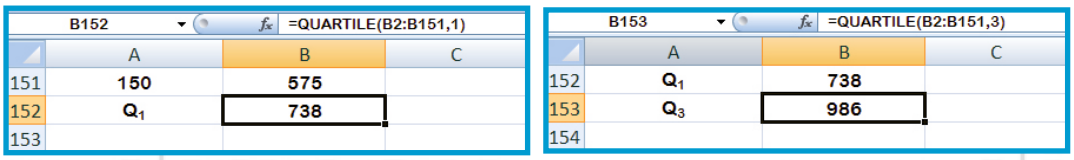
$$\text{Cofficient of QD} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \qquad \ldots (5)$$

Here we describe the computation of the quartile deviation and its coefficient in Excel 2007 for raw data.

7.4.1 For Raw Data

We consider Problem 2 and explain the quartile deviation and its coefficient for raw data in MS Excel 2007. In order to compute the quartile deviation and its coefficient, we follow the steps given below:

Step 1: We compute the 1st and 3rd quartiles in Cells B152 and B153, respectively, as explained in Sec. 6.7 of Lab Session 6 (see Fig. 7.6).



(a) (b)

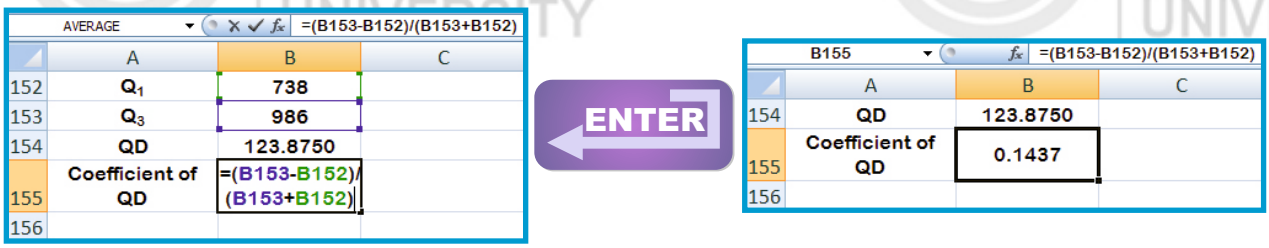
Fig. 7.6

Step 2: The formula for quartile deviation is given in equation (4). The values of Q_1 and Q_3 are calculated in Cells B152 and B153, respectively. So we type "=(B153-B152)/2" in Cell B154 and press ***Enter*** to compute the quartile deviation as shown in Fig. 7.7.

(a) (b)

Fig. 7.7

Step 3: To compute the coefficient of quartile deviation, we type "=(B153-B152)/(B153+B152)" in Cell B155 and press ***Enter*** as shown in Fig. 7.8.



(a) (b)

Fig. 7.8

You have learnt how to compute the quartile deviation and its coefficient for the raw data. You should try to compute the quartile deviation and its coefficients for

discrete as well as continuous frequency distributions before going to the next section.

Activity 2

Compute the quartile deviation and coefficient of quartile deviation for the discrete as well as continuous frequency distributions constructed in Secs. 2.4 and 2.5, respectively, of Lab Session 2 with the help of MS Excel 2007.

7.5 MEAN DEVIATION

You have learnt in Unit 2 of MST-002 that the range and quartile deviation are the positional measures of dispersion and these measures are not based on all observations. Also these do not measure the scatter in observations around an average. To overcome this drawback, we use mean deviation. Here we briefly mention the main points to calculate the mean deviation as follows:

- We consider any arbitrary value around which the mean deviation is to be computed since the mean deviation is defined as the average of the sum of the absolute values of deviation from any arbitrary value, viz., mean, median, mode, etc.
- **For Raw Data**

  The deviation of an observation x_i from the assumed arbitrary value, say, A, is defined as $(x_i - A)$. Therefore, the mean deviation of n observations, say $x_1, x_2, ..., x_n$ is given by

  $$MD = \frac{1}{n}\sum_{i=1}^{k}|x_i - A| \qquad \ldots (6)$$

  where A – any arbitrary value, e.g., mean, median, mode, etc.
- **For Frequency Distributions**

  If $x_1, x_2, ..., x_k$ are k observations (or mid values in case of class intervals) with corresponding frequencies $f_1, f_2, ..., f_k$, respectively, the mean deviation is defined as

  $$MD = \frac{\sum_{i=1}^{k} f_i|x_i - A|}{N} \quad \text{where } N = \sum_{i=1}^{k} f_i. \qquad \ldots (7)$$

It is often suggested that mean deviation be calculated from the median because it has the least value when measured from the median.

Here we describe the computation of the mean deviation about mean in Excel 2007 for raw data, discrete and continuous frequency distribution.

7.5.1 For Raw Data

We consider Problem 2 and explain the mean deviation in case of raw data in Excel 2007. In order to compute the mean deviation about mean, we follow the steps given ahead:

Step 1: We enter the data of Table 3 of Lab Session 2 in an Excel spreadsheet and compute the arithmetic mean in Cell B152 as

explained in Sec. 6.3 of Lab Session 6. The output is shown in Fig. 7.9.

B152 =AVERAGE(B2:B151)

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

Fig. 7.9

The formula with dollar sign ($) is used for an absolute reference.

Step 2: We compute the value of the deviation of the first sample from the mean, i.e., $(x-\bar{x})$ by typing "=B2-B152" in Cell C2 as shown in Fig. 7.10a. Note that here we are using absolute reference, i.e., BB152 to fix Cell B152 as it contains the value of the arithmetic mean. We drag down Cell C2 up to Cell C151 as shown in Fig. 7.10b.

C2 =B2-B152

	A	B	C	D
1	S.No.	Life of Bulb (in hours) (x)	$(x-\bar{x})$	
2	1	1087	218.9400	
3	2	1289		

(a)

DRAG IT DOWN

C2 =B2-B152

	A	B	C	D
1	S.No.	Life of Bulb (in hours) (x)	$(x-\bar{x})$	
2	1	1087	218.9400	
3	2	1289	420.9400	
4	3	876	7.9400	
5	4	725	-143.0600	
6	5	900	31.9400	
7	6	1080	211.9400	
8	7	952	83.9400	
9	8	741	-127.0600	
10	9	1000	131.9400	
11	10	900	31.9400	

(b)

Fig. 7.10

Step 3: To compute the absolute deviation from the mean for the first sample, we refer to Fig. 7.11 and

1. select Cell D2,
2. click on the ***Formulas*** tab, and
3. click on the ***Math & Trig*** → ***Abs***.

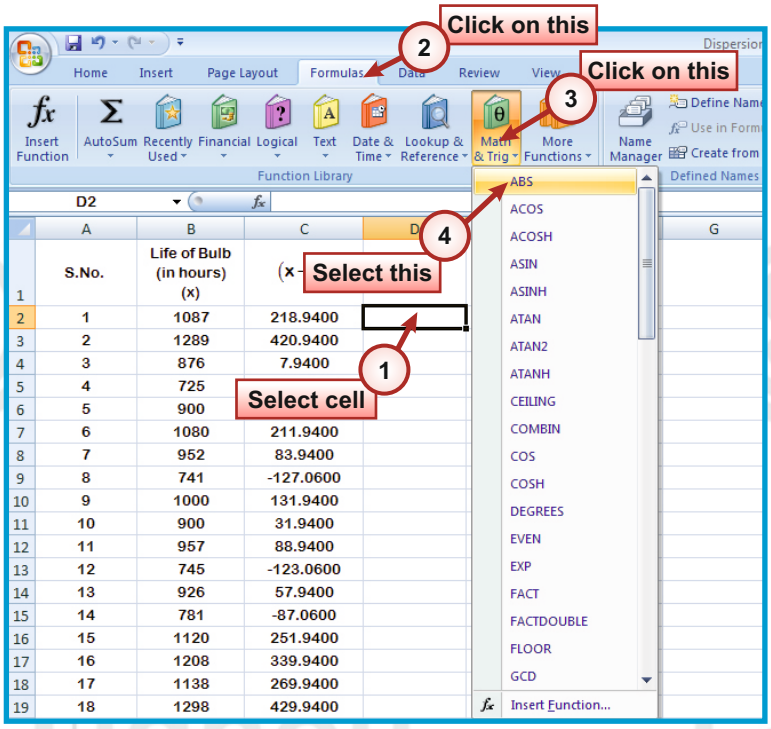
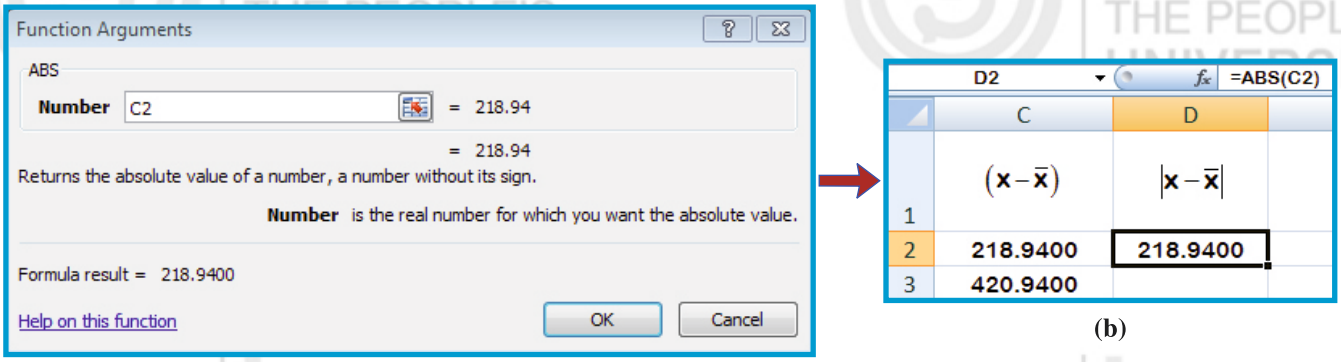


Fig. 7.11

Step 4: A new dialog box opens. We select Cell C2 and click on ***OK*** as shown in Fig. 7.12a. The value of the absolute deviation from the mean for the first sample is obtained in Cell D2 as shown in Fig. 7.12b.



