UNIT 16 GRAPHICAL PRESENTATION OF DATA-II

Structure

16.1 INTRODUCTION

In Unit 15 of this block, we have discussed some of the techniques of graphical presentation of data. In that unit, we have restricted ourselves to the graphical methods which are used for representing frequency distributions. The present unit discusses the graphical methods for time series data. A time series graph is frequently used for analysing and presenting the time series data. Range chart is a type of time series graph which is used for showing the range of variation. The Band chart is another type of time series graph which shows the total for successive time periods broken up into subtotals for each of the component parts of the total. In this unit, we shall also discuss as to how data are represented by plotting stem-and-leaf displays and box plots. Stem-and-leaf display is like histograms, but here the additional feature is that the given value of each individual is also shown in these displays. Further, in this unit we shall discuss box plots to represent the data through five-number summary.

Objectives

After studying this unit, you should be able to:

- describe a time series graph;
- describe the method of drawing a time series graph;
- draw a range chart and band chart;
- describe the method of drawing a stem-and-leaf display;
- describe the box plot and the different parts of the box plot; and
- draw the box plots.

16.2 TIME SERIES GRAPHS

A time series graph is drawn to analyse the time series data. It brings out the pattern of fluctuations in time series data and facilitates in obtaining meaningful results about its future behaviour. To draw a time series graph, time is measured along the horizontal axis and the observed data along the vertical axis. After then points are plotted against the magnitudes corresponding to each successive time period and finally these points are joined by the line segments. The resulting zigzag curve is called the time series graph. For examples, a time series graph for the production (in tons) of a commodity during the period 2001-2009 given in following table is shown in Fig. 16.1

Years	2001	2002	2003	2004	2005	2006	2007	2008	2009
Production	30	25	35	40	25	30	40	45	20

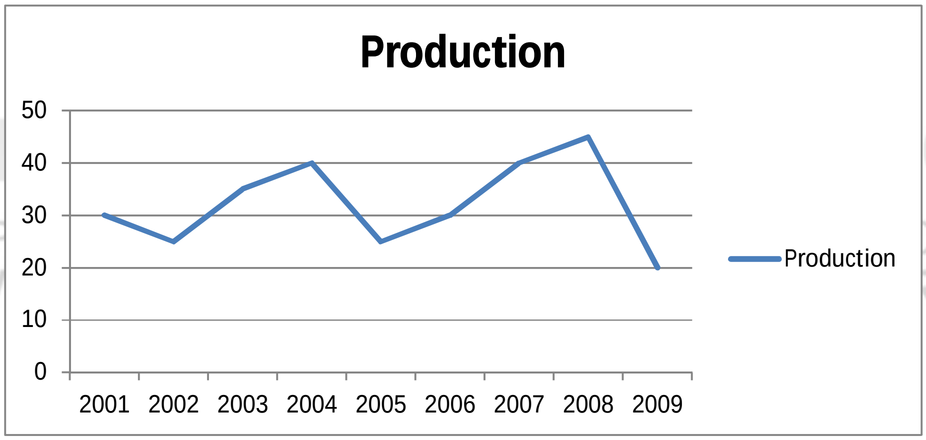


Fig. 16.1 Production of a Commodity from 2001 to 2009

In rest of this section we shall discuss on method of drawing a time series graph in subsection 16.2.1 and types of time series graph in subsection 16.2.2.

16.2.1 Method of Drawing a Time Series Graph

A time series is formed by the observations of a variable under study at different phases of time. The time series graph is extremely helpful in analysing the fluctuations in the values of the variable under study at different phases of time. We generally take time variable along x-axis and the values or magnitude of the observations of the variable under study along y-axis. After plotting the values of the variable against the corresponding values of the time variable as points, we join such points by line segments. The graph so drawn is known as a **time series graph** or **line graph**. Such type of graph is simplest to understand, easiest to draw and most widely used in practice. With the help of these graphs several variables can be shown on the same graph and a comparison can also be made. Following are some points which should be followed while constructing a time series graph:

> **Kink:** Refer Note 3 on page number 52 of Unit 15 of this course.

a) The scale of the y-axis should begin from zero even if the lowest y-value associated with any x-period or value is far from zero. If necessary concept of kink can be used.

b) If the unit of measurement is same, we can represent two or more variables on the same graph.

c) Not more than 5 or 6 number of variables is shown on the same graph, otherwise the chart/graph becomes quite confusing.

d) When two or more variables are to be shown on the same graph, it is advised to use different designs of lines to distinguish between the variables.

16.2.2 Types of Time Series Graph

There are two types of time series graphs:

(i) Range Chart (ii) Band Chart

Let us discuss these one by one:

(i) Range Chart

A range chart is a very useful method of showing the series of range of variation or fluctuation between the maximum and minimum values of a variable at the same point of time. For example, if we are interested in showing the minimum and maximum prices of a commodity for different periods of time or the minimum and maximum marks obtained by the students in different years, etc. the range chart would be the appropriate option.

For drawing a range chart, we take time variable along x-axis and the value of other variable on the y-axis. Then we draw two line graphs together by plotting the maximum and minimum observations in the given data. One curve representing the highest values at different point of time of the variable and the other one representing the lowest values at the same different point of time. The gap between both the curves represents the range of variation. For highlighting the difference between the lowest and highest values, the use of some colour or shade should be made. Let us take an example of drawing a range chart.

Example 1: Represent the following data by range chart.

Days	Max. Temp. $(\text{in } ^\circ\text{C})$	Min. Temp. $(\text{in } ^\circ\text{C})$
Monday	38	12
Tuesday	41	16
Wednesday	35	14
Thursday	42	15
Friday	44	18
Saturday	45	20
Sunday	46	21

Solution: Since there are two variables with same scales of measurement, both the variables are shown on the same graph as in Fig. 16.2.

