Know the

Data interpretation

Know the DI

OBJECTIVES:

To familiarize you with the different types of data interpretation problems in CAT and other entrance exams.

To enable you to observe graphs and grasp data.

Methods of representing data

1. TABLE

The most simple and common form of representing data is in tables. Table is a two-dimensional form of representing data in rows and columns. It is a systematic method of grouping data based on various factors. The headings of the various rows and columns provide information about the data represented by various cells in a table.

An example of a table:

Table 1.0

NUMBER OF PERSONS VISITING A MUSEUM IN THOUSANDS

	2004	2005	2006	2007	2008*
Adults	8	10	10	12	15
Students	10	12	15	19	20
Pre-schoolers	2	2	3	4	6

Question:

What is the percentage increase in the number of adults from 2004 to 2008?

A. 20% B. 30% C.40% D. None of these

This table is used to represent the data on the total number of adults, students and pre-schoolers who visited a museum in the years 2004 to 2008.

By observing the row headings, we can find that the three groups of people represented in the table are adults, students and pre-schoolers.

The column headings correspond to the years 2004 to 2008.

From the above table we can find the number of persons from each group who visited the museum in a particular year.

For example, the number of students who visited the museum in 2004 is 10,000. This can be found out by drawing a line from the row representing students and another line from the column representing the year 2004. The cell where these two lines meet, gives the required information.

What should one look out for in a table?

1. TITLE:

The title is a brief description about the table. It gives us concise information about the data represented using the table.

The title of the table 1.0 is:

'NUMBER OF PERSONS VISITING A MUSEUM IN THOUSANDS'

The title clearly states that the table presents data on the number of persons (in thousands) visiting a library.

In the above title, one should take note of the fact that the data in the table is represented in thousands. Therefore, the number of students who visited the library in the year 2005 is 12,000 and **not 12.**

2. ROW HEADINGS & COLUMN HEADINGS

The row headings and the column headings present information on the data present in the corresponding row and column of the table.

In the above table, the row headings are adults, students and pre-schoolers. The column headings are the various years for which data corresponding to the groups has been represented.

2. LINE GRAPH

A line graph is a pictorial representation of data. It consists of a line plotted by joining the points indicating the variation of a quantity with respect to the variables represented by the X-axis and Y-axis respectively. The line establishes the relationship between the variables that the X-axis and Y-axis represent.

A pictorial representation such as a line graph gives a clear indication of the trends in the variation of one quantity with respect to another.

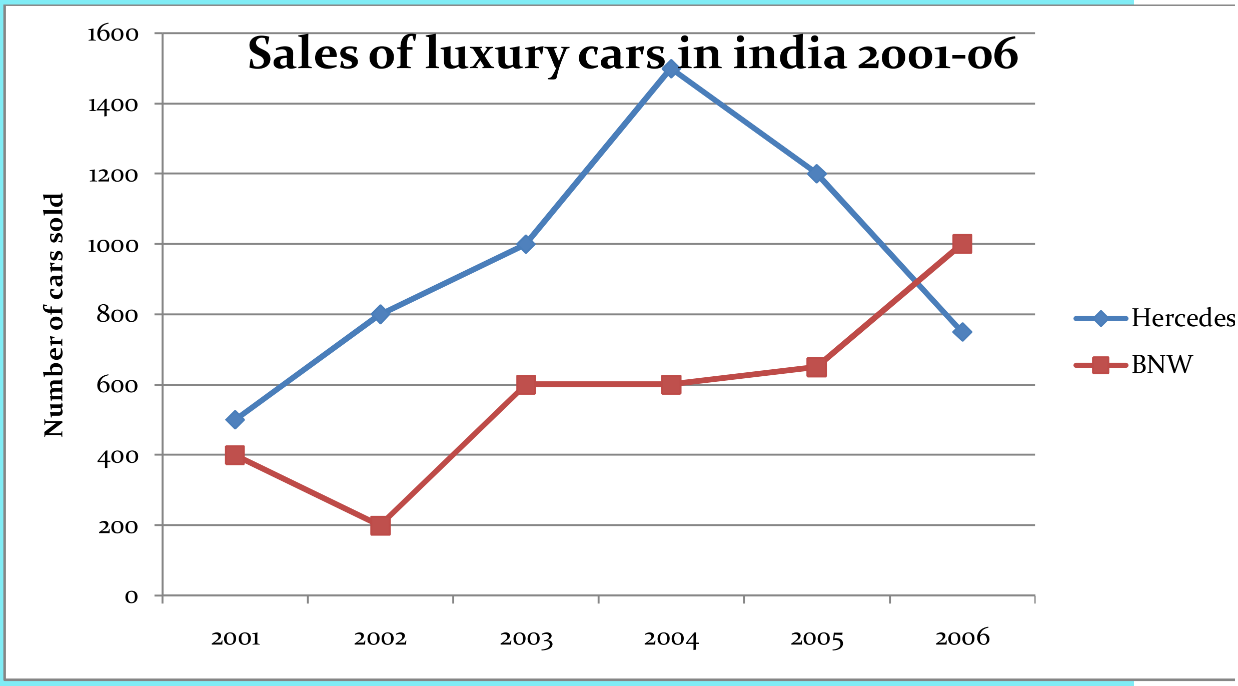


Figure 1.1

Question:

In which year is the ratio of sales of Hercedez to BNW maximum?

A. 2001 B. 2003 C.2004 D.2005

The line graph represents the annual sales of two brands of luxury cars in India in the years 2001 to 2006.

The Y-axis represents the number of cars sold.

The X–axis represents the year of sales.

The line graph is obtained by joining the points indicating the sales of the two brands of cars in each year.

From figure 1.1, we can find the number of cars of the two brands of luxury car sold in each year from 2001 to 2006.

For example, we can find that the number of BNW cars sold in 2001 is 400 and the number of Hercedes cars sold in 2002 is 800.

What to look for in a LINE GRAPH?

TITLE:

Read the title of the line graph to understand the nature of data represented by it.

In figure 1.1, the title of the line graph is '**Sales of luxury cars in India, 2001-2006**'

From the title, it is evident that the line graph is about the sales of luxury cars in India. It is also evident that the data represented corresponds to the sales from 2001 to 2006 and not for any other time.

X-axis & Y-axis label

The label of the X-axis and Y-axis gives us the information about the two parameters on the basis of which the line graph has been plotted.

In Figure 1.1, the X- axis label is '**Year of Sales**' and the Y-axis label is '**Number of cars sold'.** Always check for the units of the label while observing what the label of each axis represents.