(a) (b)

Fig. 7.12

Step 5: We drag down Cell D2 up to Cell D151 to determine the absolute deviation from the mean for the remaining samples (Fig. 7.13).

	C	D	E
1	$(x-\bar{x})$	$\lvert x-\bar{x}\rvert$	
2	218.9400	218.9400	
3	420.9400	420.9400	
4	7.9400	7.9400	
5	-143.0600	143.0600	
6	31.9400	31.9400	
7	211.9400	211.9400	
8	83.9400	83.9400	
9	-127.0600	127.0600	
10	131.9400	131.9400	
11	31.9400	31.9400	
12	88.9400	88.9400	
13	-123.0600	123.0600	
14	57.9400	57.9400	

Fig. 7.13

Step 6: The mean deviation is the average of all absolute deviations from the mean obtained in Cells D2:D151 [equation (6)]. So we use the ***Average*** function in Cell B153 and select Cells D2:D151 as explained in Sec. 6.3 of Lab Session 6 to compute the mean deviation about the mean as shown in Fig. 7.14.

B153 =AVERAGE(D2:D151)

	A	B	C	D
151	150	575	-293.0600	293.0600
152	$\bar{x}$	868.0600		
153	MD	144.0373		
154				

Fig. 7.14

Step 7: We can also directly compute the mean deviation about mean using ***Avedev*** function as shown in Fig. 7.15. For this, we

1. select Cell B154,
2. click on the ***Formulas*** tab, and
3. click on ***More Functions*** → ***Statistical*** → ***Avedev***.

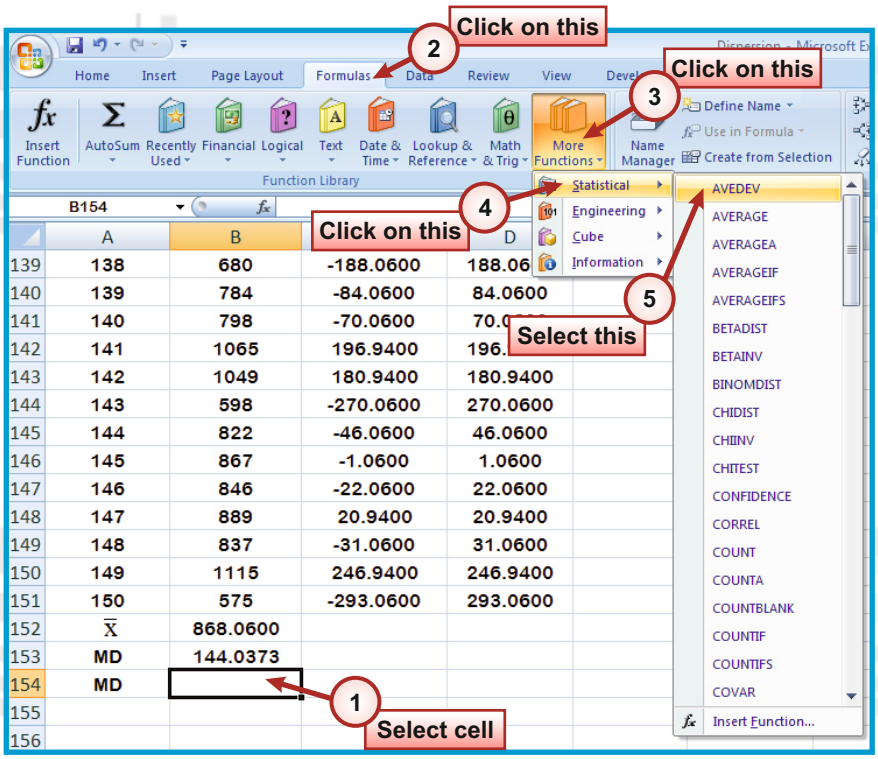
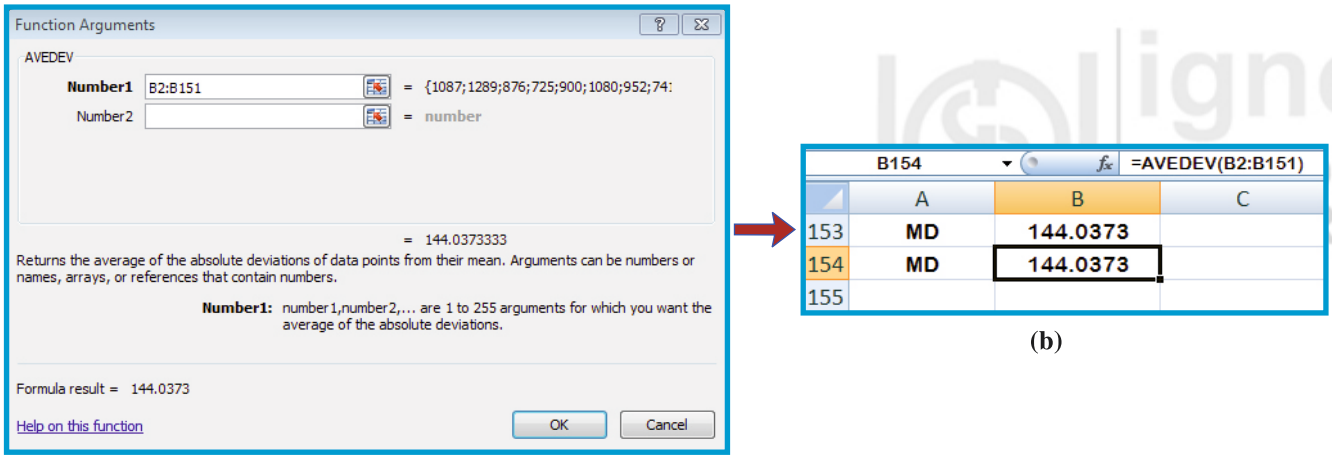


Fig. 7.15

Step 8: A new dialog box opens. We select Cells B2:B151 and click on ***OK*** as shown in Fig. 7.16a. We get the mean deviation about the mean in Cells B154 as shown in Fig. 7.16b.



(a) (b)

Fig. 7.16

Note from Fig. 7.16b that the value of mean deviation about the mean computed in Cell B153 is the same as that computed using ***Avedev*** function in Cell B154.

We now explain the computation of the mean deviation for a discrete frequency distribution.

7.5.2 For Discrete Frequency Distribution

We consider Problem 1 and explain the mean deviation for a discrete frequency distribution in MS Excel 2007. Note that Excel does not have any built-in function to compute the mean deviation for the frequency distributions. In order to compute the mean deviation, we follow the steps given below:

Step 1: To compute the mean deviation about the mean for Problem 1, we repeat Steps 1 to 6 of Sec. 6.3.2 of Lab Session 6. The output is shown in Fig. 7.17.

	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	Total	36	120	509	
11	$\bar{x}$	4.2417			
12					

Fig. 7.17

The formula with dollar sign ($) is used for an absolute reference.

Step 2: We compute the deviation in the values from the mean, i.e., $(x-\bar{x})$ by typing "=B2-B11" in Cell E2 and dragging it down up to Cell E9. We also compute the absolute deviation from the mean, i.e., $|x-\bar{x}|$ as explained in Steps 3 to 5 of Sec. 7.5.1. Note that here we are using absolute reference, i.e., B11 to fix Cell B11 as it contains the value of arithmetic mean. The output is shown in Fig. 7.18.