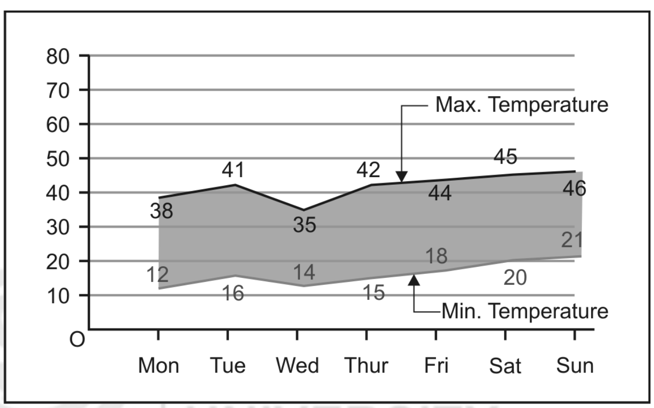


Fig. 16.2 Range Chart for the Maximum and Minimum Temperature in a Week

(i) Band Graph

Another type of a time series graph which shows the total for successive time periods broken up into sub divisions for each of the components of the total is known as band graph. In other words, the band graph represents the range of the components of total and shows as to how and what to be distributed. The different components of the total are plotted as line graph one over the other in band graph and the gaps between the successive lines are represented and differentiated with filling up them by different shades, colours, etc., so that the graph looked like the series of bands. The band graph is especially useful in representing the components which divide such as total costs, total sales, total production, etc. into their various respective components. For example, total production may be divided into its components like nature of commodity, machines, plants, etc. Band graph is also used where the data are put in percentage form. The whole graph depicts 100% and the bands as the percentage of the various components of the total. Let us take an example of drawing a band graph.

Example 2: Present the data on the amount of production (in million tones) of various plants from 2000-01 to 2007-08 given in the following table:

Year	Plant-1	Plant-2	Plant -3	Plant -4
2000-01	42.6	38.3	26.5	31.7
2001-02	48.7	36.4	28.6	34.7
2002-03	47.2	32.4	25.3	30.8
2003-04	44.8	38.8	30.1	29.6
2004-05	46.7	37.2	27.4	32.6
2005-06	45.2	34.9	25.2	35.4
2006-07	49.1	37.8	29.1	38.2
2007-08	48.2	37.8	28.4	33.5

Solution: The above data can be most suitably presented through a band graph. We proceed for constructing such a graph as follows:

We take the time on the x-axis and the other variable on the y-axis. Then we plot the various points for different years for Plant-1 and join them by straight line segments. This is represented by curve A (see Fig. 16.3). Now we add the values of production of Plant-2 for various years to the values of production of Plant-1 and plot the values and finally join them by straight line segments. This is represented by curve B (see Fig. 16.3). The difference between the curves B and A, gives the production of Plant-2. Now we add the values of production of Plant-1 and Plant-2 to that of Plant-3 and plot the various points. This is represented by curve C (see Fig. 16.3). The difference between curve C and curve B represents production of plant -3. After that we add the values of production of Plant-1, Plant-2 and Plant-3 to the values of production of Plant -4 and draw a curve. This is represented by curve D (see Fig. 16.3). The difference between curve C and curve D represents the production in Plant-4. Using these steps required band graph is shown in Fig. 16.3.

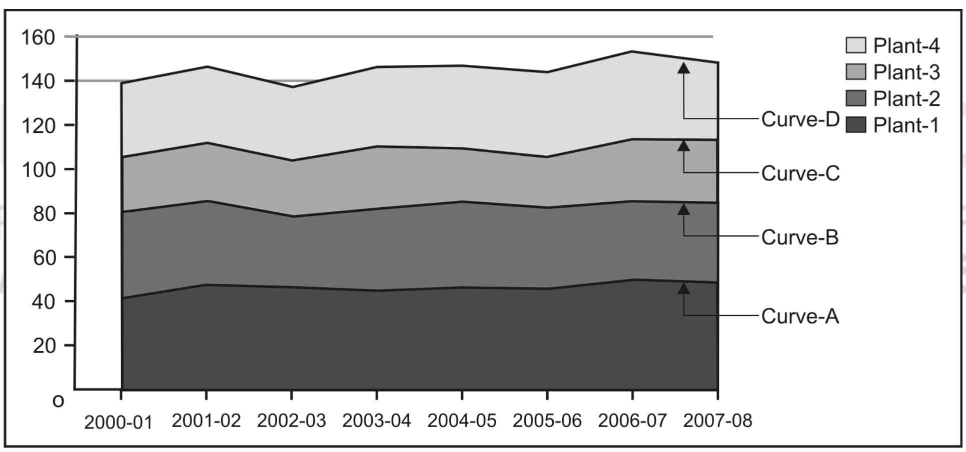


Fig. 16.3: Band Graph for the Production in four Plants

Now, you can try the following exercises.

E1) Draw a range chart for the following data:

Class:	1	2	3	4	5	6	7	8	9	10
Max: Marks:	58	65	74	61	87	65	78	92	67	84
Min. Marks:	15	21	25	32	26	16	19	22	24	17

E2) Draw a band graph for the following data of quarterly results for profit (in lakhs of rupees):

Quarters:	Plant-I	Plant-II	Plant-III
Quarter-I	34	43	46
Quarter-II	41	47	41
Quarter-III	38	39	44
Quarter-IV	51	57	53

16.3 STEM-AND-LEAF DISPLAYS

In previous units of this block as well as in previous Sec. of this unit, we have seen that data can be represented in a variety of ways including graphs, charts and tables. In this Sec. we discuss another type of graph named stem-and-leaf display. A stem-and-leaf display is very similar to a histogram but shows more information. The stem-and-leaf display summarises the shape of a set of data and provides the details regarding individual values. A stem-and-leaf display quickly summarises data while maintaining the individual data points. Now a day's use of stem-and-leaf displays is increasing, so let us formally define it in the next paragraph with some examples.