Legend

Usually, line graphs are used to plot the variation of more than one quantity of the same type.

For example in Figure 1.1, annual sales of both Hercedes and BNW cars are represented in the same graph.

A legend is used to indicate differentiate between the two quantities being compared in the same line graph.

In figure 1.1, the line joining the diamond shaped points indicates the sales of Hercedes cars and the line joining the square shaped points indicates the sales of BNW cars.

Patterns

A line graph is generally used to analyse the trends and patterns in variation of a quantity. From one point of data representation to another, the quantity being plotted may increase, decrease or remain unchanged.

If the line joining two points in a line graph is parallel to the X-axis, there is no change in the quantity being plotted. In figure 1.1, we can observe that the line joining the number of BNW cars sold in 2003 and 2004 is parallel to X-axis. This shows that the number of BNW cars sold in 2003 and 2004 is the same.

We can also observe whether a quantity increases or decreases from a line graph.

If the line joining two points of observation in the line graph is sloping down, it represents a decrease. If the line joining the two points of observation in the line graph is sloping upwards, it represents an increase. In figure 1.1, we can observe that the line joining the number of BNW cars sold in 2001 and 2002 is sloping down; the number of cars sold has decreased from 400 to 200. The line joining the number of BNW cars sold in 2002 and 2003 is sloping upward; the sales have increased from 200 in 2002 to 600 in 2003.

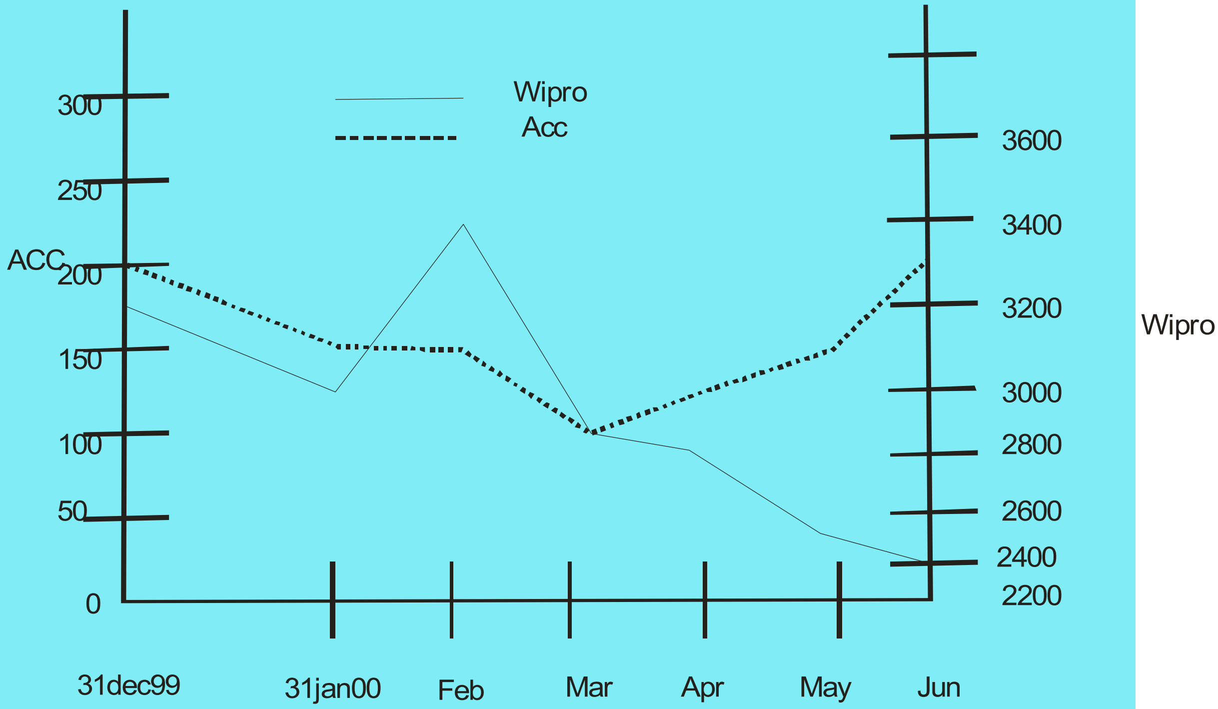
Variations in Line Graph

A line graph may have two different Y-axis with different scales.

An example is given below.

In this graph the share prices of two companies Wipro and ACC is documented on 6 month end starting Jan 2000.

Figure 1.2



Question:

In which month did Wipro grow at a faster rate than ACC? (compare the growth rate)

A. 31 Dec 99 – 31 Jan 00

B. 31 Jan 00 – Feb

C. Feb – Mar

D. Mar – Apr

Here the price is represented using a line graph with two y axis (the left one is for ACC and the right one is for Wipro). The above graph can be represented in table as