	A	B	C	D	E	F
1	S.No.	Number of Mobile Phones (x)	No. of Families (f)	fx	$(x-\bar{x})$	$\lvert x-\bar{x}\rvert$
2	1	1	6	6	-3.2417	3.2417
3	2	2	16	32	-2.2417	2.2417
4	3	3	22	66	-1.2417	1.2417
5	4	4	24	96	-0.2417	0.2417
6	5	5	22	110	0.7583	0.7583
7	6	6	16	96	1.7583	1.7583
8	7	7	9	63	2.7583	2.7583
9	8	8	5	40	3.7583	3.7583

Fig. 7.18

Step 3: We now type "=C2*F2" in Cell G2 to determine the multiplication of f and $|x-\bar{x}|$ as shown in Fig. 7.19a. We drag Cell G2 down up to Cell G9 as shown in Fig. 7.19b.

(a)

G2 f_x =C2*F2

	C	D	E	F	G	H
1	No. of Families (f)	fx	$(x-\bar{x})$	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$	
2	6	6	-3.2417	3.2417	19.4500	
3	16	32	-2.2417	2.2417		
4	22	66	-1.2417	1.2417		
5	24	96	-0.2417	0.2417		
6	22	110	0.7583	0.7583		

DRAG IT DOWN

(b)

	C	D	E	F	G	H
1	No. of Families (f)	fx	$(x-\bar{x})$	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$	
2	6	6	-3.2417	3.2417	19.4500	
3	16	32	-2.2417	2.2417	35.8667	
4	22	66	-1.2417	1.2417	27.3167	
5	24	96	-0.2417	0.2417	5.8000	
6	22	110	0.7583	0.7583	16.6833	
7	16	96	1.7583	1.7583	28.1333	
8	9	63	2.7583	2.7583	24.8250	
9	5	40	3.7583	3.7583	18.7917	

Fig. 7.19

Step 4: In Cell G10, we compute $\sum f|x-\bar{x}|$ using "=Sum(G2:G9)" function as explained in Step 7 of Sec. 2.3 of Lab Session 2 (see Fig. 7.20).

G10 =SUM(G2:G9)

	D	E	F	G	H
1	fx	$(x-\bar{x})$	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$	
2	6	-3.2417	3.2417	19.4500	
3	32	-2.2417	2.2417	35.8667	
4	66	-1.2417	1.2417	27.3167	
5	96	-0.2417	0.2417	5.8000	
6	110	0.7583	0.7583	16.6833	
7	96	1.7583	1.7583	28.1333	
8	63	2.7583	2.7583	24.8250	
9	40	3.7583	3.7583	18.7917	
10	509		$\sum f\lvert x-\bar{x}\rvert$	176.8667	
11					

Fig. 7.20

Step 5: We use equation (7) to compute the mean deviation about the mean, type "=G10/C10" in Cell B12 (Fig. 7.21a) and press ***Enter*** to get the result shown in Fig. 7.21b.

MINVERSE =G10/C10

	A	B	C	D	E	F	G
9	8	8	5	40	3.7583	3.7583	18.7917
10	Total	36	120	509		$\sum f\lvert x-\bar{x}\rvert$	176.8667
11	$\bar{x}$	4.2417					
12	MD	=G10/C10					
13							

(a)

ENTER

	A	B	C	D	E	F	G
9	8	8	5	40	3.7583	3.7583	18.7917
10	Total	36	120	509		$\sum f\lvert x-\bar{x}\rvert$	176.8667
11	$\bar{x}$	4.2417					
12	MD	1.4739					
13							

(b)

Fig. 7.21

We have explained the computation of mean deviation about the mean for discrete frequency distribution. You can now try to compute it for continuous frequency distribution and also about other arbitrary values such as median and mode before going to the next section.

Activity 3

1) Consider the continuous frequency distribution of Sec. 2.5 of Lab Session 2 and compute mean deviation about the mean in the same way as explained in Sec. 7.5.2.

2) Compute the median and mode as explained in Lab Session 6 for Problems 1 and 2 and also determine the mean deviation about the median and mode.

7.6 VARIANCE, STANDARD DEVIATION AND COEFFICIENT OF VARIATION

In Unit 2 of MST-002, you have learnt that variance and standard deviation are the most popular measures of dispersion. Here we briefly mention the main points for calculating these measures as follows:

For Raw Data

- Variance is the average of the square of deviations of the values from mean. If we have n observations $x_1, x_2, ..., x_n$, variance can be calculated as

$$\text{Var}(x) = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \ldots (8)$$

- Standard deviation (SD) is defined as the positive square root of the variance and is given by

$$\text{SD} = \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \qquad \ldots (9)$$

For Frequency Distributions

- If we have k observations (or mid values in case of continuous frequency distribution) $x_1, x_2, ..., x_k$ with corresponding frequencies $f_1, f_2, \ldots, f_k$, respectively, the variance can be computed as

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{k} f_i(x_i - \bar{x})^2 \quad \text{where } N = \sum_{i=1}^{k} f_i. \qquad \ldots (10)$$

- The formula for the standard deviation is given by

$$\text{SD} = \sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i(x_i - \bar{x})^2}{N}} \qquad \ldots (11)$$

Coefficient of Variation

- We use coefficient of variation as it is a relative measure of variability to compare two data series. We define the coefficient of variation (CV) as

$$\text{CV} = \frac{\sigma}{\bar{x}} \times 100 \qquad \ldots (12)$$

Here we describe the computation of variance and standard deviation in Excel 2007 for raw data, discrete and continuous frequency distributions. We first consider Problem 2 and explain how to compute variance and standard deviation for raw data in MS Excel 2007.

7.6.1 For Raw Data

In order to compute the variance and standard deviation for raw data, we follow the steps given ahead:

Step 1: We enter the data of Table 3 of Lab Session 2 in an Excel spreadsheet. To compute the variance for the given data, we refer to Fig. 7.22 and

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

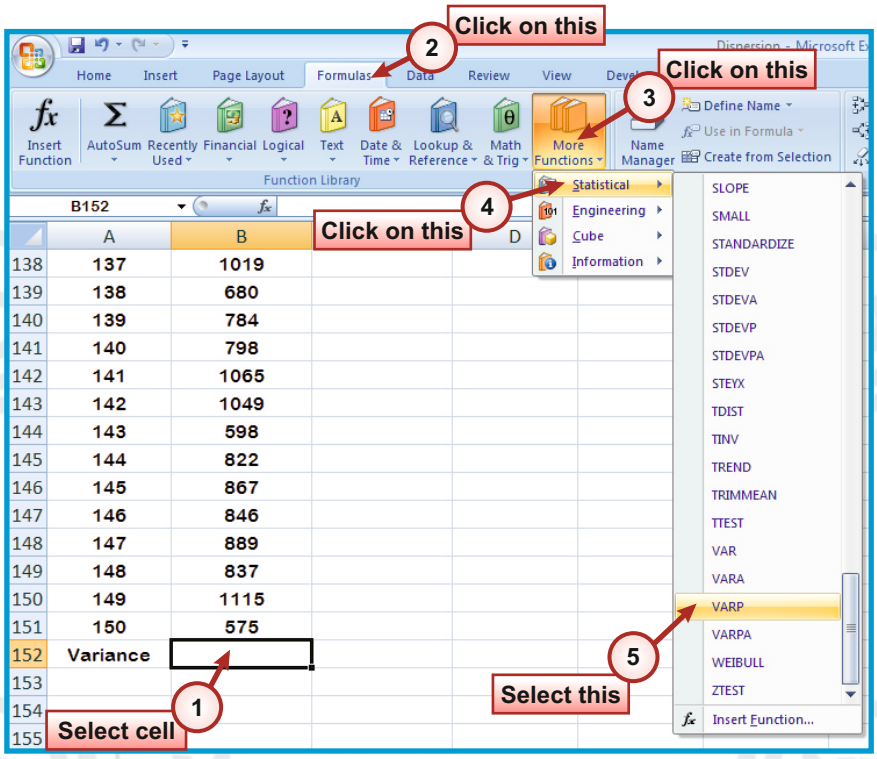
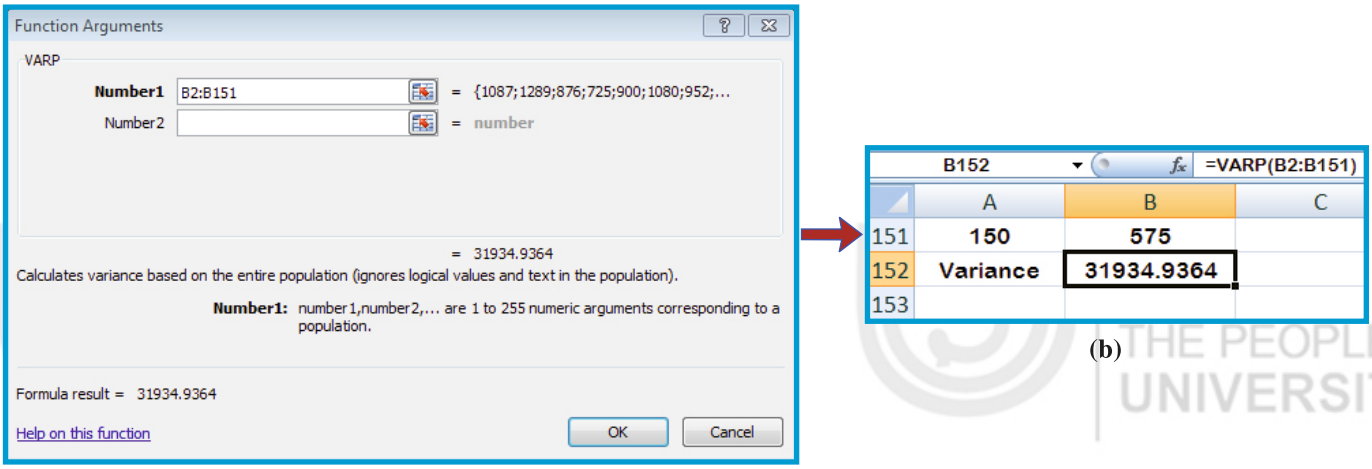


Fig. 7.22

Here we are using n in the denominator to calculate variance, so we use ***Varp*** function in Excel. If we have (n – 1) in the denominator for a sample as discussed in MST-004, we use ***Var*** function in Excel.