It has a vertical line of numbers obtained after removing the last digits (i.e. unit digits) from the given numbers called starting parts and for each starting part there is a horizontal line of numbers, i.e. the digits at the unit places of the given numbers called leaves. And each complete horizontal line including starting part and leaves is known as stem. The data displayed like this is nothing but known as stem-and-leaf display. The distance between the lowest values that are recorded in two consecutive stems is known as **stem width** or **category interval**, which plays very important role in stem-and-leaf displays.

Stem-and-leaf displays are used in many situations like series of scores of sports teams, series of temperature or rainfall over a period of time, series of classroom test scores, etc.

Following example will illustrate the above discussion more clearly.

Example 3: We have a set of values of the test scores of 22 students in a class as 11, 2, 28, 33, 48, 0, 42, 17, 24, 14, 0, 18, 26, 29, 35, 42, 22, 8, 28, 8, 46, 14. Draw a simple stem-and-leaf display by taking stem width 10.

Solution: Simple stem-and-leaf display for the given data can be shown as follows:

0	2 0 0 8 8	(5)
1	1 7 4 8 4	(5)
2	8 4 6 9 2 8	(6)
3	3 5	(2)
4	8 2 2 6	(4)

Total observations = 22

After arranging the leaves in ascending order of magnitude, we have

0	0 0 2 8 8
1	1 4 4 7 8
2	2 4 6 8 8 9
3	3 5
4	2 2 6 8

Here starting parts show the 'tens digits' and the leaves show the ‘ones digits’ in the above stem-and-leaf display. At a glance, one can see that 4 students got marks in the 40's in their test out of 50. Out of these four students two got 42 marks each, whereas the other two got 46 and 48 marks in the test. Fourth row (i.e. fourth stem) indicates that two students got marks in 30’s in their test out of 50. And actual marks of these two students are 33 and 35. Similarly, we can get the information about the marks of the other students from successive rows (stems). When you count the total numbers of leaves, you may know how many students appeared in the test. The information is nicely organised when a stem-and-leaf display is used. Stem-and-leaf display provides a tool for specific information in large sets of data, otherwise one would have a long list of marks to arrange and analyse.

16.3.1 Stem-and-Leaf Display for more than one Set of Data

Stem-and-leaf display is also used to compare two sets of data. That is known as 'back to back' stem-and-leaf display. For example, if you want to compare the batting scores of two cricket players, then stem-and-leaf display is right way to represent the data.

Example 4: Draw a stem-and-leaf display for batting scores of two players given below.

Player A	102, 61, 82, 88, 90, 63, 69, 85, 105, 93, 65, 94, 107, 97, 67
Player B	104, 62, 83, 95, 106, 95, 108, 63, 108, 82, 93, 109

Solution: The scores of two players can be compared with the help of back to back stem-and-leaf display as follows:

Leaf (Player A)	Starting part	Leaf (Player B)
1 3 5 7 9	6	2 3
2 5 8	8	2 3
0 3 4 7	9	3 5 5
2 5 7	10	4 6 8 8 9

Here column of starting parts is now in the middle and the leaves columns are to the right (player B) and left (player A) of the column of starting parts. You can see that the player B has more innings with a highest score than the player A. The player B has only 2 innings with scores of 62 and 63, while the player A has 5 innings with the scores of 61, 63, 65, 67 and 69. You can also see that player B has the highest score of 109, compared to player A with highest score of 107. Thus we see that presentation of the data by stem-and-leaf display provides us lot of information in very quick time.

In above two examples stem width or category interval was 10. Now we take an example in which stem width is 5 instead of 10.

Example 5: Arrange the numbers 47, 35, 37, 20, 43,15, 15, 26,46, 25, 29, 12, 39, 44, 21, 24, 16, 40, 19, 46, 30, 34, 17, 39, 16, 40, 31, 21, 14, 42,16, 43, 22, 11, 24, 25, 31, 27, 40, 33 in a stretched stem-and-leaf display that has single-digit starting parts and leaves, but has stem width of 5.

Solution: A simple stem-and-leaf display has a unique starting part for each stem with stem width 10, while the stretched stem-and-leaf display shown below has a stem width 5 which means we have stretched the stem (of stem width 10) into two stems each of width 5, and the same starting part is used for both stems (i.e. for stem 1 we used 1a and 1b, for stem 2 we used 2a and 2b, etc., it is also explained below the display). The required stretched stem-and-leaf display is given as follows:

1a	2 4 1
1b	5 5 6 9 7 6 6
2a	0 1 4 12 4
2b	6 5 9 5 7
3a	0 4 1 1 3
3b	5 7 9 9
4a	3 4 0 0 2 3 0
4b	6 5 5

In this stem-and-leaf display 'a' stands for the interval 0-4 and 'b' stands for 5-9. The values between10-19 of stem 1 are now represented into two stems 1a and 1b which include values between 10-14 and 15-19 respectively. Similarly, values between 20-29 of stem 2 are now represented by two stems 2a and 2b, which include values between 20-24 and 25-29 respectively, and so on.

16.3.2 Merits of Stem-and-Leaf Display

Following are some merits of stem-and-leaf display:

(i) Stem-and-leaf display arranges the data in place values.

(ii) Total number of observations and mode can easily be obtained from stem-and-leaf display (see Example 3).

(iii) Summarises the shape of a set of data (the distribution) and provides the detail regarding individual values.

(iv) Stem-and-leaf display also enables you to find quantiles such as median, quartiles (i.e. Q_1, Q_2, Q_3), deciles (i.e. D_1, D_2, D_3, ... , D_9), percentiles (i.e. P_1, P_2, P_3, ..., P_{99}), etc. As discussed below.