Date	ACC	Wipro
31-dec-99	200	3200
31-jan-00	150	3000

Feb	150	3400
Mar	100	2900
Apr	125	2800
May	150	2500
Jun	200	2400

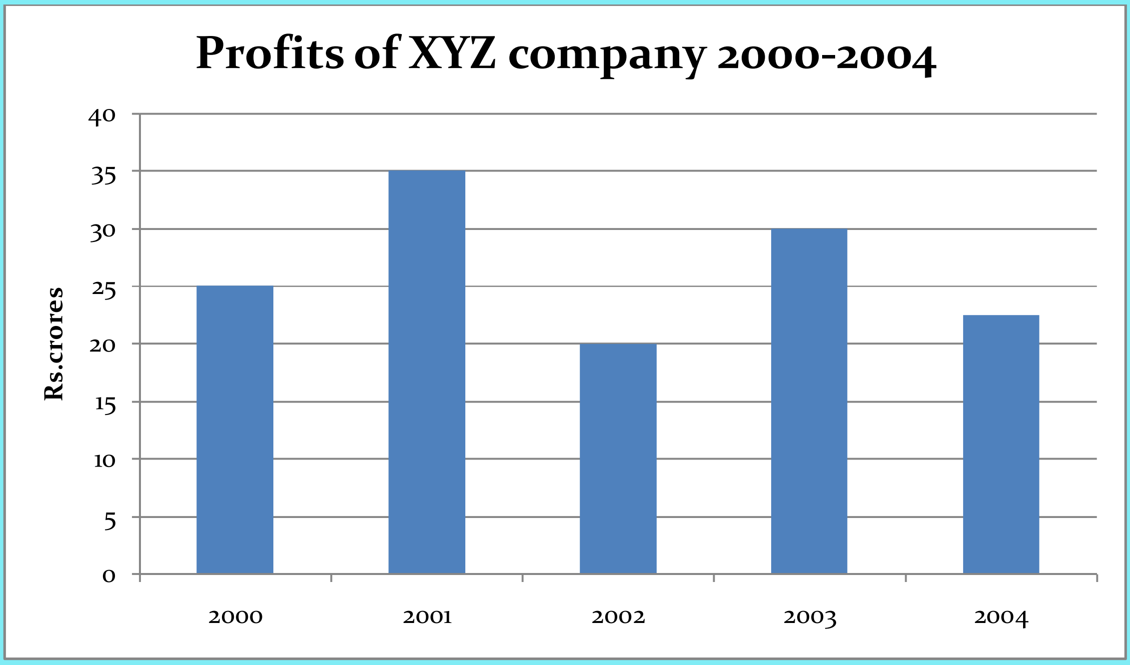
3. BAR GRAPHS

A bar graph is used to represent discrete data. In contrast, a line graph is used to represent continuous data. While the line graph is useful in observing the relative change, a bar graph is more useful in determining the absolute change in the variables being represented by the graph.

In a bar graph, the height or the length of the bar graph determines the value of the data being represented. The width of the bar graph is not of any importance.

A bar chart representation may be horizontal or vertical depending upon the orientation of the bars used to represent the data.

Figure 1.3



Question:

In which year did the profits grow the least?

A.2001

B. 2002

C. 2003

D.2004

The bar graph represents the profits of XYZ Company in the years 2000 to 2004.

The Y-axis represents the profit in crores.

The X–axis represents the years of operation.

From the Figure 1.3, we can find the profit of XYZ Company in each year from 2000 to 2004.

For example, we can find that the profit of XYZ Company in 2000 was 25 crores.

What to look for in a BAR GRAPH?

TITLE:

Read the title of the bar graph to understand the nature of data represented by it.

In figure 1.3, the title of the bar graph is '**profits of XYZ Company in the years 2000 to 2004.**'

From the title, it is evident that the bar graph is about the profits of XYZ Company. It is also evident that the data represented corresponds to the sales from 2000 to 2004 and not for any other time.

X-axis & Y-axis label

The label of the X-axis and Y-axis gives us the information about the two parameters on the basis of which the bar graph has been plotted.

In Figure 1.3, the X- axis label is '**Years of operation**' and the Y-axis label is '**Profit in Rs.crores'.** Always check for the units of the label while observing what the label of each axis represents.

VARIATIONS IN BAR GRAPH

COLUMN CHART

Like a line graph, a bar graph may also be used to represent the data about more than one varying quantity. In a column bar chart, quantities whose variation is measured against the same parameter are grouped together.

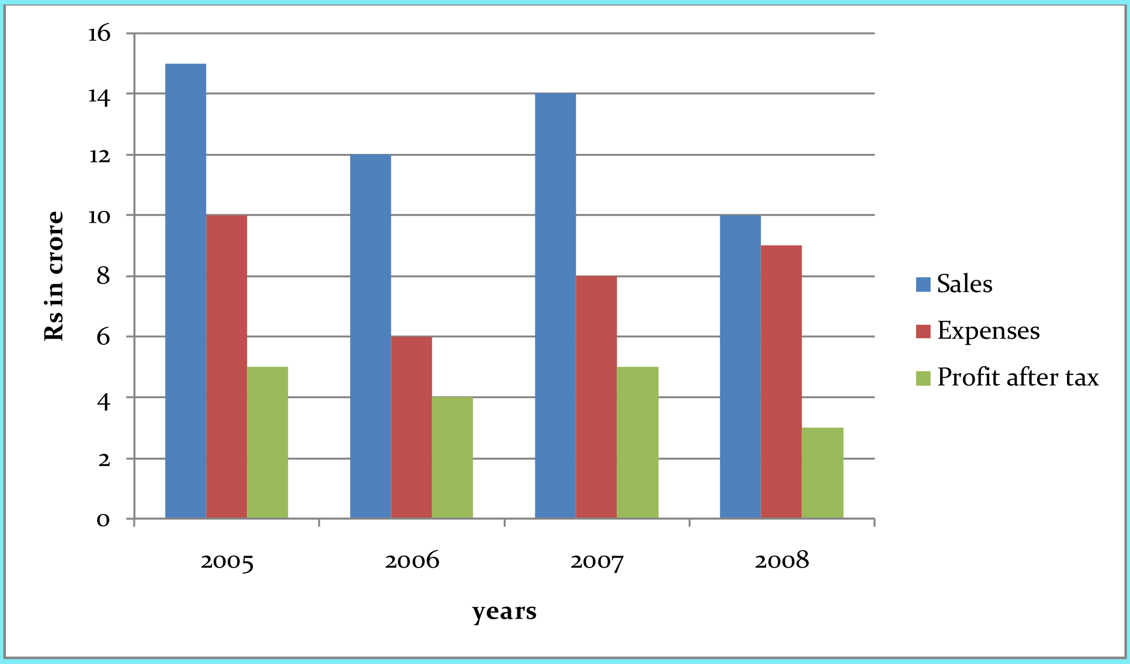


Figure 1.3

Question:

In which year is the ratio of expense to sales the least?

A.2008

B. 2005

C. 2006

D.2007

In figure1.3, the variation in three types of quantities: sales, expenses and profit after tax for the years 2005-2007 is represented in the same graph. The data representing the sales, expenses and profit after tax for each year has been clubbed together.

From the graph we can observe that in 2005, sales was 15 crores, expenses was 10 crores and profit after tax was 5 crores.

HORIZONTAL BAR CHARTS

In a column chart, the Y-axis is generally used to represent the measure of quantities and the X-axis is used to represent the categories. In a horizontal bar chart, it is done vice-versa.

The measure of the quantities is represented in the X-axis and the groupings are represented in the Y-axis.