Step 2 : A new dialog box appears. We select Cells B2:B151and click on ***OK*** as shown in Fig. 7.23a. We obtain the value of variance in Cell B152 as shown in Fig. 7.23b.



(a) (b)

Fig. 7.23

Step 3: We can compute the standard deviation of the given data directly by taking the square root of variance. For this, we refer to Fig. 7.24 and

1. select Cell B153,
2. click on the ***Formulas*** tab, and
3. click on the ***Math & Trig*** → ***Sqrt***.

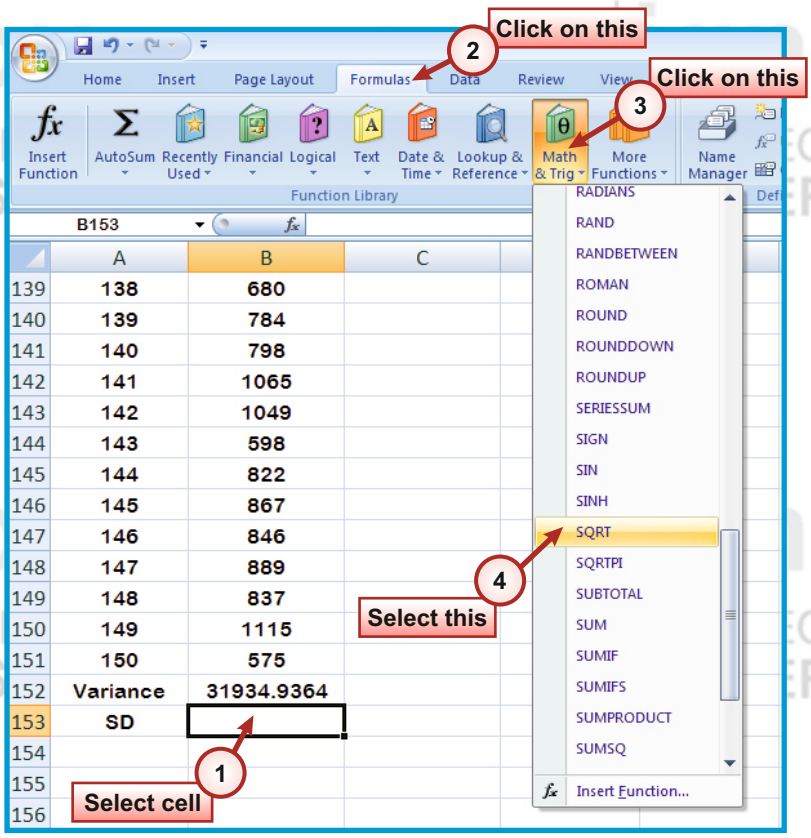
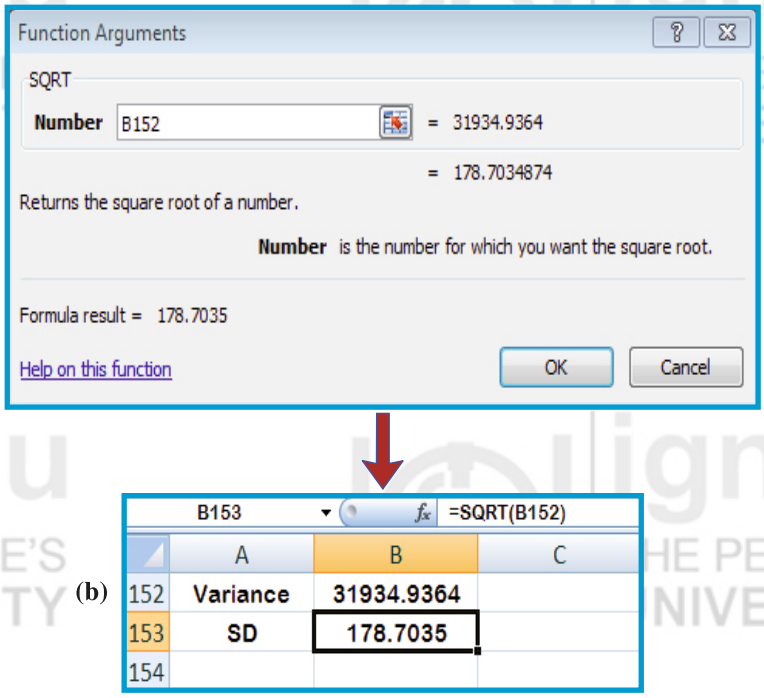


Fig. 7.24

Step 4: A new dialog box opens. We select Cells B152 and click on ***OK*** as shown in Fig. 7.25a. The value of the standard deviation is shown in Fig. 7.25b.



(a)

(b)

Fig. 7.25

Step 5: We can also directly compute the standard deviation using ***Stdevp*** function as shown in Fig. 7.26. For this, we

1. select Cells B154,
2. click on the ***Formulas*** tab, and
3. click on ***More Functions*** → ***Statistical*** → ***Stdevp***.

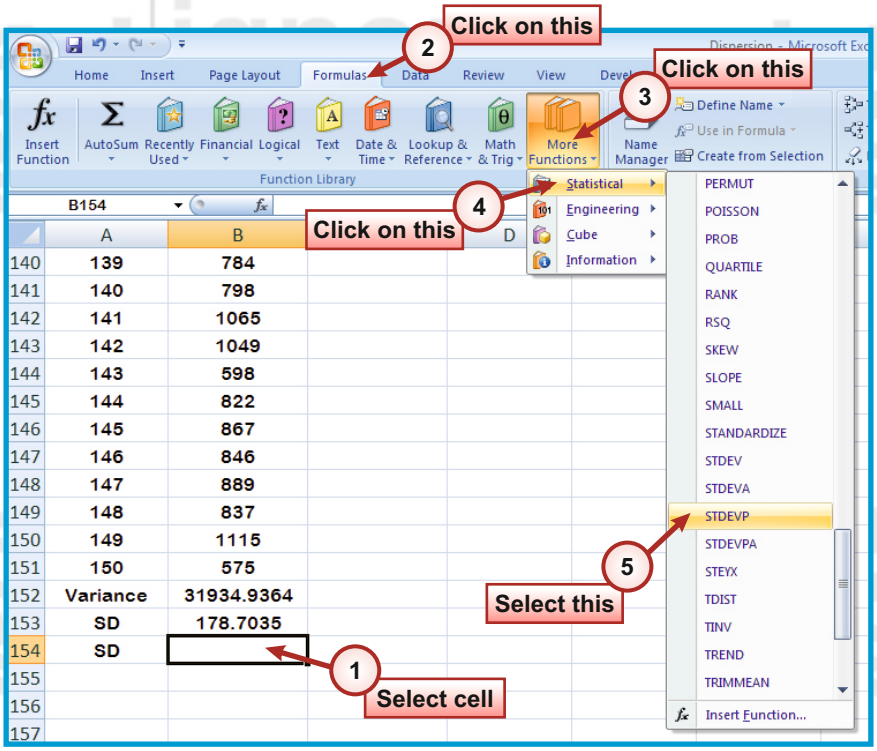
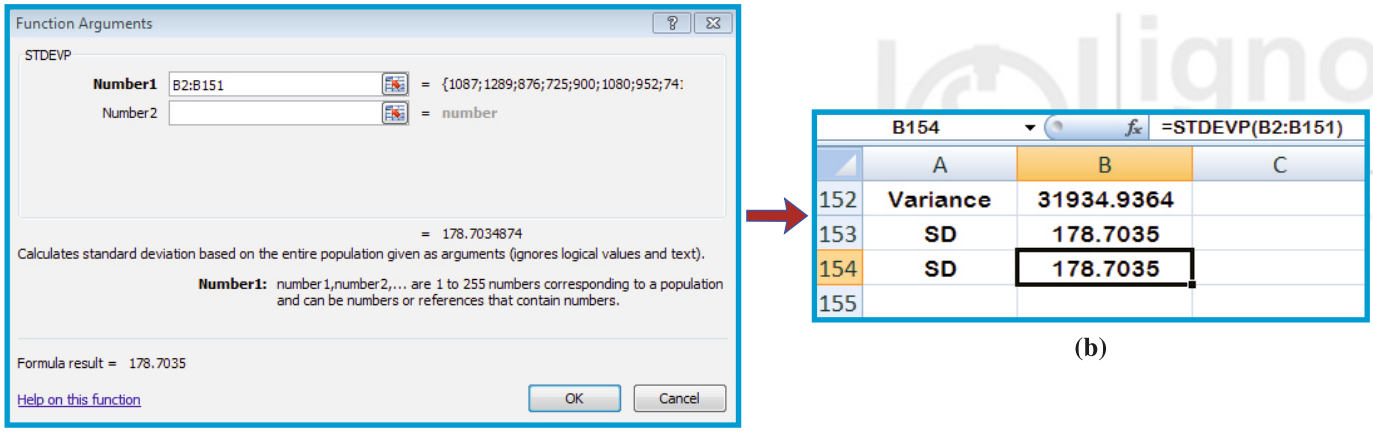