Formula for Calculating Quantiles: First of all given observations are arranged in ascending order of magnitude. Then j m^{ths} quantiles denoted by $Q_{j/m}$ (e.g. 7/10 of the data are below $Q_{7/10}$) is given by

$Q_{j/m} = x_i$, where x_i is that value of the variable below which j m^{ths} observations lie and

$$i = \left(\frac{j \times n}{m}\right) + \frac{1}{2} \qquad \text{... (16.1)} \quad \text{,where n = total number of observations}$$

For example, let us apply this formula for finding median for the data of Example 3:

Median = Second quartile = $Q_{2/4}$:

$$i = \left(\frac{j \times n}{m}\right) + \frac{1}{2} = \left(\frac{2 \times 22}{4}\right) + \frac{1}{2} = 11.5$$

$\therefore$ median = $x_{11.5}$ = 11th observation in the array

+ 0.5(12th observation – 11th observation)

= 22 + 0.5(24 – 22) = 23

Now, you can try the following exercises.

E3) Draw a stem-and-leaf display with the following marks obtained by 30 students.
77, 80, 82, 68, 65, 59, 61, 57, 50, 62, 61, 70, 69, 64, 67,
70, 62, 65, 65, 73, 76, 87, 80, 82, 83, 79, 79, 71, 80, 77
Also determine the median for the marks.

E4) Draw a stem-and-leaf display for the following data.
31, 42, 22, 27, 33, 57, 67, 58, 64, 44, 65, 59, 46, 61, 35, 26, 63
Also find seventh decile.

E5) Draw a stem-and-leaf display for the given data:
141, 137, 105, 139, 107, 144, 110, 135, 117, 125, 147,113, 109, 120, 132, 110, 130, 112
Also find sixty seventh percentile.

16.4 BOX PLOTS

In previous section we have discussed the stem-and-leaf displays. Now let us discuss another type of plot which is known as Box plot in this section. In descriptive statistics, a box plot also known as a box-and-whisker plot is a convenient way of graphically representing numerical data. It represents the data through five-number summary, i.e. the smallest observation (sample minimum), lower quartile (Q_1), median (Q_2), upper quartile (Q_3), and the largest observation (sample maximum). Box plots are used to describe a distribution generally when it is extremely skewed or multimodal. A box plot also indicates which observation(s), if any, might be considered as outliers. A box plot is a quick graphic approach for examining one or more sets of data.

Box plots display differences between populations without making any assumptions of the distributions. The spacing between the different parts of the box helps in indicating the degree of dispersion (spread) and skewness in the data, and identifies outliers. Box plots can be drawn either horizontally or vertically. Here we will draw box plots vertically.

16.4.1 Method of Construction of the Box Plots

Box plots are useful for identifying key values while comparing two or more distributions. To understand more clearly the method of constructing the box plots, let us consider the following data of 37 students in a class who were examined by a game with a box containing some balls. Their task was to select a ball from one box placed at one corner and put it in another blank box placed at another corner of the class as quickly as possible, and their times (in seconds) were recorded. The scores were compared for 16 boys and 21 girls who participated in game. Observed data are given in the following table:

Time (in seconds) for completing the given task

Boys	18, 19, 20, 22, 24, 25, 26, 16, 17, 19, 25, 27, 28, 23, 23, 31
Girls	15, 17, 18, 19, 20, 21, 23, 14, 17, 18, 19, 20, 21, 24, 19, 16, 17, 18, 20, 22, 28

The method of construction of separate box plots for the data of boys and girls is discussed below:

There are several ways of constructing a box plot. The first relies on the quartiles, lowest and greatest values in the distribution of scores. Fig. 16.4 shows how these three statistics are used for the above example. We draw a box plot for each gender extending from the 1st quartile to the 3rd quartile. The 2nd quartile is drawn inside the box. Therefore,

(i) The bottom of each box is the 1st quartile,

(ii) The top is the 3rd quartile,

(iii) The line in the middle is the 2nd quartile.

(iv) A line extending from the point corresponding to the smallest observation to 1st quartile is drawn and known as **lower whisker**.

(v) A line extending from 3rd quartile to the point corresponding largest observation is drawn and is known as **upper whisker**.

Let us arrange the given data for each of the gender in ascending order as shown in following table to find out the above components, i.e. Q_1, Q_2, Q_3, smallest observation and largest observation.

Gender	Times (in Seconds)
Boys	16, 17, 18, 19, 19, 20, 22, 22, 23, 23, 24, 25, 25, 27, 28, 31
Girls	14, 15, 16, 17, 17, 17, 18, 18, 18, 19, 19, 19, 20, 20, 20, 21, 21, 22, 23, 24, 28

For Boys

The lowest or smallest observation = x_s = 16.

First quartile $(Q_1) = \left(\frac{16+1}{4}\right)^{th}$ item

= 4.25th item = 4th item + 0.25 (5th item – 4th item)

= 19 + 0.25 (19 – 19) = 19

Second quartile = $(Q_2) = 2\left(\frac{16+1}{4}\right)^{th}$ item

= 8.5th item = Mean of 8th and 9th items

= $\frac{22+23}{2}$ = 22.5

Third quartile $(Q_3) = 3\left(\frac{16+1}{4}\right)^{th}$ item

= 12.75th item = 12th item + 0.75 (13th item – 12th item)

= 25 + 0.75 (25 – 25) = 25

The largest observation = x_1 = 31

For Girls

The lowest or smallest observation = x_s = 14

First quartile $(Q_1) = \left(\frac{21+1}{4}\right)^{th}$ item

= 5.5th item = mean of 5th and 6th items = $\frac{17+17}{2}$ = 17

Second quartile $(Q_2) = 2\left(\frac{21+1}{4}\right)^{th}$ item = 11th item = 19

Third quartile $(Q_3) = 3\left(\frac{21+1}{4}\right)^{th}$ item

= 16.5th item = mean of 16th and 17th item = $\frac{21+21}{2}$ = 21

The largest observation = x_l = 28

Box plots for boys and girls on the basis of the above findings are shown below in Fig. 16.4.