The data in column chart shown in Figure 1.3, can be represented using a horizontal bar chart in the following manner:

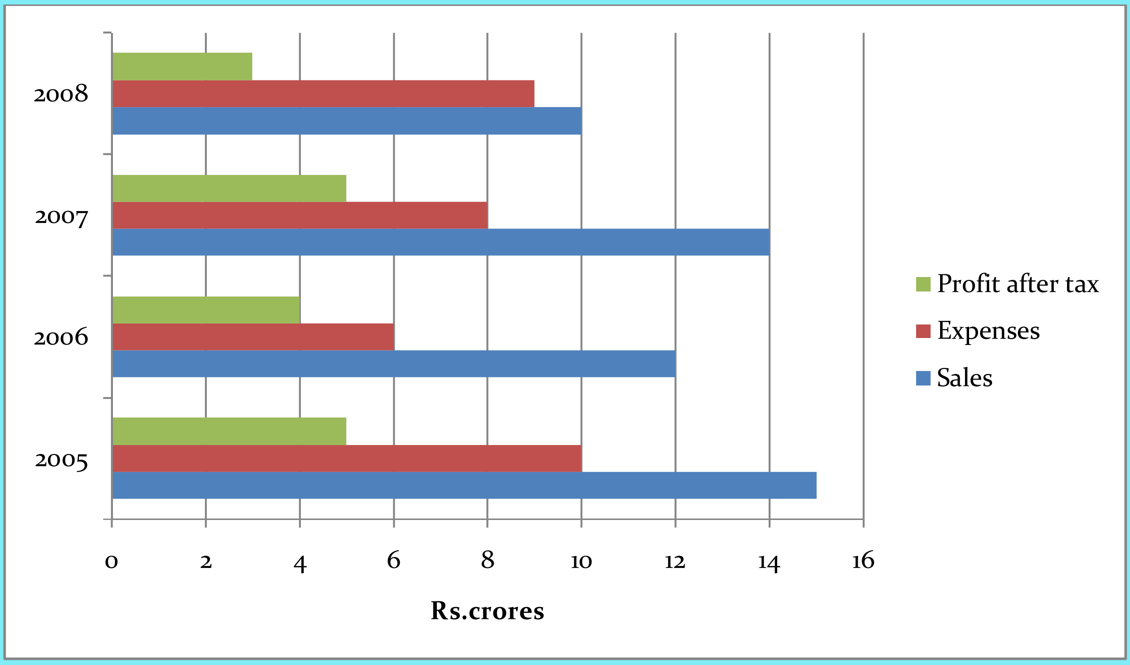


Figure 1.4

Question:

In which year did the profit after tax grow the least?

A.2005

B. 2006

C. 2007

D.2008

CUMULATIVE BAR GRAPH

In a cumulative bar graph, the measure of different quantities is represented by stacking the bars representing the different categories represented in the graph one on top of the other. In the cumulative bar graph given below, the bars corresponding to sales, revenue and profit are stacked on top of each other.

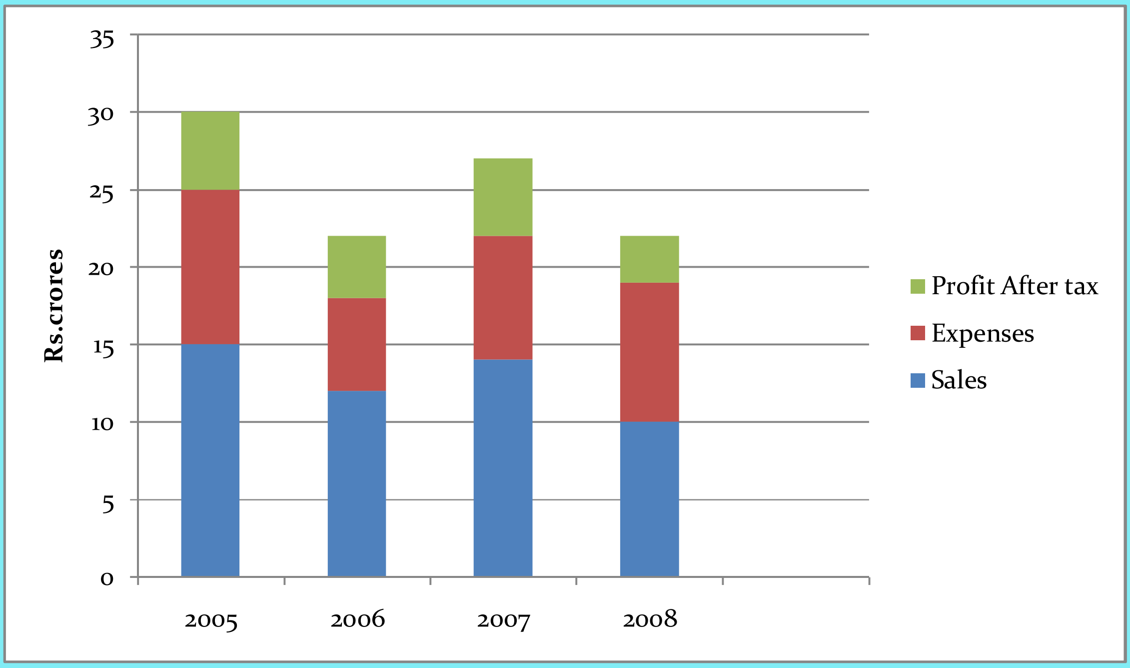


Figure 1.5

Question:

In which year is the ratio of expenses to sales the largest?

A.2005

B. 2006

C. 2007

D.2008

In figure 1.5, the values of sales, expenses and profit after tax are represented by a single bar. In such cases, the value of each quantity is obtained by calculating the difference between the upper limit and the lower limit of the stack representing the corresponding quantity.

For example, we can obtain the value of expenses in 2005 by calculating the difference between the upper limit and the lower limit of the bar representing the expenses. The upper limit is 25 crores and the lower limit is 15 crores. Hence the value of expenses in 2005 is 25-15= 10 crores. Similarly, we can obtain the measure of the other quantities in each year.

4. PIE CHART

A pie-chart is used to represent the distribution of data. It is a form of representation in which the variables are distributed over 360 degrees of a circle and the complete circle represents the total quantity of the variables.

Pie-charts are generally used when there is a need to represent the break-up of different contributing components. Some examples of instances when pie-charts are used to represent the data are: market share of various competitors, revenue from different units of a company, voting pattern etc.

The distribution of the various quantities may be represented in a pie-chart using percentage values, absolute values or the area that corresponds to the contribution from each component in degrees.

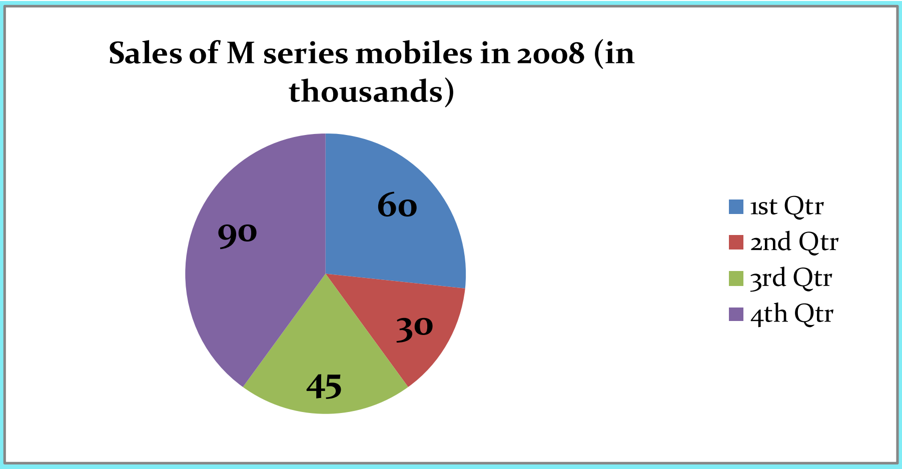


Figure 1.6

Question:

Which quarter had the highest growth?

A. 1st

B. 2nd

C. 3rd

D. 4th

The pie-chart shown above represents the sales of M series mobiles in the four quarters in the year 2008 (in thousands).

From the pie-chart, we can find the number of M series mobiles sold in each quarter in 2008.

For example, the number of M series mobiles sold in 3rd quarter is 45000.

We can also find the total number of M series mobiles sold in 2008 by adding the number of those sold in each quarter.

The same pie-chart can also be represented in percentage terms. It will be as given below.

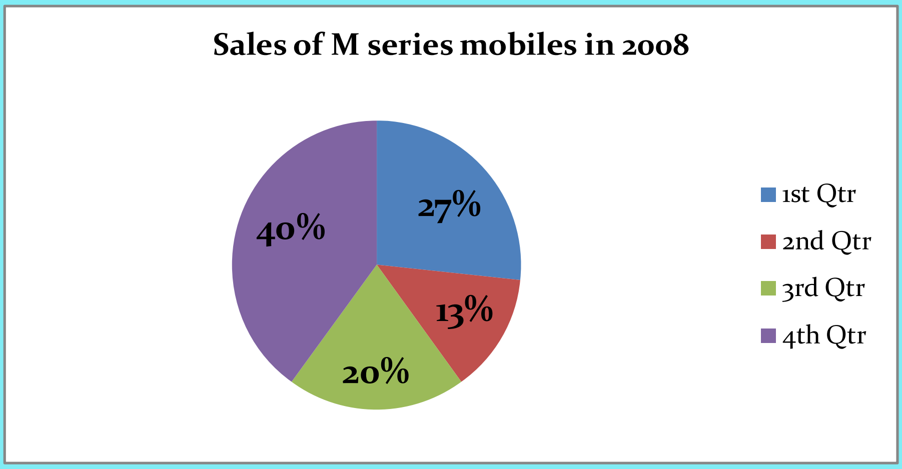


Figure 1.7

Question:

If sales in 2nd quarter is Rs.4000 crores then what is the sales in 3rd quarter?(in crores)

A.Rs.5000

B.Rs. 3000

C.Rs. 2000

D.Rs.1200

While the previous pie-chart was based on the absolute value, the pie-chart above represents the percentage of M series mobiles sold in each quarter in

2008.

Same pie-chart in degrees

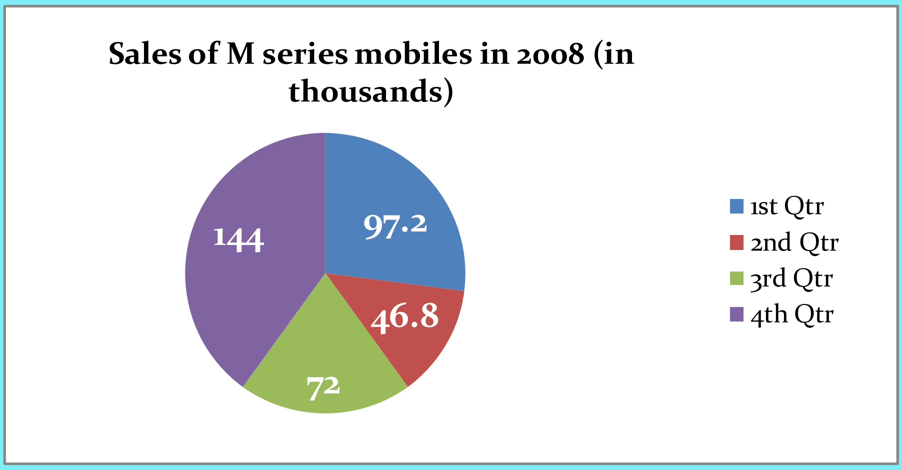


Figure 1.8

If sales in 2nd quarter is Rs.4000 crores then what is the sales in 3rd quarter?(in crores)

A.Rs.5000

B.Rs. 3000

C.Rs. 2000

D.Rs.1200

The sum of all the angles should be equal to 360'

What to look for in a PIE-CHART?

Title

As is the case with any form of data representation, the title is important in getting to know about the nature of data represented.

Legend

It is important to check the legend of the pie-chart to find out which portion corresponds to which category. It is very crucial in interpreting the data.

Method of representation:

One should pay attention to the method of representation in a pie-chart by observing whether the data represented is in absolute numbers, percentage or in terms of degrees.

5. RADAR GRAPH

Radar graph uses a radial grid to display data. This graph is so called because it resembles a radar screen. It is also sometimes called as web chart or spider chart.

In a radar graph, the centre of the graph is taken as the reference point. Values increase in a gradual manner as we move away from the center of the graph. The grid lines are present around the centre point. Each variable is represented by one grid line. The radial grid forms an equilateral polygon. The number of variables determines the number of sides in the polygon.

An example of a radar graph

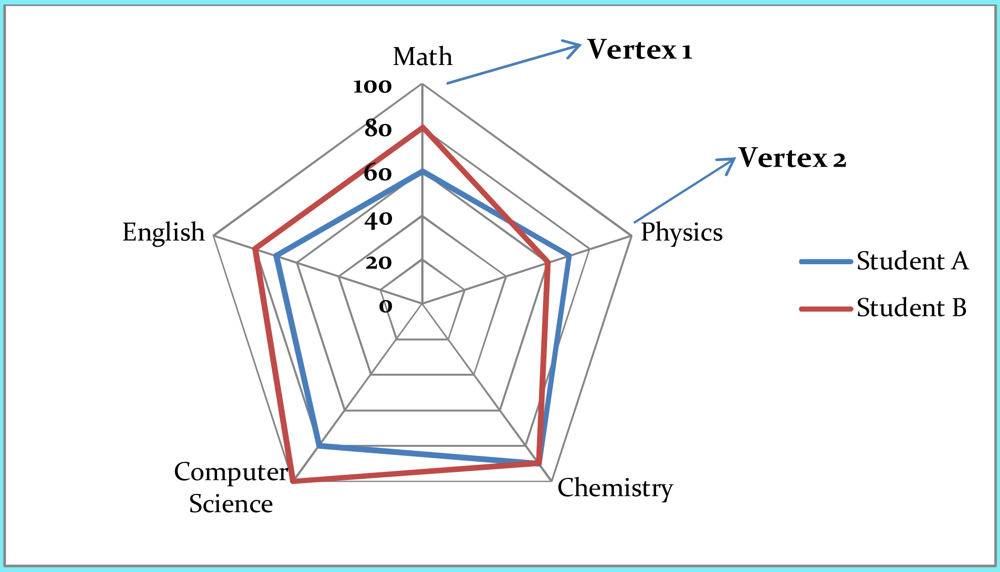


Figure 1.9

Student A scored more in how many subjects than student B scored)

A.1

B.2

C.3

D. more than 3

The radar graph shown above gives the information about the marks scored by two students A and B in 5 subjects – Math, Physics, Chemistry, Computer Science and English.

The centre of the graph is taken as the reference point or zero marks. The line joining each vertex to the centre of the polygon serves as the scale for the variables being represented – the marks scored in the different subjects. For example the line joining the top vertex (**Vertex 1**) of the pentagon to its centre is the scale representing the marks scored in Math by student A and student B. As we move away from the centre, the value increases. In figure 1.9, there is a gradient of 20 marks between two successive points on the scale.

How to interpret data in a radar graph?

While a radar graph may appear complex on observing it for the first time, it is easy to interpret it if one understands how the data is represented.

From figure 1.9, let us find the marks scored in Math by student A and student B. The line joining the top vertex (**Vertex 1**) of the pentagon to its centre is the scale for the marks scored in Math by student A and student B. The line graph representing student A touches the scale indicating the marks scored in Math at 60. So we can find out that student A has scored 60 marks in Math. Similarly we can find that student B has scored 80 marks in Math.

Let us take another example. The line joining Vertex 2 to the centre is the scale for representing the marks scored by the students in physics. The line graph representing student A meets this scale at a point between the grid lines representing 60 and 80. So we can conclude that student A has scored 70 marks in physics. Similarly we can find that student B has scored 60 marks in physics.

What to look for in a RADAR GRAPH?

Legend

Read the legend carefully to find out which line represents which quantity. In figure 1.9, the blue line graph represents student A and the red line graph represents student B.

Variables represented

It is important to understand the variables represented in a radar graph. In Figure 1.9, there are five variables – marks scored in the five subjects: Math, physics, chemistry, computer science and English. Since 5 variables are represented in the radar graph, the radial grid lines enclose a pentagon. If 6 variables were used, the grid lines will enclose a hexagon.

Scale/Axis

Observe carefully as to which part of the radial graph represents which variable. In figure 1.9, the line joining Vertex 1 to the centre of the circle is the scale used to represent the marks scored in Math.

Also observe the gradient in the radar graph shown in Figure 1.8, the gradient between two successive points on the scale is 20. Clear understanding of the scale is important to determine the values that lie between two radial grid lines. For example, the line graph representing student A meets the scale representing marks scored in physics at a point between the grid lines representing 60 and 80. So we can conclude that student A has scored 70 marks in physics.

6. SCATTER CHART

A scatter chart displays data as a set of points. Values are represented by the position of the points on the chart. Categories are differentiated by representing the points indicating their values uniquely. For example, a small cross may be used instead of a dot, or a square shaped point etc. Scatter charts are typically used to compare aggregated numeric values, such as scientific or statistical data.

The scatter chart shown below is used to represent the rainfall in Colombo in a particular month.

Rainfall in the month of September, 2008

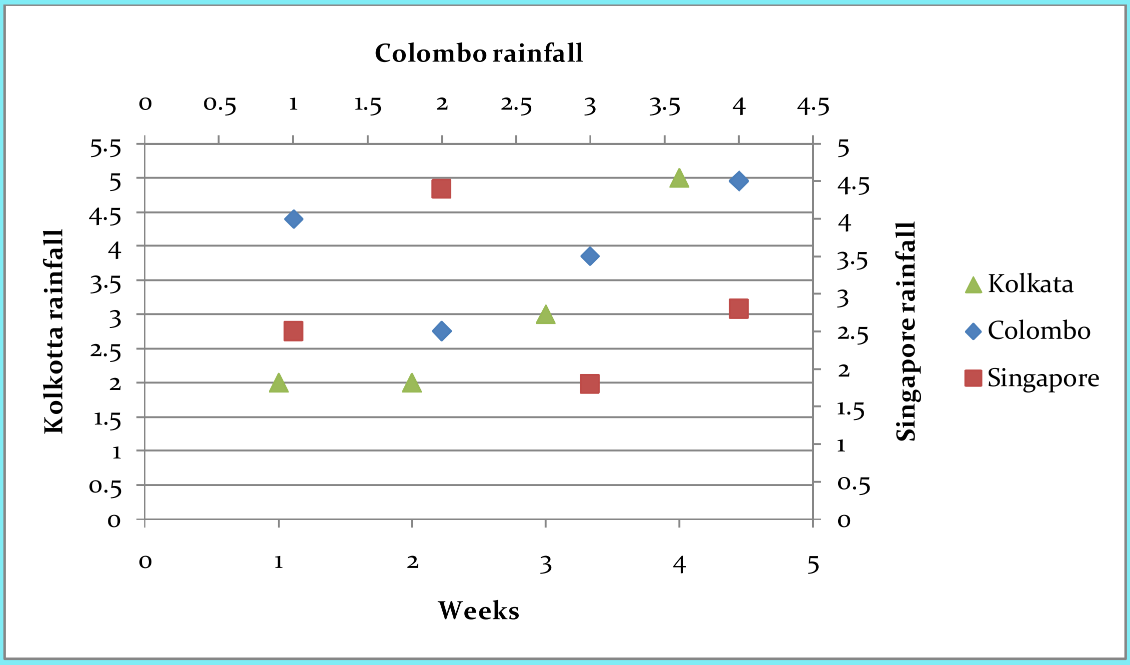


Figure 1.10

Which city had the highest rainfall in the 4th week (percentage wise)?

A.Singapore

B.Kolkotta

C.Colombo

D. Both Colombo and Singapore

The scatter graph shown above displays the rainfall in Kolkata, Colombo and Singapore in the month of September, 2009.

The X-axis represents the 4 weeks in the month for which the rainfall data was recorded.

The X–axis represents the rainfall received (in cm).

From the Figure 1.10, we can find the rainfall in Kolkata, Colombo and

Singapore in each week of September 2009.

For example, we can find that Kolkata received 2 cm rainfall in the first week of September 2009.

TITLE:

Read the title of the line graph to understand the nature of data represented by it.

In figure 1.10, the title of the scatter graph is '**Rainfall in the month of September, 2008.**'

X-axis & Y-axis label

The label of the X-axis and Y-axis gives us the information about the two parameters on the basis of which the graph has been plotted.

In Figure 1.10, the X- axis label corresponds to the Weeks for which the data was recorded and the Y-axis refers to the rainfall recorded (in cm).

Always check for the units of the label while observing what the label of each axis represents.

Markers

In a scatter graph, the various quantities being compared are differentiated by the markers used to represent them. For example, in figure 1.10, a triangular marker is used to represent data corresponding to Kolkata; a square marker is used to represent data corresponding to Singapore and a diamond shaped marker is used to represent data corresponding to Colombo.

7. Area Charts

An area chart displays data in a continuous format. Each category is represented by an area enclosed within an upper boundary and a lower boundary. The boundary lines correspond to the upper and the lower limit of the values for a category that is being represented. The area chart is generally used to represent the fluctuations in the value of the categories being represented over a period of time.

The area chart is a useful method of representing data where the total data as well as the proportion of each contributing component to the total needs to be analysed.

The stacked area chart shown below represents the variations in the share prices of three different companies between January – April, 2009.

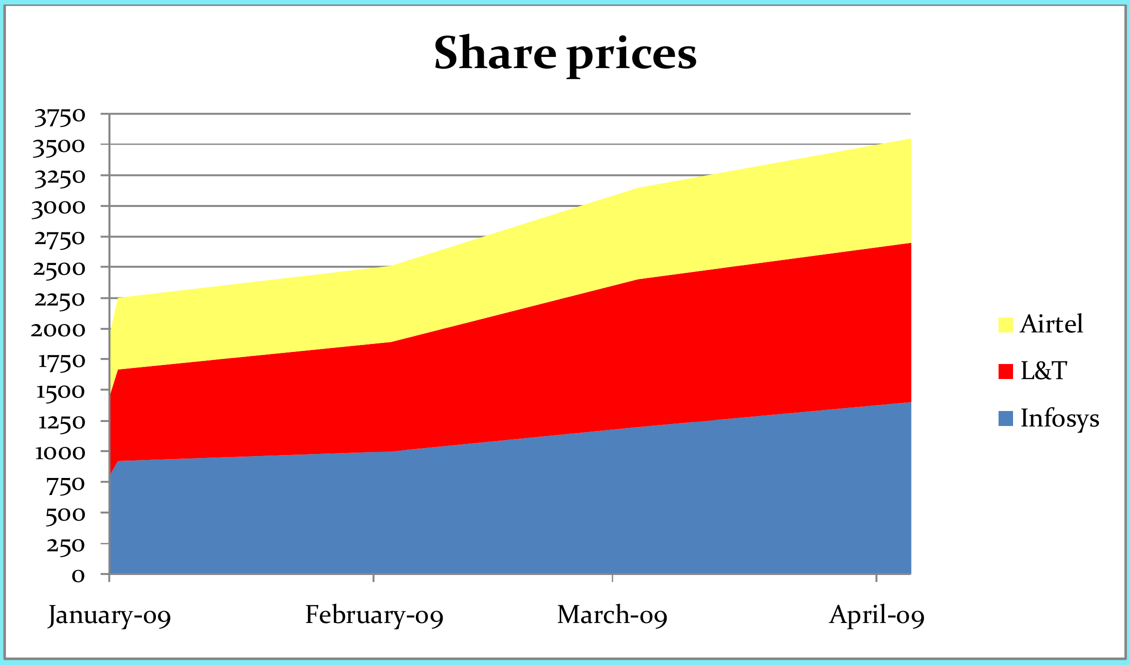


Figure 1.11

In which month did the share price of Infosys increase highest percentage wise?

A.January

B.February

C.March

D. April

From the area chart shown above, the least and the highest values and the range over which the value of the categories being compared varies is evident.

For example, it is evident from figure 1.11 that the share prices of Infosys and varies between 800 and 1400.

While calculating the values for other categories, we need to calculate the difference between the upper limit and the lower limit of the area representing the category.

For example the lower limit of the area representing L&T is 650 and the upper limit is 1950. Therefore share price varies by 1300 between January-April 2009.

What to look out for in area chart?

Legend

Check the legend to see which portion of the graph corresponds to which category. In figure 1.11, the area filled with blue colour corresponds to Infosys, area filled with red colour corresponds to L&T and the area filled with yellow colour correspond to Airtel.

Upper limit and lower limit

Always check the upper limit and the lower limit of the area corresponding to the category under consideration to calculate the values. Do not take the values mentioned in the graph as such.

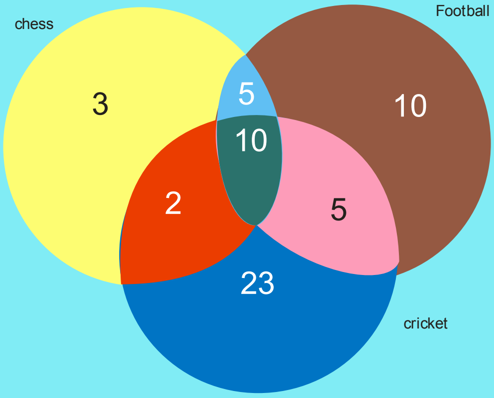
8. VENN DIAGRAMS

A Venn diagram is used to represent the logical relationship between various groups of things or sets. There are formed by representing information about a particular group within an enclosed region which is usually in the form of a circle.

Venn diagrams are formed by drawing two or more overlapping circles that are used to represent information about a group and the relationship between the various groups that are being represented.

An example of a Venn diagram is shown below.

The Venn diagram shown below gives the information about number of students playing cricket, football and chess in a particular class.



What is the total number of people playing at least one game?

A.20

B.30

C.40

D. None of these

The three circles represent the three distinct groups being represented.

The circle on the top right represents those playing football, the circle on the bottom represents those playing cricket, and the circle on the top left indicates those playing chess.

The students playing only football is represented by the region shaded in brown. It does not share any common area with the other two circles.

Similarly, those playing only cricket is represented by the region in dark blue and the students playing only chess is represented by the region shaded in yellow.

The region of intersection between two groups gives the number of elements common to both groups. For example, the region shaded in red is common to cricket and chess. It represents the number of students playing both cricket and chess.

Similarly, the region shaded in purple is common to both cricket and football. It represents the number of students playing both cricket and football.

The region shaded in green is common to all three circles. It represents the number of students playing all the three games.

From the above Venn diagram, we can find out the following information.

Number of students playing all three games (region in light green) = 10

Number of students playing **only** cricket **and not any other game** (region in dark blue) = 23

Number of students playing **only** football **and not any other game** (region in brown) = 10

Number of students playing **only** chess **and not any other game** (region in yellow) = 3

Number of students playing cricket and football (region shaded in purple) = 5

Number of students playing cricket and chess (region shaded in red) = 2

Number of students playing football and chess (region shaded in light blue) = 5

Total number of students playing cricket is given by

Number of students playing cricket only (region in dark blue) = 23

+

Number of students playing cricket and football (region shaded in purple)= 5

+

Number of students playing cricket and chess (region shaded in red) = 2

+

Number of students playing all three games (region in light green) = 10

which equals 23+5+2+10 = 40.

Similarly total number of students playing football = 30

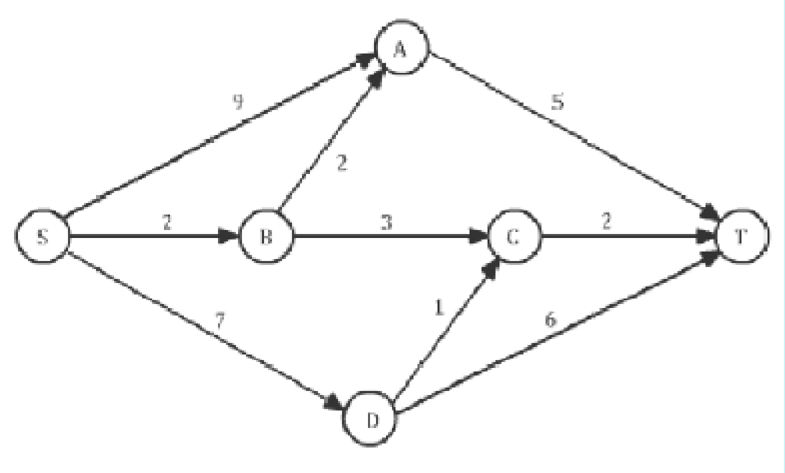
Total number of students playing chess = 20

9. NETWORK GRAPHS/ FLOW DIAGRAMS

Flow diagrams are used to represent information where weights are attached to the connectors between two nodal points. Some of the examples where flow diagrams are used to represent the information are: the routes between two places, shortest distance between two points, supply of commodities through a pipeline from a common storage point to other places etc.

Example of a simple flow diagram is shown below:

A significant amount of traffic flows from point S to point T in the one-way street network shown below. Points A, B, C, and D are junctions in the network, and the arrows mark the direction of traffic flow. The fuel cost in rupees for travelling along a street is indicated by the number adjacent to the arrow representing the street.



What is the most economical root?(economical= fuel cost is the least)

A.SBCT

B.SAT

C.SDCT

D. SDT

From the above flow diagram, it is evident that traffic flow is possible only in one direction.

We can calculate the minimum cost for travelling between two points based on the information provided in the network graph.

For example, if we want to move from S to T at the minimum cost, we need to go from S to B, B to C and C to T. The cost incurred will be 7 rupees.

What to look out for in a network graph/flow diagram?

Check the direction of the connectors; see whether the arrows are unidirectional or bi-directional.

Check for the weights associated with each connector.

In some cases, restriction may be imposed on using a particular connector, take note of them while analysing the problem.

10. CASELET FORM OR NON-GRAPHICAL FORM OF DATA INTERPRETATION

In this form of data representation, the information is presented in the form of short passage of random data. So the data presented in a caselet is not obvious by mere observation.

In order to make better sense from a caselet, we can structure the data by drawing a table or graph so that information retrieval becomes simpler.

Consider the caselet given below:

Volta limited tripled its sales from that of 2006 to that in 2008. However, the aggregate profit was one-eigth the total sales in 2006. The expenses of the company grew by 5 crores in 2007 and 3 crores in 2008. Sales in 2008 were 6 crore more than expenses in that year. Sales less expenses is profit. The company sold products worth 8 crores in 2006.

What is the total profit (in crores)?

A.1

B.2

C.3

D. 4

It is given that the company sold products worth 8 crores in 2006 and it tripled in 2008. So sales in 2008 was 24 crores.

In 2008, sales was 6 crore more than expenses. It is also given that Sales – expenses = Profitr. So profit in 2008 was 6 crores and expenses were 18 crore.

In 2008, expenses increased by 3 crores. So expenses in 2007 were 15 crores.

In 2007 expenses had increased by 5 crores, so expenses in 2006 were 10 crores.

In 2006, expense was 10 crores, sales was 8 crores. So profit = sales – expenses = (-2) crores

Total profit = one-eight sales in 2006 = (1/8)*8= 1 crore.

We have found that profit in 2006 = (-2) crores, profit in 2008 = 6 crores.

Also sum of profits in 2006, 2007 and 2008 = 1 crore

So profit in 2007 = 1 – [6 + (-2)] = (-3) crore

So sales in 2007 was 12 crores

Tabulating these results, we get

Year	Sales (IN CRORES)	Expenses (IN CRORES)	Profit = Sales- Expenses (IN CRORES)
2006	8	10	-2
2007	12	15	-3

2008	24	18	6

What to look for in a caselet

In caselets, the data given is usually not in the most simplified form. You may have to construct tables or graphs in order to interpret the data.

Look for statements that will give you clues to unravel the entire picture and help you in the process of simplyfying the data representation.

Log on to www.semanticslearning.com to download the next DI chapter - "conversion"