Fig. 7.26

It should be noted that sum of squares of deviations is least when deviations are measured from the mean. This means $\Sigma(x_i - A)^2$ is least when A = Mean.

Step 6: A new dialog box appears. We select Cells B2:B151and click on ***OK*** as shown in Fig. 7.27a. We get the standard deviation in Cell B154 as shown in Fig. 7.27b.



(a) (b)

Fig. 7.27

Note from Fig. 7.27b that the value of standard deviation computed manually in Cell B153 is the same as that computed using ***Stdevp*** function in Cell B154.

The values of SD, MD and variance cannot be negative.

Step 7: To determine the coefficient of variation of the given data, we need to compute the arithmetic mean as explained in Sec.6.3 of Lab session 6. We calculate the arithmetic mean in Cells B155 as shown in Fig. 7.28.

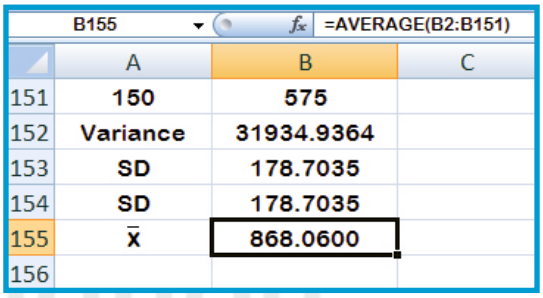


Fig. 7.28

Step 8: To determine the coefficient of variation of the given data, we use equation (12), type "=(B154/B155)*100" in Cell B156 (Fig. 7.29a) and press ***Enter*** as shown in Fig. 7.29b.

(a)

MINVERSE | =(B154/B155)*100

	A	B	C
153	SD	178.7035	
154	SD	178.7035	
155	$\bar{x}$	868.0600	
156	CV	=(B154/B155)*100	
157			

ENTER

(b)

B156 | =(B154/B155)*100

	A	B	C
154	SD	178.7035	
155	$\bar{x}$	868.0600	
156	CV	21	
157			

Fig. 7.29

We now consider the computation of variance and standard deviation for a discrete frequency distribution.

7.6.2 For Discrete Frequency Distribution

We consider Problem 1 and explain the variance and standard deviation for a discrete frequency distribution in MS Excel 2007. Note that Excel does not have any built-in function to compute variance and standard deviation for frequency distributions. In order to compute variance and standard deviation for discrete frequency distribution, we follow the steps given below:

Step 1: To compute variance or standard deviation for Problem 1, we first calculate the mean of the data by repeating Steps 1 to 6 of Sec. 6.3.2 of Lab Session 6. The output is shown in Fig. 7.30.

	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	Total	36	120	509	
11	$\bar{x}$	4.2417			

Fig. 7.30

The formula with dollar sign ($) is used for an absolute reference.

Step 2: We compute the square of the deviation in the values from the mean, i.e., $(x-\bar{x})^2$ by typing "=(B2-B11)^2" in Cell E2 and pressing ***Enter*** (Fig. 7.31a). Note that here we are using absolute reference, i.e., B11 to fix Cell B11 as it contains the value of the arithmetic mean.

STDEVP | =(B2-B11)^2

	B	C	D	E	F
1	Number of Mobile Phones (x)	No. of Families (f)	fx	$(x-\bar{x})^2$	
2	1	6	6	=(B2-B11)^2	
3	2	16	32		

(a)

ENTER

	B	C	D	E	F
1	Number of Mobile Phones (x)	No. of Families (f)	fx	$(x-\bar{x})^2$	
2	1	6	6	10.5084	
3	2	16	32		

(b)

Fig. 7.31

Step 3: We drag Cell E2 down up to Cell E9 to determine the square of the deviation from the mean for the remaining values as shown in Fig. 7.32.

	B	C	D	E	F
1	Number of Mobile Phones (x)	No. of Families (f)	fx	$(x-\bar{x})^2$	
2	1	6	6	10.5084	
3	2	16	32	5.0251	
4	3	22	66	1.5417	
5	4	24	96	0.0584	
6	5	22	110	0.5751	
7	6	16	96	3.0917	
8	7	9	63	7.6084	
9	8	5	40	14.1251	

Fig. 7.32

Step 4: We now type "=C2*E2" in Cell F2 to determine the product of f and $(x-\bar{x})^2$ as shown in Fig. 7.33a and drag Cell F2 down up to Cell F9 as shown in Fig. 7.33b.

F2 | =C2*E2

	C	D	E	F	G
1	No. of Families (f)	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$	
2	6	6	10.5084	63.0504	
3	16	32	5.0251		
4	22	66	1.5417		
5	24	96	0.0584		

(a)

DRAG IT DOWN

	C	D	E	F	G
1	No. of Families (f)	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$	
2	6	6	10.5084	63.0504	
3	16	32	5.0251	80.4011	
4	22	66	1.5417	33.9182	
5	24	96	0.0584	1.4017	
6	22	110	0.5751	12.6515	
7	16	96	3.0917	49.4678	
8	9	63	7.6084	68.4756	
9	5	40	14.1251	70.6253	

(b)

Fig. 7.33

Step 5: In Cell F10, we compute $\sum f(x-\bar{x})^2$ using "=Sum(F2:F9)" function as explained in Step 7 in Sec. 2.3 of Lab Session 2 (see Fig. 7.34).

F10 =SUM(F2:F9)

	E	F	G
9	14.1251	70.6253	
10	$\sum f(x-\bar{x})^2$	379.9917	
11			

Fig. 7.34

Step 6: We use equation (10) to compute the variance, type "=F10/C10" in Cell B12 (Fig. 7.35a) and press ***Enter*** to get the result (Fig. 7.35b).

(a)

MINVERSE =F10/C10

	A	B	C	D	E	F	G
9	8	8	5	40	14.1251	70.6253	
10	Total	36	120	509	$\sum f(x-\bar{x})^2$	379.9917	
11	$\bar{x}$	4.2417					
12	Variance	=F10/C10					
13							

ENTER

(b)

B12 =F10/C10

	A	B	C
9	8	8	5
10	Total	36	120
11	$\bar{x}$	4.2417	
12	Variance	3.1666	
13			

Fig. 7.35

Step 7: We can compute the standard deviation of the given data directly by taking square root of the variance (see Fig. 7.36) as explained in Steps 3 and 4 of Sec. 7.6.1.

B13 =SQRT(B12)

	A	B	C
10	Total	36	120
11	$\bar{x}$	4.2417	
12	Variance	3.1666	
13	SD	1.7795	
14			

Fig. 7.36

Step 8: To determine the coefficient of variation of the given data, we type "=(B13/B11)*100" in Cell B14 as shown in Fig. 7.37.

B14 =(B13/B11)*100

	A	B	C
11	$\bar{x}$	4.2417	
12	Variance	3.1666	
13	SD	1.7795	
14	CV	42	
15			

Fig. 7.37

You have learnt how to compute the variance, standard deviation and coefficient of variation for raw data as well as discrete frequency distributions. You should try to compute the variance, standard deviation and coefficient of variation for continuous frequency distributions before going to the next section.

Activity 4

Compute the variance, standard deviation and coefficient of variation for the continuous frequency distributions formed in Sec. 2.5 of Lab Session 2 with the help of MS Excel 2007.