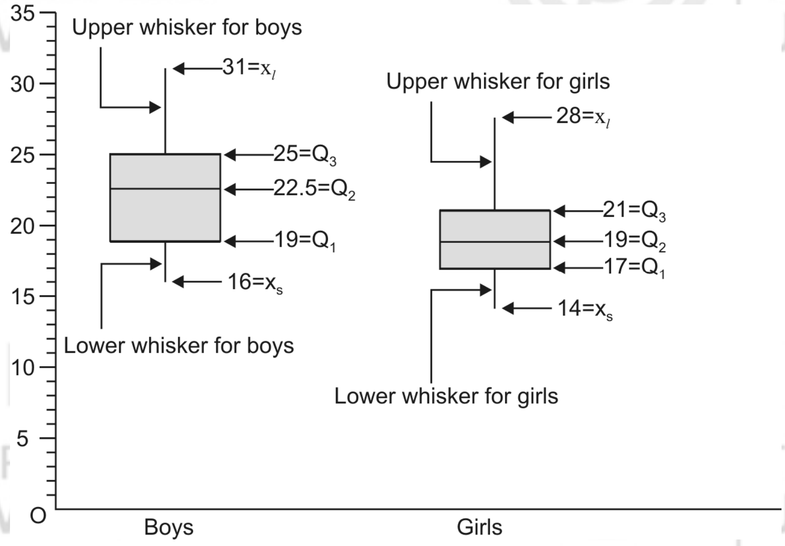


Fig. 16.4: The Box Plots with the Whiskers for Boys and Girls

16.4.2 Components of a Box Plot

Let us now discuss the various components and terminologies which are used in drawing the various types of box plots.

1. **Upper and Lower Hinges**

   The upper and lower hinges are constructed to represent the third quartile and first quartile respectively. For the example discussed above, the values of upper and lower hinges are 21 and 17 for the girls whereas for the boys they are 25 and 19 respectively.

2. **H-Spread**

   This is calculated by taking the difference between the upper and lower hinges. The H-Spread shows the spreadness of the elements of the data between the first quartile and third quartile. For the data related to boys in the example discussed above, the H-spread is 25–19 = 6 whereas for girls it is 21–17 = 4.

3. **Whiskers**

   The lines extending above and below the box are called whiskers. Lower whisker extends from the point corresponding to the smallest observation to Q_1 and the upper whisker extends from Q_3 to the largest observation.

4. **Step**

   The step is calculated by multiplying the H-spread by 1.5. For the data of boys in the above example, the value of 1 step is $1.5\times 6 = 9$, whereas for girls it is $1.5 \times 4 = 6$.

5. **Upper and Lower Inner Fences**

   The upper and lower inner fences are calculated by adding one step to the upper hinge and subtracting 1 step from the lower hinge respectively. In other words, upper inner fence is equal to the 1 step ahead from upper hinge whereas the lower inner fence is one step before the lower hinge. For the data of boys in the example discussed above, the upper and lower inner fences are 25 + 9 = 34 and 19 – 9 = 10 whereas for the data of girls the upper and lower inner fences are 21 + 6 = 27 and 17 – 6 = 11 respectively.

6. Upper and Lower Outer Fences

The upper and lower outer fences are calculated by adding two steps to the upper hinge and subtracting 2 steps from the lower hinge respectively. For the data of boys in the example discussed above the upper and lower outer fence are $25 + 2\times 9 = 43$ and $19 - 2\times 9 = 1$, whereas for the data of girls they are, $21 + 2\times 6 = 33$ and $17 - 2\times 6 = 5$, respectively.

7. Upper and Lower Adjacent

The upper and lower adjacent are used to represent the largest and the smallest observations of the data. For the data of boys in the example discussed above upper and lower adjacent are 31 and 16 whereas for the girls data they are 28 and 14 respectively.

8. Outside Value

The outside value is a value which is beyond an inner fence but not beyond an outer fence. This is used to represent the scattered values of the data. To represent these values circles are used.

9. Far Out Value

A value which is beyond the upper outer fence or lower outer fence is known as far out value. To represent these values, the asterisks are used.

Box plot with the components discussed above for data of boys is shown in Fig 16.5.

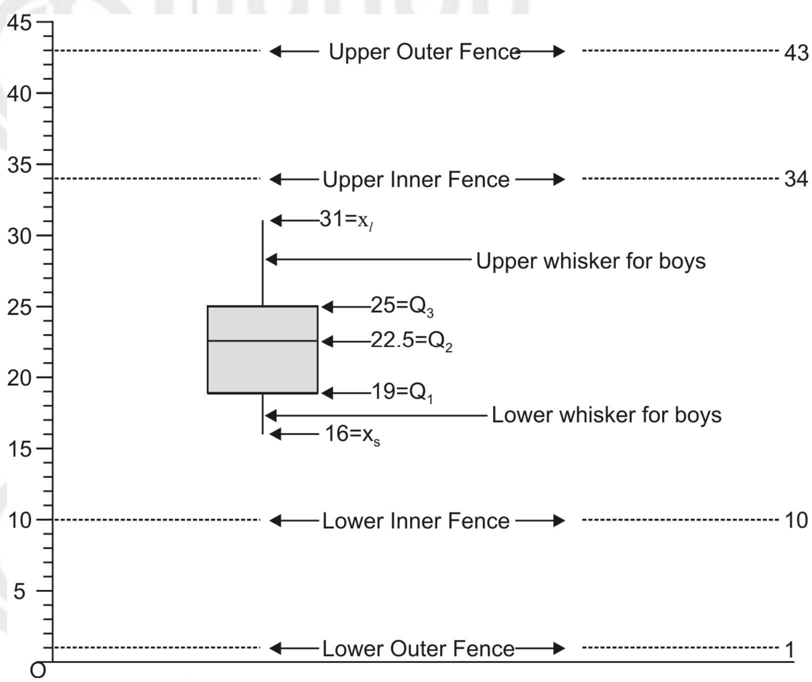


Fig. 16. 5 Box Plot for Boys with Inner and Outer Fences

Box plot with the components discussed above for data of girls is shown on next page in Fig 16.6.

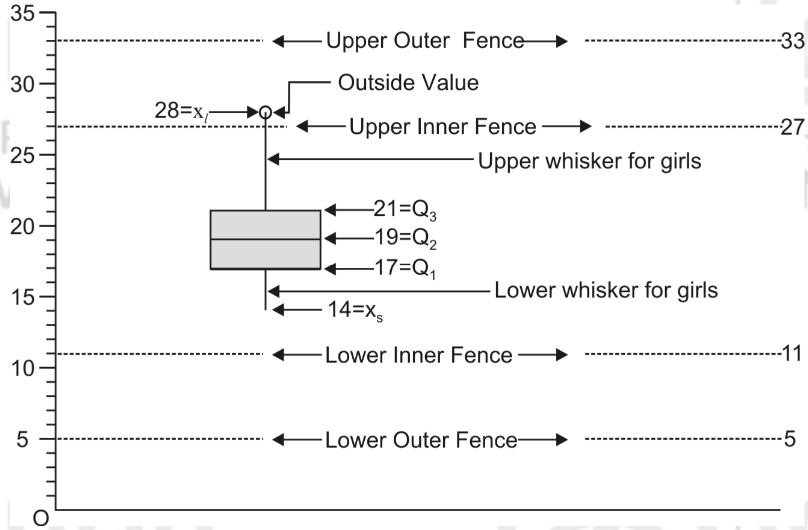


Fig. 16.6 Box Plot for Girls with Inner and Outer Fences

16.4.3 Box Plots with Outliers

Outside value(s) and far out value(s) are known as extreme observations. If the extreme observations are present in the data then those observations may be represented in box plot by an individual mark. The extreme observations are described as outliers when they are represented in box plots. The outside value is a value which is beyond an inner fence but not beyond an outer fence whereas the far out value is a value which is beyond the lower and upper outer fences. The individual marks for extreme values can be plotted above and below to the whiskers in box plot. Specially, outside values are indicated by small circles. In the data of girls in the above given example 28 is only far out value, whereas in the data of boys no value is beyond the lower or upper inner fence. The box plot for girls data indicating outside value by a circle is shown in Fig. 16.6.

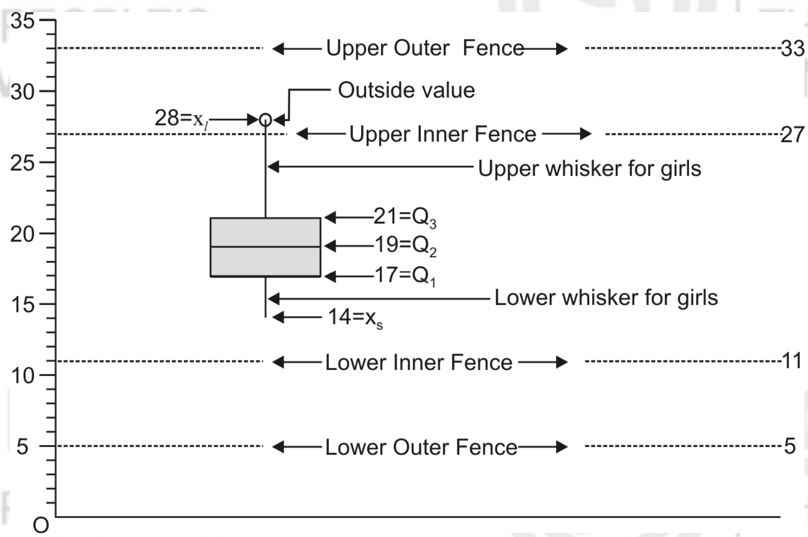


Fig. 16.6: The Box Blot for Girls with the Outlier.

16.4.4 Box Plots with + Signs

One more component which is to be included in box plots are the value of some important parameters like, mean, mode, etc. We indicate the mean score for a group of values by inserting a plus sign in box plots. For the example discussed in sub-section 16.4.1 mean in case of boys data is 20.875 and mean in case of girls data is 19.33, which are shown by a plus sign in Fig. 16.7 as given below.

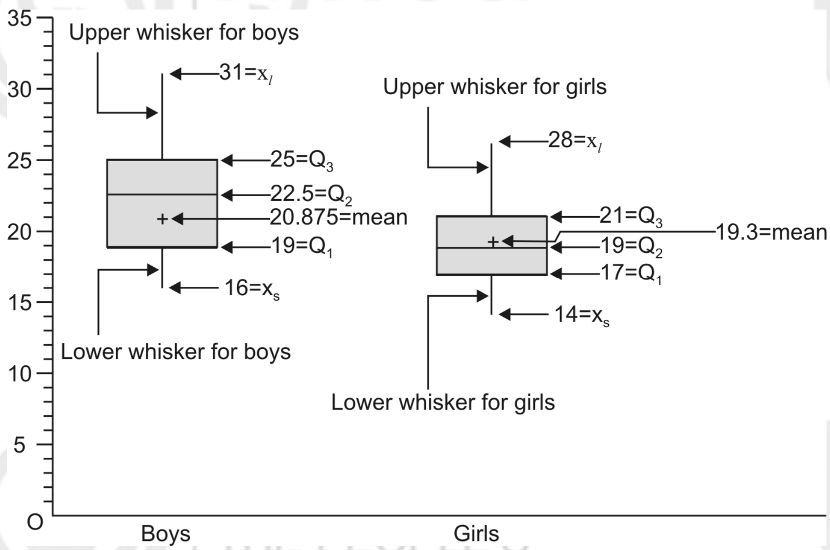


Fig. 16.7: The Box Plot with Whisker and + Signs for Boys and Girls Data.