7.7 MOMENTS, SKEWNESS AND KURTOSIS

In Units 3 and 4 of MST-002, you have learnt the computation of moments, skewness and kurtosis. Here we briefly mention the main points as follows:

Moments about the origin

- If we have n observations $x_1, x_2, ..., x_n$, we define the rth order moment about the origin (also known as raw moments) for ungrouped or raw data as

$$\mu'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r \qquad \ldots (13)$$

- If we have k observations (or mid values in case of class intervals) $x_1, x_2, ..., x_k$, with corresponding frequencies $f_1, f_2, \ldots, f_k$, respectively, the rth order moment about the origin for discrete or continuous frequency distributions is given by

$$\mu'_r = \frac{1}{k}\sum_{i=1}^{n} f_i x_i^r \qquad \ldots (14)$$

Moments about the mean

- The first four moments about the mean are also known as central moments and can be obtained with the help of the moments about the origin as follows :

$$\mu = 0 \qquad \ldots (15a)$$

$$\mu_2 = \mu'_2 - \mu'^2_1 \qquad \ldots (15b)$$

$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1 \qquad \ldots (15c)$$

$$\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'^2_1 - 3\mu'^4_1 \qquad \ldots (15d)$$

Coefficient of Skewness

- Karl Pearson defined two coefficients of skewness, β_1 and γ_1, based upon the central moments. These are given as follows:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \qquad \ldots (16)$$

Since β_1 does not give the direction of skewness, i.e., whether it is positive or negative, another measure of skewness (γ_1) is defined as

$$\gamma_1 = \pm\sqrt{\beta_1} \qquad \ldots (17)$$

The sign of skewness would depend upon whether μ_3 is positive or negative.

- **Description:**
  1. If $\gamma_1 = 0$, the distribution is symmetric.
  2. If $\gamma_1 < 0$, the distribution is negatively skewed.
  3. If $\gamma_1 > 0$, the distribution is positively skewed.

Coefficient of Kurtosis

- Karl Pearson also defined two coefficients of kurtosis β_2 and γ_2 based upon the central moments:

  β_2 measures kurtosis and is defined as

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \qquad \ldots (18)$$

- The other measure of kurtosis (γ_2) is defined as

$$\gamma_2 = \beta_2 - 3 \qquad \ldots (19)$$

- **Description:**

  If $\beta_2 = 3$ or $\gamma_2 = 0$, the distribution is mesokurtic.

  If $\beta_2 < 3$ or $\gamma_2 < 0$, the distribution is platykurtic.

  If $\beta_2 > 3$ or $\gamma_2 > 0$, the distribution is leptokurtic.

Here we describe the computation of moments, skewness and kurtosis using Excel 2007 for raw data, discrete and continuous frequency distributions.

7.7.1 For Raw Data

We consider Problem 2 and explain the computation of moments, skewness and kurtosis for raw data using MS Excel 2007. In order to compute the moments, skewness and kurtosis, we follow the steps given below:

Step 1: We enter the data of Table 3 of Lab Session 2 in an Excel spreadsheet. We type "=B2^2" in Cell C2 (Fig. 7.38a), "=B2^3" in Cell D2 (Fig. 7.38b) and "=B2^4" in Cell E2 (Fig. 7.38c) to compute the values of x^2, x^3 and x^4, respectively, for the first sample.

(a)

C2 | =B2^2

	B	C	D
1	Life of Bulb (in hours) (x)	x^2	
2	1087	1181569	
3	1289		

(b)

D2 | =B2^3

	C	D	E
1	x^2	x^3	
2	1181569	1284365503	
3			

(c)

E2 | =B2^4

	D	E	F
1	x^3	x^4	
2	1284365503	1396105301761	
3			

Fig. 7.38

Step 2: We select Cell C2:E2 as shown in Fig. 7.39a and drag down these cells up to Row 151 to determine the values of x^2, x^3 and x^4 for the remaining samples as shown in Fig. 7.39b.

	B	C	D	E	F
1	Life of Bulb (in hours) (x)	x^2	x^3	x^4	
2	1087	1181569	1284365503	1396105301761	
3	1289				
4	876				

(a)

DRAG THEM DOWN

	B	C	D	E	F
1	Life of Bulb (in hours) (x)	x^2	x^3	x^4	
2	1087	1181569	1284365503	1396105301761	
3	1289	1661521	2141700569	2760652033441	
4	876	767376	672221376	588865925376	
5	725	525625	381078125	276281640625	
6	900	810000	729000000	656100000000	
7	1080	1166400	1259712000	1360488960000	
8	952	906304	862801408	821386940416	
9	741	549081	406869021	301489944561	
10	1000	1000000	1000000000	1000000000000	
11	900	810000	729000000	656100000000	
12	957	915849	876467493	838779390801	
13	745	555025	413493625	308052750625	
14	926	857476	794022776	735265090576	

(b)

Fig. 7.39

Step 3: To get the first four moments about the origin (μ'_1, μ'_2, μ'_3 and μ'_4) for the raw data, we use equation (13) and compute the average of the series x, x^2, x^3 and x^4 given in Cells B2:B151, Cells C2:C151, Cells D2:D151 and Cells E2:E151, respectively, using ***Average*** function in Cells B153 to B156, as explained in Sec. 6.3 of Lab Session 6 (Fig. 7.40).

B153 =AVERAGE(B2:B151)

	A	B	C
152			
153	μ'_1	868.0600	
154			

(a)

B154 =AVERAGE(C2:C151)

	A	B	C
153	μ'_1	868.0600	
154	μ'_2	785463.1000	
155			

(b)

B155 =AVERAGE(D2:D151)

	A	B	C
154	μ'_2	785463.1000	
155	μ'_3	738499562.6200	
156			

(c)

B156 =AVERAGE(E2:E151)

	A	B	C
155	μ'_3	738499562.6200	
156	μ'_4	719029405462.1400	
157			

(d)

Fig. 7.40

Step 4: You have studied that the first moment about the mean is always zero [equation (15a)]. So we type "0" in Cell B157 (Fig. 7.41).

B157 0

	A	B	C
156	μ'_4	719029405462.1400	
157	μ_1	0	
158			

Fig. 7.41

Step 5: We use equation (15b) and type "=B154-B153^2" in Cell B158 to determine the second moment about the mean as shown in Fig. 7.42.

B158 f_x =B154-B153^2

	A	B	C
157	μ_1	0	
158	μ_2	31934.9364	
159			

Fig. 7.42

Step 6: We use equation (15c) and type "=B155-3*B154*B153+2*B153^3" in Cell B159 to determine the third moment about the mean as shown in Fig. 7.43.

B159 f_x =B155-3*B154*B153+2*B153^3

	A	B	C
158	μ_2	31934.9364	
159	μ_3	1227582.2512	
160			

Fig. 7.43

Step 7: We use equation (15d) and type "=B156-4*B155*B153+6*B154*B153^2-3*B153^4" in Cell B160 to determine the fourth moment about mean (Fig. 7.44).

B160 f_x =B156-4*B155*B153+6*B154*B153^2-3*B153^4

	A	B	C
159	μ_3	1227582.2512	
160	μ_4	2579008046.6851	
161			

Fig. 7.44

Step 8: We use equations (16) and (17) and type "=(B159^2)/(B158^3)" and "=Sqrt(B161)" in Cells B161 and B162 to determine the values of β_1 and γ_1, respectively, as shown in Figs. 7.45a and b.

B161 f_x =(B159^2)/(B158^3)

	A	B	C
160	μ_4	2579008046.6851	
161	β_1	0.0463	
162			

(a)

B162 f_x =SQRT(B161)

	A	B	C
161	β_1	0.0463	
162	γ_1	0.2151	
163			

(b)

Fig. 7.45

Notice from Fig. 7.45 that the value of γ_1 is 0.2151 and so the data for the life of bulbs are slightly positively skewed.

Step 9: We use equations (18) and (19) and type "=(B160)/(B158^2)" and "=B163-3" in Cells B163 and B164 to determine the values of β_2 and γ_2, respectively, as shown in Figs. 7.46a and b.

B163 f_x =(B160)/(B158^2)

	A	B	C
162	γ_1	0.2151	
163	β_2	2.5288	
164			

(a)

B164 f_x =B163-3

	A	B	C
163	β_2	2.5288	
164	γ_2	-0.4712	
165			

(b)

Fig. 7.46

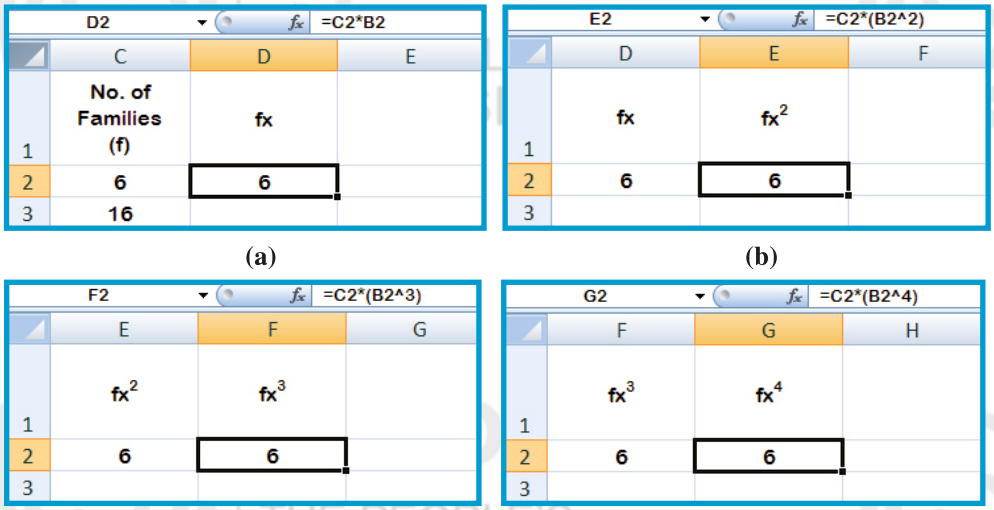
Notice from Fig. 7.46b that the value of γ_2 is – 0.4712 and so the data for the life of bulbs are platykurtic.

We now consider the computation of moments, skewness and kurtosis for discrete frequency distribution.

7.7.2 For Discrete Frequency Distribution

We consider Problem 1 and explain how to compute the moments, skewness and kurtosis for discrete frequency distribution using MS Excel 2007. In order to compute the moments, skewness and kurtosis, we follow the steps given below:

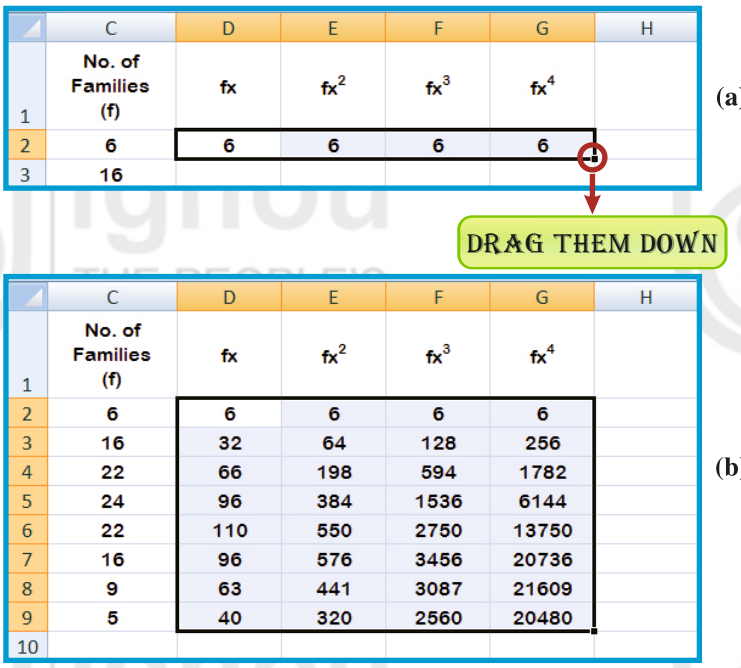
Step 1: We consider the discrete frequency distribution formed in Sec. 2.4 of Lab Session 2. We type "=C2*B2" in Cell D2 (Fig. 7.47a), "=C2*(B2^2)" in Cell E2 (Fig. 7.47b), "=C2*(B2^3)" in Cell F2 (Fig. 7.47c) and "=C2*(B2^4)" in Cell G2 (Fig. 7.47d) to compute the values of fx, fx^2, fx^3 and fx^4, respectively, for the first sample.



(a) (b) (c) (d)

Fig. 7.47

Step 2: We select Cell D2:G2 as shown in Fig. 7.48a and drag down these cells up to Row 9 to determine the values of fx, fx^2, fx^3 and fx^4 for the remaining samples as shown in Fig. 7.48b.



	C	D	E	F	G	H
1	No. of Families (f)	fx	fx^2	fx^3	fx^4	
2	6	6	6	6	6	
3	16					

(a)

	C	D	E	F	G	H
1	No. of Families (f)	fx	fx^2	fx^3	fx^4	
2	6	6	6	6	6	
3	16	32	64	128	256	
4	22	66	198	594	1782	
5	24	96	384	1536	6144	
6	22	110	550	2750	13750	
7	16	96	576	3456	20736	
8	9	63	441	3087	21609	
9	5	40	320	2560	20480	
10						

(b)

Fig. 7.48

Step 3: We determine the values of $\sum f$ in Cells C10 using the ***Sum*** function as shown in Fig. 7.49a and drag right Cell C10 up to Cell G10 to compute the values of $\sum fx$, $\sum fx^2$, $\sum fx^3$ and $\sum fx^4$, respectively, as shown in Fig. 7.49b.

(a)

C10 | =SUM(C2:C9)

	B	C	D
9	8	5	40
10	Total	120	
11			

DRAG IT RIGHT

(b)

C10 | =SUM(C2:C9)

	B	C	D	E	F	G	H
9	8	5	40	320	2560	20480	
10	Total	120	509	2539	14117	84763	
11							

Fig. 7.49

Step 4: We use equation (14) for r = 1 and type "=D10/C10" in Cell B12 to determine the first moment about the origin (Fig. 7.50).

B12 | =D10/C10

	A	B	C
11			
12	μ_1'	4.2417	
13			

Fig. 7.50

Step 5: We use equation (14) for r = 2 and type "=E10/C10" in Cell B13 to determine the second moment about the origin (Fig. 7.51).

B13 | =E10/C10

	A	B	C
12	μ_1'	4.2417	
13	μ_2'	21.1583	
14			

Fig. 7.51

Step 6: We use equation (14) for r = 3 and type "=F10/C10" in Cell B14 to determine the third moment about the origin (Fig. 7.52).

B14 | =F10/C10

	A	B	C
13	μ_2'	21.1583	
14	μ_3'	117.6417	
15			

Fig. 7.52

Step 7: We use equation (14) for r = 4 and type "=G10/C10" in Cell B15 to determine the fourth moment about the origin (Fig. 7.53).

B15 | =G10/C10

	A	B	C
14	μ_3'	117.6417	
15	μ_4'	706.3583	
16			

Fig. 7.53

Step 8: We compute the first four moments about the mean, i.e., μ_1, μ_2, μ_3 and μ_4 in the same way as explained in Steps 4 to 7 of Sec. 7.7.1.

We also determine the values of β_1, γ_1 , β_2, and γ_2 as explained in Steps 8 and 9 of Sec. 7.7.1 (see Fig. 7.54).

	A	B	C
16	μ_1	0.0000	
17	μ_2	3.1666	
18	μ_3	1.0318	
19	μ_4	23.3146	
20	β_1	0.0335	
21	γ_1	2.3251	
22	β_2	0.1831	
23	γ_2	-0.6749	
24			

Fig. 7.54

Notice from Fig. 7.54 that the value of γ_1 is 2.3251 and so the data of mobile phones are positively skewed. Also the value of γ_2 is -0.6749 and so the data of mobile phones are platykurtic.

You have learnt how to compute the moments, skewness and kurtosis for raw data as well as discrete frequency distributions. You should try to compute the moments, skewness and kurtosis for continuous frequency distributions before going to the next section.

Activity 5

Compute the first four moments about the origin, moments about the mean, β and γ coefficients of skewness and kurtosis for the continuous frequency distributions formed in Sec. 2.5 of Lab Session 2 with the help of MS Excel 2007.

7.8 DESCRIPTIVE STATISTICS USING DATA ANALYSIS TOOLPAK

We can also use the ***Data Analysis ToolPak*** for determining descriptive statistics in Excel. Here we consider Problem 2, i.e., data of Table 3 of Lab Session 2 to explain the procedure. We follow the steps given below:

Step 1: We enter the data of Table 3 of Lab Session 3 in an Excel spreadsheet. To compute descriptive statistics using ***Data Analysis ToolPak*** in Excel 2007, we

1. click on the ***Data*** tab,
2. select ***Data Analysis*** as shown in Fig. 7.55a,
3. select the ***Descriptive Statistics*** option,
4. click on ***OK*** as shown in Fig. 7.55b, and
5. get a new dialog box as shown in Fig. 7.55c.

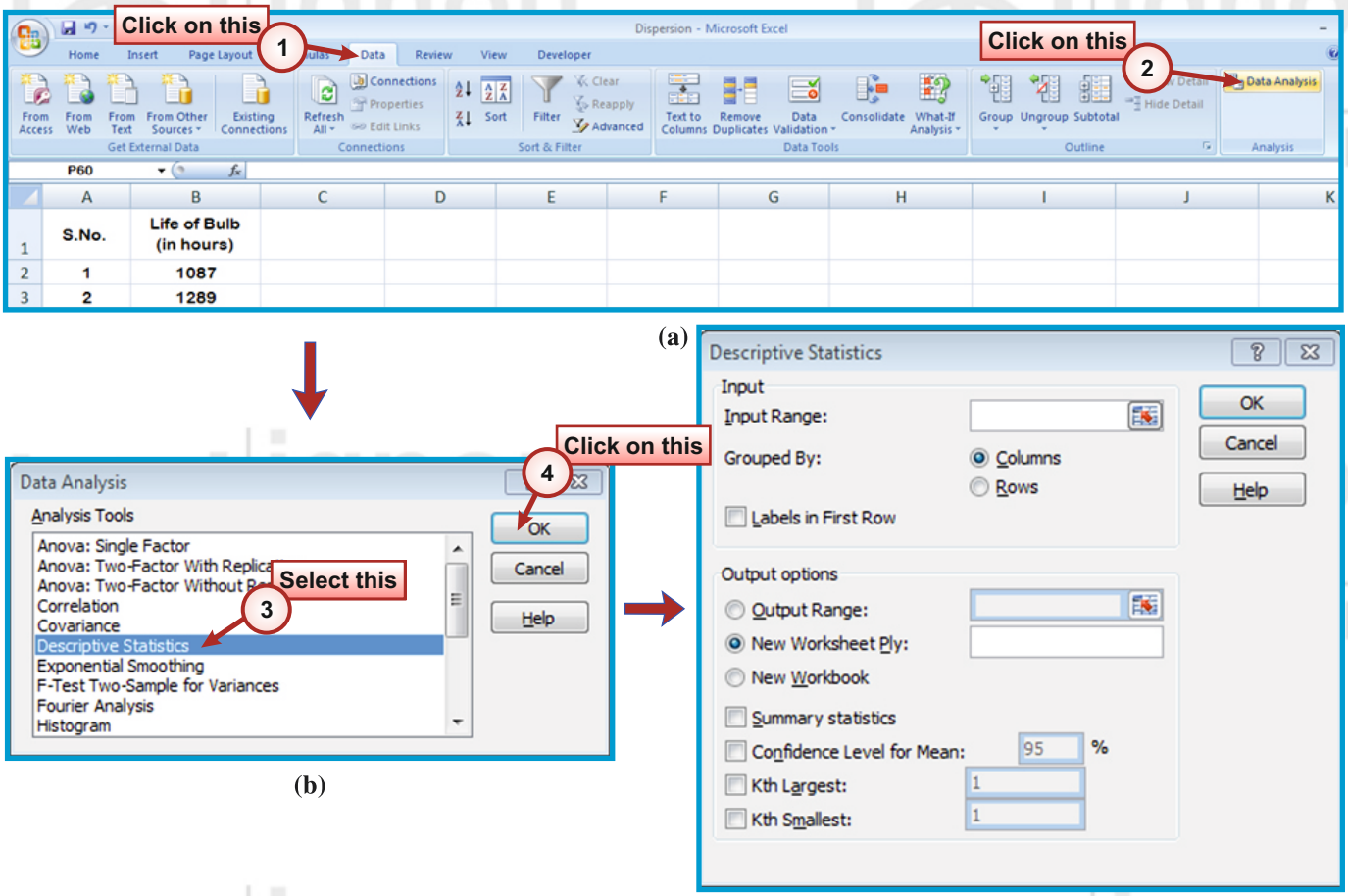


Fig. 7.55

Step 2: As shown in Fig. 7.56, we

1. specify the data with the label **Life of Bulb** in ***Input Range***, i.e., Cells B1:B151,
2. tick on the ***Labels*** box since we have included data labels in the first row, i.e., **Life of Bulb (in hours)** given in Cell B1 in the input range,

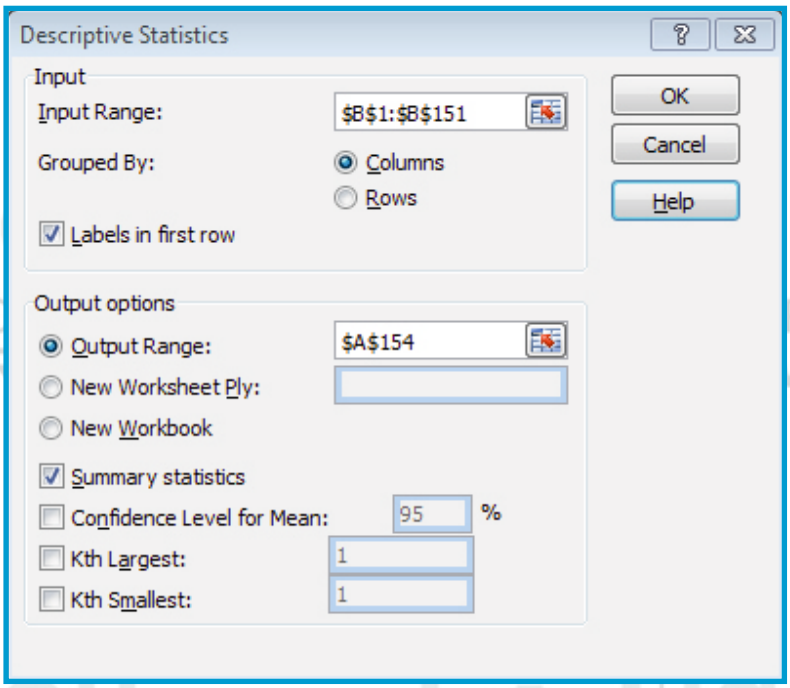


Fig. 7.56

3. provide the ***Output Range*** under the ***Output Options***. Here we select Cell A154 for the output, and
4. tick on the ***Summary statistics*** box and then click on ***OK***.

Step 3: After completing Step 2, we obtain the output shown in Fig. 7.57.

	A	B	C
153			
154	*Life of Bulb (in hours)*		
155			
156	Mean	868.06	
157	Standard Error	14.6400	
158	Median	860.5000	
159	Mode	754	
160	Standard Deviation	179.3022	
161	Sample Variance	32149.2648	
162	Kurtosis	-0.4462	
163	Skewness	0.2173	
164	Range	797	
165	Minimum	501	
166	Maximum	1298	
167	Sum	130209	
168	Count	150	
169			

Fig. 7.57

Notice from Fig. 7.57 that the ***Descriptive Statistics*** option under ***Data Analysis ToolPak*** provides the following information:

- ✓ Arithmetic mean in Cell B156, which is the same as shown in Fig. 6.4 of Lab Session 6.
- ✓ Standard error in Cell B157, which you have studied in Unit 1 of MST-004 (Statistical Inference).
- ✓ Median and mode in Cells B158 and B159, respectively.
- ✓ Standard deviation and variance in Cells B160 and B161, respectively. Note that the ***Data Analysis ToolPak*** determines the sample standard deviation and variance, which has (n – 1) in the denominator instead of n, as explained in Unit 1 of MST-004.
- ✓ Kutosis and skewness in Cells B162 and B163, respectively. These values differ from the values shown in Figs. 7.45 and 7.46. This is because Excel as well as major statistical software packages include an adjustment for the sample size and provide the adjusted coefficient of skewness and kurtosis. Since it is beyond the scope of this course, we shall not discuss this aspect.
- ✓ The range, minimum and maximum values of the data, which are given in Cells B164, B165 and B166, respectively.
- ✓ The sum, i.e., total of the given values and count, i.e., the number of observations in the given data in Cells B167 and B168, respectively.

You can now try the following exercises for practice.

Activity 6

Work out the following exercises with the help of MS Excel 2007:

A1) Examples 1 to 10 given in Unit 2 of MST-002.

A2) Exercises E1 to E8 given in Unit 2 of MST-002.

A3) Example 1 given in Unit 3 of MST-002.

A4) Exercises E1 to E3 given in Unit 3 of MST-002.

A5) Exercise E4 given in Unit 4 of MST-002.

Match the results with the manual computation of the data carried out in Units 2, 3 and 4 of MST-002.

Continuous Assessment 7

1. Consider the discrete frequency distribution constructed for the data of Table 5 of Lab Session 2 and compute

   i) the range, quartile deviation and its coefficients,

   ii) the mean deviation, variance, standard deviation and coefficient of variation, and

   iii) moments, skewness and kurtosis.

2. Consider the raw data as well as continuous frequency distribution of Table 6 of Lab Session 2 and compute

   i) the range, quartile deviation and its coefficients,

   ii) the mean deviation, variance, standard deviation and coefficient of variation, and

   iii) moments, skewness and kurtosis.

Home Work: Do It Yourself

1) Follow the steps explained in Secs. 7.3 to 7.8 to compute the measures of dispersion, moments, skewness and kurtosis. Take the screenshots and keep them in your record book.

2) Develop the spreadsheets for the exercises given in "Continuous Assessment 7" as explained in this lab session. Take the screenshots of the final spreadsheets.

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