Fig. 16.7 provides a revealing summary of the data. Since half of the scores are between the hinges (recall that the hinges are the first and third quartiles), we see that half of the girl's times are between 17 and 21, whereas half of the boy's times are between 19 and 25.

16.4.5 Box Plots with Whisker, + Sign and Outliers

On the basis of data of the example discussed in sub-section 16.4.1 we see that girls generally dropped the balls from one box to another faster than boys. We also see that one boy was slower than almost all of the women (except 3). Fig. 16.8 shows the box plot for the girl's data with whisker, + sign and outliers.

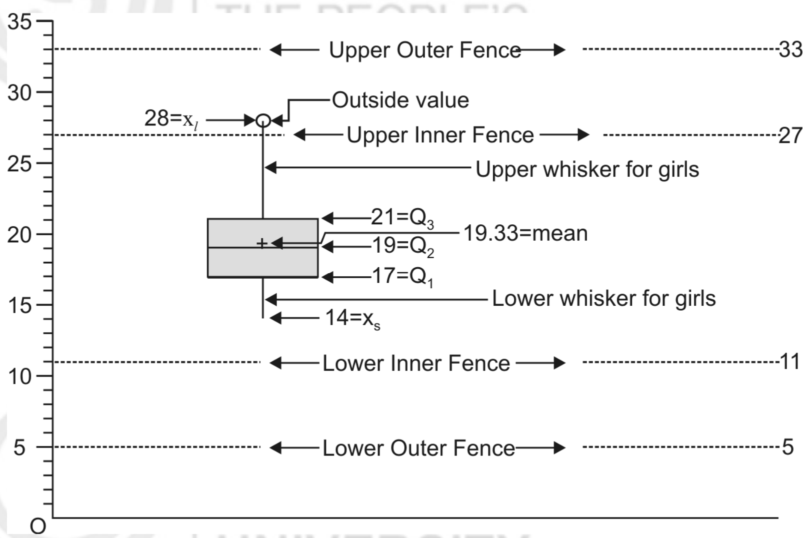


Fig. 16.8: A Box-Plot for the Girl's Data with Whisker, + Sign and Outliers

Note 2: If some learner is interested to know more about the topics discussed in Secs. 16.4 and 16.5 he/she may refer chapters 6 and 7 of the book written at serial number 5 in the reference books listed below the introduction of MST-001 on page number 4 of block 1.

Now, you can try the following exercises.

E6) Draw a box plot for the given data:

17, 15, 17, 20, 13, 15, 15, 16, 16, 15, 19, 12, 19, 14, 11, 14, 16, 10, 19, 18, 20, 14, 17, 19, 16, 22, 21, 23, 14, 12, 18, 13, 12, 25, 14, 15, 31, 17, 10, 21

E7) Draw a box plot for the given data:

31, 42, 22, 27, 33, 27, 37, 28, 34, 44, 25, 39, 26, 31, 26, 33, 46, 48, 50

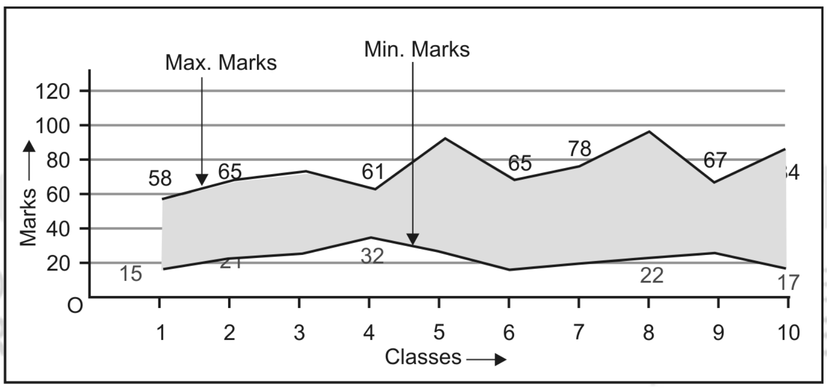
16.7 SUMMARY

In this unit, we have discussed:

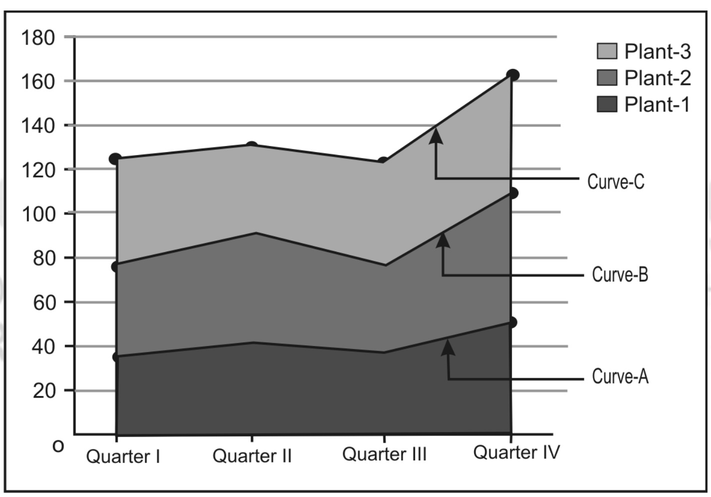
1) Time series data and methods of drawing a time series graph.
2) How to draw a range chart and band chart.
3) Methods of drawing a stem-and-leaf displays.
4) The box plots and the different components of the box plot.

16.8 SOLUTIONS / ANSWERS

E1) Range chart of the given data is given below.



E2) Required band graph of the given data is given below.



E3) Stem-and-leaf display of the given data of marks obtained by 30 students is given below.

5	9 7 0
6	8 5 1 2 1 9 4 7 2 5 5
7	7 0 0 3 6 9 9 1 7
8	0 2 7 0 2 3 0

After arranging the leaves in ascending order of magnitude, we have

5	0 7 9
6	1 1 2 2 4 5 5 5 7 8 9
7	0 0 1 3 6 7 7 9 9
8	0 0 0 2 2 3 7

Median = second quartile = $Q_{2/4}$:

$$i = \left(\frac{j \times n}{m}\right) + \frac{1}{2} = \left(\frac{2 \times 30}{4}\right) + \frac{1}{2} = 15.5$$

$\therefore$ median = $x_{15.5}$ = 15^{th} value in the array + 0.5(16^{th} value – 15^{th} value)

= 70 + 0.5(70 – 70) = 70

E4) Stem-and-leaf display of the given data is given below.

2	2 7 6
3	1 3 5
4	2 4 6
5	7 8 9
6	7 4 5 1 3

After arranging the leaves in ascending order of magnitude, we have

2	2 6 7
3	1 3 5
4	2 4 6
5	7 8 9
6	1 3 4 5 7

D_7 = seventh decile = $Q_{7/10}$:

$$i = \left(\frac{j \times n}{m}\right) + \frac{1}{2} = \left(\frac{7 \times 17}{10}\right) + \frac{1}{2} = 12.4$$

$\therefore D_7$ = $x_{12.4}$ = 12^{th} value in the array + 0.4(13^{th} value – 12^{th} value)

= 59 + 0.4(61 – 59) = 59.8

E5) Stem-and-leaf display of the given data is given below.

10	5 7 9
11	0 7 3 0 2
12	5 0
13	7 9 5 2 0
14	1 4 7

After arranging the leaves in ascending order of magnitude, we have

10	5 7 9
11	0 0 2 3 7
12	0 5
13	0 2 5 7 9
14	1 4 7

P_{67} = sixty seventh percentile = $Q_{67/100}$:

$$i = \left(\frac{j \times n}{m}\right) + \frac{1}{2} = \left(\frac{67 \times 18}{100}\right) + \frac{1}{2} = 12.56$$

$\therefore P_{67} = x_{12.56} = 12^{th}$ value in the array + 0.56(13^{th} value – 12^{th} value)

= 132 + 0.56(135 – 132) = 133.68

E6) After arranging the given data in ascending order of magnitude, we have

10, 10, 11, 11, 12, 12, 12, 13, 13, 14,14,14,14, 15, 15, 15, 15, 15, 16, 16, 16, 16, 17, 17, 17, 17, 18, 18, 19, 19, 19, 19, 20, 20, 21, 21, 22, 23, 25, 25

The lowest or smallest observation = x_s = 10

First quartile $(Q_1) = \left(\frac{40+1}{4}\right)^{th}$ item

= 10.25^{th} item = 10^{th} item + 0.25 (11^{th} item –10^{th} item)
= 14 + 0.25 (14 – 14) = 14

Second quartile $(Q_2) = 2\left(\frac{40+1}{4}\right)^{th}$ item

= 20.5^{th} item = Mean of 20^{th} and 21^{st} items

$$= \frac{16+16}{2} = 16$$

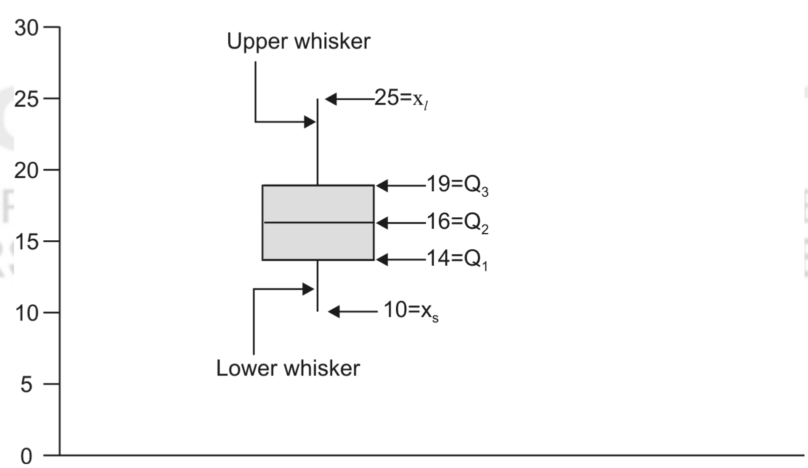
Third quartile $(Q_3) = 3\left(\frac{16+1}{4}\right)^{th}$ item

= 30.75^{th} item = 30^{th} item + 0.75 (31^{st} item – 30^{th} item)

= 19 + 0.75 (19 – 19) = 19

The largest observation = x_l = 25

Using above calculations box plot based on five-number summary (i.e. smallest observation x_s, first quartile (Q_1), second quartile (Q_2), third quartile Q_3), largest observation x_l) is given below:



E7) After arranging the given data in ascending order of magnitude, we have

22, 25, 26, 26, 27, 27, 28, 31, 31, 33, 33, 34, 37, 39, 42, 44, 46, 48, 50

The lowest or smallest observation = x_s = 22

First quartile $(Q_1) = \left(\frac{19+1}{4}\right)^{th}$ item = 5^{th} item = 27

Second quartile $(Q_2) = 2\left(\frac{19+1}{4}\right)^{th}$ item = 10^{th} item = 33

Third quartile $(Q_3) = 3\left(\frac{19+1}{4}\right)^{th}$ item = 15^{th} item = 42

The largest observation = x_1 = 50

Using above calculations box plot based on five-number summary (i.e. smallest observation x_s, first quartile (Q_1), second quartile (Q_2), third quartile Q_3), largest observation x_1) is given below: