WEB PAGE FOR CHAPTER 7

SPSS ACTIVITY

Using the SPSS Chapter 7 Data File B

1 Obtain frequencies, descriptive statistics, and a variety of graphs and charts for the data for the variables *age, hours worked per week* and *main method of transport*. Choose the descriptives, graphs and charts you deem most appropriate.

2 Make new variables called '*qTotal*' and *qMean* which will provide the total and mean for each of the two questionnaire items *Qu1* and *Qu2*.

3 When you have done this, play around with the data. The best way to learn how to use SPSS is simply play around and see what it will do.

MULTIPLE CHOICE QUESTIONS

1 Statistics that describe typical scores are called:
(a) models
(b) standard deviations
(c) Z scores
(d) measures of central tendency

2 Statistics which describe the dispersion of scores around a typical score are measures of:
(a) deviation
(b) variability
(c) dispersibility
(d) skew

3 The term average is synonymous with the statistic called:
(a) the mode
(b) the mean
(c) the median
(d) the variance

4 The sum of deviations around the mean must equal:
(a) 1.0
(b) –1.0
(c) 0
(d) a value that varies with the number of scores

5 The difference between the highest and lowest values is called:
(a) interquartile range
(b) the range
(c) the quartile range
(d) the semi interquartile range

6 When all scores have the same value:
(a) the distribution is abnormal
(b) the mean, median and mode are unequal
(c) the standard deviation is equal to zero
(d) the distribution is unimodal

7 For which of the sets of values is the variance largest?
(a) 2,2,3,4,4
(b) 3,3,3,3,3
(c) 1,1,3,5,5
(d) 1,2,3,4,5

8 If the variance is 16 then the SD is:
(a) 2
(b) 4
(c) it depends on sample size
(d) 8

9 For which of the following sets of scores would the value of SD be the smallest:
(a) 1114777
(b) 1234567
(c) 7777777
(d) 1112333

10 The disadvantage of using the range as a measure of dispersion is:
(a) It is difficult to calculate
(b) It is heavily influenced by extreme values
(c) It is determined by only two points in the data set
(d) It changes from one sample to the next

11 A mean can be calculated on data from which of the following scales?
(a) nominal and ordinal
(b) interval and ordinal
(c) interval only
(d) ratio and interval

12 In checking for input errors, which of the following would you use?
(a) a box plot
(b) a stem and leaf plot
(c) a histogram
(d) all the above

13 If you have ordinal data, which measure of central tendency would you use?
(a) the mode
(b) the standard deviation
(c) the median
(d) the mean

ADDITIONAL QUESTIONS AND ACTIVITIES

1 On what type of graph would it be appropriate to show the number of gold, silver and bronze medals won by a country's athletes at the Asian Games.

2 Explain how a stem and leaf display contains more information than a grouped frequency distribution.

3 What kind of chart can be used to display negative quantities? Give two examples of negative quantities that arise in business.

4 Show the data below in a pie chart after calculating the percentages.

Household expenditure per week

Food	\$130.00
Housing	\$536.00
Clothing	\$ 34.00
Travel	\$100.00

5 A survey of where overseas tourists like to spend their vacations in your country was taken to determine what attractions should be advertised. The results were

national parks	15%
beaches	48%
cities	8%
golf courses	25%
mountain climbing	4%

Construct a bar chart to display the results.

6 A study of the number of 'bugs' in software programmes analysed over the last eight years were:

8.5%, 8%, 7.5%, 7%, 6.4%, 6%, 5.6%, 5%.

Construct a line graph and comment on the chart.

7 A sample of N = 20 scores has a mean of 5. What is the $\sum X$ for this sample.?

8 Find the median of this distribution:

3, 10, 8, 4, 10, 7, 6

9 In a recent survey comparing picture quality of three brands of colour television, 63 people preferred brand A, 29 preferred B and 58 preferred brand C. What is the mode for this distribution?

10 Find the mode of the following distribution:

5, 6, 9, 11, 5, 11, 8, 14, 2, 11

11 Taking a Traffic Census. Which is the most commonly used method of personal transport in your area? Which are the busiest times of day? Develop a traffic census form for your own area. Produce sufficient copies to monitor every 30 minute period from 8 am to 5 pm. and arrange a rota to cover each period at the same point on a main road with separate recorders on each side of the road. Produce a report with appropriate tables, graphs ands charts on:
(i) the popularity of each particular type of transport
(ii) the pattern of traffic flow during the day
Stages in this project are:
(a) Preparation of the census form
(b) Selection of an appropriate census point. Bear in mind the safety of the census takers during rush hour
(c) Testing the form out on a half-hour period to discover any weakness
(d) Drawing up a rota of recorders to cover the time period
(e) The taking of the actual census
(f) Analysis of data to find the usage of various types of transport
(g) Analysis of data to reveal hour-by-hour movement of traffic
(h) Production of the report

12 Survey of the Price of Houses. Obtain a copy of the local paper which contains house sales. Draw up a list of housing categories based on (a) the number of bedrooms, e.g. 1 bedroom, 2 bedroom.... over 5 bedrooms, and (b) types of housing, e.g. apartment, single storey house, double storey house….. Record the asking price for every house in each category. Calculate the mean price for each category. Display your data using appropriate tables, graphs and charts.

13 Examine some recent newspapers and magazines. Find some graphs, histograms etc. and consider whether they are providing information in a clear and simple way and that they are not conveying an erroneous impression. If you find any that do not appear to be clear, simple and intend to deceive, work out better ways to display the information.

14 Measure some variable that every person in your class possesses, e.g. height, weight, shoe size. Form a histogram or bar chart (depending on the type of data) and comment on any features of interest in the distribution.

15 Survey 40 people (20 males and 20 females) on the way they feel about the economic policies of the government. Use the following categories: excellent, above average, average, below average, poor. Construct a histogram showing the differences in opinion between males and females.

16 Explain in your groups what way the standard deviation as a measure of dispersal is different from the measure of range.

17 The representative of the employees' union in a small company complains that the average salary there is only £18,000. The owner of the company counters with a statement that the average salary is a whopping £25,000. The salaries are as follows: Discuss in your groups who is lying?

X	*f*
£70,000	1
39,000	1
21,000	1
19,000	1
18,000	5
11,000	1

ADDITIONAL MATERIAL

WAYS OF MISREPRESENTING NUMBERS – Misuse of descriptive statistics and graphical representations in reports

Numbers presented in chart or graph form often cause people to suspend judgement and common sense, *Homo sapiens becomes homo credens*... ready to believe anything. 'A picture is worth a thousand words' it is often said. But what if it is a wrong or misleading picture? A good piece of advice for all users of other people's numbers, charts and graphs is: *'Do not believe everything you see, especially the parts you cannot see!'*

Advocacy statistics

What has to be acknowledged at the outset is that a great many descriptive statistics are used not in order to reach a valid conclusion but to prove the case, however invalid, the presenter wants to make. The real difficulty with much of the statistics presented for business, political, and social purposes lies in the difference between *technical truth* and *advocacy* truth. Technical truth should always include qualifying statements of uncertainties. The providers of much of our daily intake of statistical data are usually advocates of causes or points of view, either as protagonists or antagonists. In a selling situation (and it does not really matter whether you are selling products, policies, presidents, or prime ministers) one is advocating a case, where not only are the qualifications and uncertainties excluded but aspects of the technical truth which might be damaging to the presentation of the case are withheld and only the positive points of importance to the case are made.

Emotive words, expressive phrases, are employed to create a more desirable impression than the bare facts warrant. Users of numbers are in the same position, and businesses concerned with the sale of products and services rely heavily on the art of advocacy. This is not to say that every statistic is intended to deceive, but there are many ways of presenting the same set of data. As a future manager or business person, the best safeguards you have is to possess sufficient knowledge and understanding of the basic techniques of collecting, processing and presenting data, including their uses and limitations, to have the confidence to question them and challenge them so that your business will not suffer through wrong decisions being made on the basis of deceptive statistics.

Here is a basic checklist of questions that a user of statistical information should employ to assess the material being displayed to them by others.

The data

- What do they tell us?
- Who said so? Is the source stated?
- With what degree of confidence?
- Relating to what time period?
- Is there a short written or verbal summary?
- How precise are the figures?
- What degree of accuracy is acceptable?
- Were there sufficient data to justify the conclusion?
- Do they answer the question we want to answer?
- What is the range of the data?
- Which average is used – mean, median or mode?
- Percentage of what?
- Compared with what?

Charts and graphs

- What does it say?
- Does it say it clearly enough? Does it make sense?
- Why this particular form of chart or graph?
- Is there a better way of displaying the data?
- Is anything missing?
- Do comparative charts use the same scale?
- Would they be better on log scale?
- Is the comparison valid?
- Is the conclusion valid?
- Is there a brief written or verbal summary?
- Is the source stated?

About samples

- What kind of sample is it (e.g., random, quota)?
- How big is it?
- When did the sampling take place?
- What degree of confidence can we have in the results?
- Are the results generally applicable?
- Are the sample details included in the report?
- Who is included in the sample and why?
- Who is excluded from it and why?
- How many people responded?
- How many did not reply?

It is not enough just to ask the questions. You have to know why you are asking them, sound as though you know why you are asking them, and understand the response and its adequacy. When putting the questions, you must give a firm impression that you know what you are about.

As a general rule no set of data, either in tabular or graphic form, should ever be presented without comment, on the basis 'Here it is – sort it out for yourself'. Researchers should 'sort it out'' and provide a brief written or verbal summary of what the data show, why the data have been summarized the way they have been – i.e., reasons for choosing the particular representative figures – and what the main conclusions are that can be drawn from them. Increasing demands are placed on the creativity of those presenting raw data, central tendencies and percentages to come up with new and innovatory displays that can mislead the unwary. But remember that in every complex chart there is a simple statistic trying to get out.

Following are some of the most common traps to be looked for.

Spurious precision – misleading means and plausible percentages

There is something very impressive and reassuring about a number taken to several places of decimals. For example, *'In a recent test, Loo brand antiseptic disinfectant killed 99.25 per cent of all known germs–dead!'* Really!! What about the other 0.75 per cent? But, in any case, does 'really matter? Or what about an advertising claim that 'Independent laboratory tests' (how reassuring, especially when you see the white-coated doctor-like figures in a clinical setting) show that children using Den Clean toothpaste have suffered 21.75 per cent fewer cavities over a trial period as compared with a control group using an *'ordinary' brand of toothpaste'*? Does it really matter whether the figure is 21 per cent or 22 per cent? The statement that the average cinema-goer (whoever he or she may be) visits the cinema 12.48 times a year is not only saying something which is physically impossible but implies a degree of precision which is meaningless. It may be quite wrong to infer that the average cinema-goer visits on average a cinema once a month, since the pattern of visits may be four times a month over a three-month dark, wet and cold winter period and no visits during the rest of the year.

These are examples of spurious statistics which purport to say something terribly meaningful but offer a degree of precision totally unnecessary for the purpose in hand

The pricing of many products provides another example of spurious precision. 'Shoes for only $59.99 per pair!' 'Drive away this car for $24,995 all-in!' What do you suppose is behind this pricing psychology''? Could it be intended to make us feel that we are getting a bargain - and so much better value than for $60 or $25,000''?

Precision should not be confused with accuracy. – Indeed, precise figures tell us nothing about their accuracy. If a survey shows that the per capital consumption of chocolate last year was 2.423812 kg, this is very precise indeed – but, is it accurate?

Many percentages are quoted out of context or with important additional or qualifying background information omitted. Here are some examples:

'Aircraft passenger deaths up by 10 per cent over last year.'

Yes, but what was the increase in the number of aircraft passenger miles flown during the two periods being compared. Were there any unusual or special factors operating such as one major crash which can distort figures?

'Consumer advertising up by more than 300 per cent between 1996 and 2006.'

So what? There must be very few things that have not trebled in value over this 10 year period. The percentage is undoubtedly the most widely used statistic. Whenever we want to compare different sizes or different shares or different rates of change, we call on percentages. Index numbers (e.g., cost of living, retail prices, wage rates, production, imports and exports), markups, discounts are all variants of the percentage. The only defence against the phoney percentage is to ask questions: 'percentage of what?' or 'of whom?'; 'Compared with what or whom?'; 'How many make up the whole 100 per cent?'; 'Who are they?'

Playing fast and loose

Loose comparisons are fairly common in marketing and advertising circles. If a dog food advertisement proclaims *'Now twice as meaty*' what does it really mean? – *'twice as meaty as what'?; twice as meatier as before?; twice as meatier in terms of' the measured meat equivalent in the average dog food, in all other dog foods, or the dog food with no meat in it at all*? What does '50% more coffee flavour' mean? What about *'Puffin cigarettes give you 10 per cent less tar'?* Product claims like these always beg the question: better than, more than, less than . . . what really is the comparison? Unless the standard or yardstick against which the comparison is made is provided, then such statements have no real statistical meaning. Such comparative statements, of course, are intended to make a good and positive impression.

Scaling the heights – keeping histograms and polygons honest

It is possible to draw graphs for the same group of data which give entirely different impressions – some are deliberately misleading. Graphs of the same data can convey entirely different impressions, as shown in Figure 1 (a) and (b), which report crime statistics for three similar suburbs. In A, cruising police patrol cars were eliminated during a 3 month trial period; B had 5 cruising cars during the period; while C was flooded with 15 cars. Your conclusions about the effects of patrol cars would probably depend on which graph you saw.

Figure 1(a) gives the impression that the presence or absence of patrol cars is associated with dramatic difference in crime rate. Note, however, that the largest difference – 1,000 versus 970 – is only 3%. Such a small difference could just as well be attributed to chance factors or to differences in crime reporting procedures. The graph is misleading because it violates the height-width rule and because the Y axis begins with a frequency of 950 crimes instead of 0 crimes.

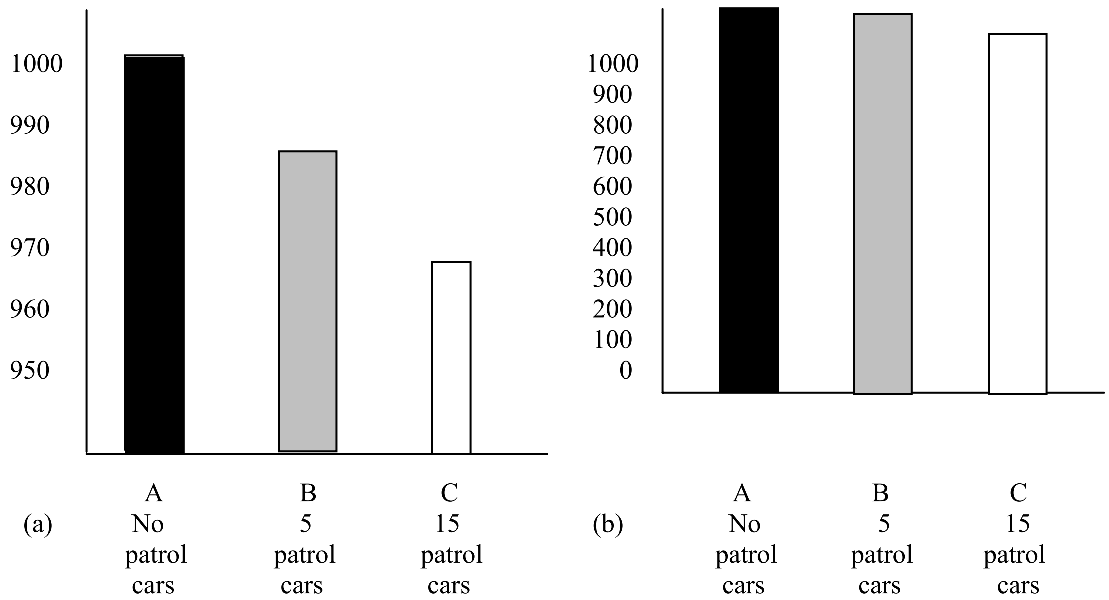

Figure 1(a) and (b) **Number of reported crimes in three similar neighbourhoods during a 3 month test period. Note how graph (a) falsely gives the impression of a great difference in crime rate across the three conditions.**

In another example, a government Finance Minister seeking a further period of office, might show the state expenditures during his term as illustrated in Figure 2. His opponent might show the same expenditures as illustrated in Figure 3

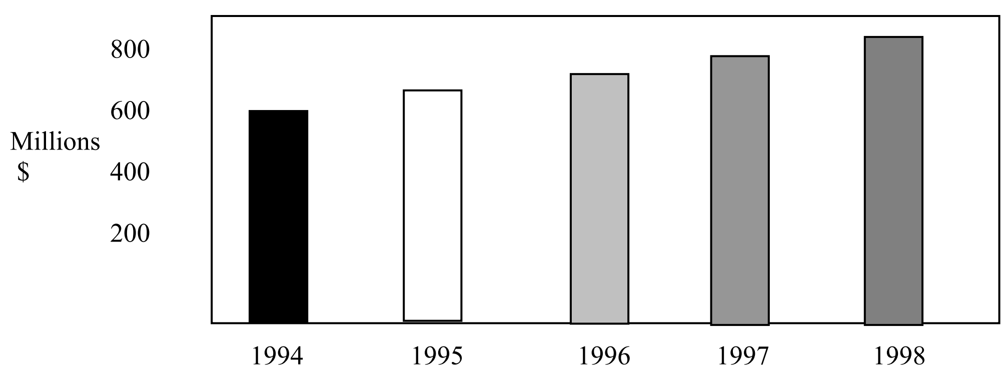

Figure 2 **Finance Minister's version**

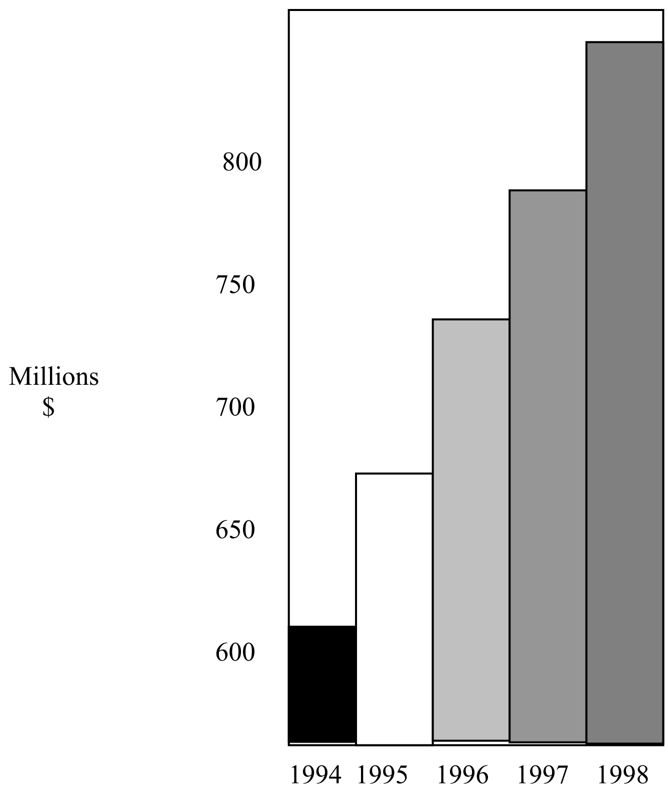

Figure 3 **Opponent’s version.**

Both histograms convey exactly the same information, but the Minister’s histogram provides an impression of minor and gradual increases in expenditure while the opponent’s emphasizes steeply rising expenditure. Essentially, the impression conveyed by a graph depends largely on the proportions chosen for the abscissa and ordinate. This illustrates why it is suggested that as a general rule a histogram or polygon should be about three-fourths as tall as it is wide, because this gives relatively undistorted proportions.

It is also recommended that the units on the abscissa and ordinate begin at the zero point. However, this is not always possible especially with score values on the abscissa. If a part of the abscissa or ordinate has been omitted, then the reader's attention should be called to this fact by means of slash marks drawn to cut the axis.

Most people in most countries of the world read horizontally but when it comes to numbers in tables and graphs they tend to think vertically. When we look at a graph we think in terms of 'up'– i.e., the vertical height of a point or line above the horizontal axis indicating the higher value or the bigger change in the value of, by convention, the dependent variable. Here is an example of distortion with a simple line graph.

Figure 4 shows three simple line graphs relating to the level of advertising in an organization's expenditure over the period from 1996 to 2006.

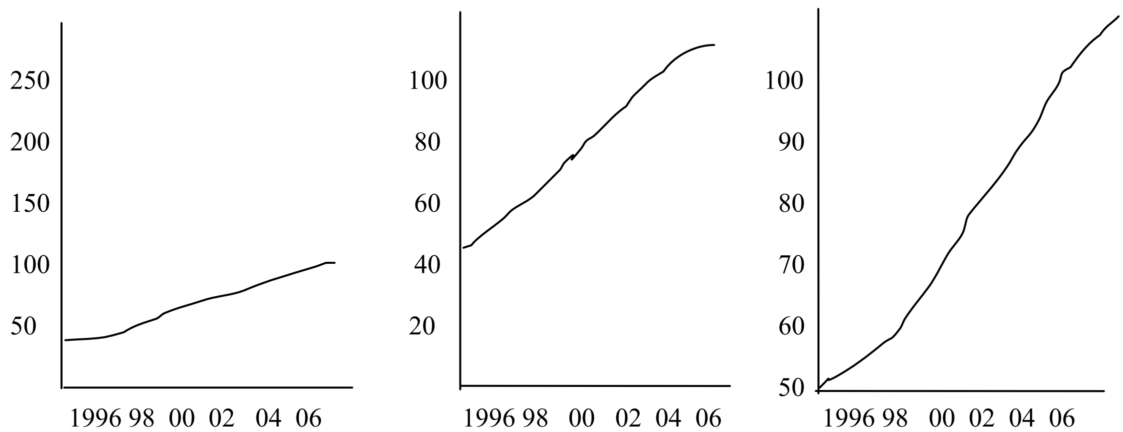

Figure 4 **Distortion with scales.**

All three graphs are based on exactly the same data but visually they tell a different story. Which graph would you expect a large shareholder who is critical of the small amount of advertising to use, and which would you expect the responsible advertising manager to use in his response? Taken in isolation, there is nothing wrong with any of the graphs, but the impression given by each is markedly different. This is done by progressively (from left to right) cutting off the extreme ends of the vertical scale, which immediately causes the line to rise more steeply. By compressing the horizontal year-line, it could be made to rise even more sharply.

Another trick which can be used is to begin the horizontal scale above the zero line. It probably would not matter too much provided any comparative graphs were lined up exactly the same way. The point is, however, that only one of the graphs would be used by the pro- and anti- advertising lobbies.

Changing the scale

When comparing graphs, they should all should use the same vertical scale, which is only practicable when the values are roughly within the same range, or they should be reduced to a uniform scale which provides a fair comparison (i.e. ratio or log scale). While every sales and marketing manager looks for the silver lining in the company's sales charts, they would be well advised to look for the scalar lining as well. In other words, beware of the 'up' line and, equally, of the 'down' line that has been doctored to appear worse, by comparison, than the facts warrant.

To take an example, suppose that two product managers are each putting forward a request for bigger marketing allocations for their separate products, based on past performance and future prospects. If each product manager makes his individual presentation to senior management, what sort of impression will they each make? The chances are of course, that the sales of each product will not be within the same range, in which case distortion will have a serious effect on making sound comparisons.

Figure 5 shows two alternative presentations to marketing management by two product managers, each responsible for products X and Y respectively. They each decide to show the other's figures on the same graph for comparison purposes when making their pitch for a bigger marketing budget.

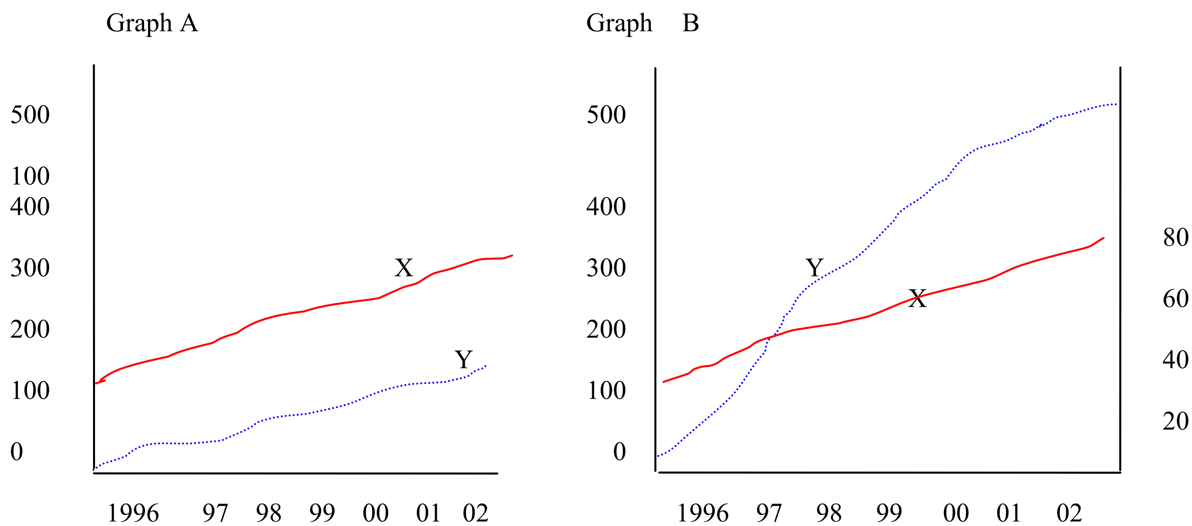

Figure 5 **Sales of two products compared (arithmetic scale).**

The manager of product X produces graph A, based on a common arithmetic scale. The manager of product Y produces graph B, again using an arithmetic scale, only this time the vertical scales are different. He tried several alternatives but settled for this one because it put his product in the most favourable light in terms of relative growth. How can a reasonable decision be made? Only by placing them both on the same log scale.

Clearly, the way in which information is presented can depend very much on the provider's or commentator's point of view and the case that they wish to advocate. There is nothing wrong in this, provided that the receiver or user of the information knows what the data represent and what has been done to them in tabular or graphic terms.

Big is beautiful

When does two times two equal eight? When it's a pictogram of the type illustrated in Figure 6. a type much used by advertisers in particular..

What has happened in Figure 6 is that the product manager has doubled all the dimensions of the washing machine symbol. The volume of the washing machine on the right-hand side of the pictogram is, in fact, eight times that of the one on the left-hand side. In other words, the pictorial ratio is 1:8, whereas the sales ratio is 1:2. Since most people will make the comparison on a volume basis, the drawing gives a totally false impression of growth. Leaving out the third dimension (depth) does not provide the answer, since the areas of the front faces of the washing machine symbols are still in the ratio of' 1:4. Even putting volume figures above each of the symbols (e.g., 10,000 and 20,000 respectively) does not overcome the deception. If the manager wishes to show that sales have doubled then the correct way is to use a simple bar chart.

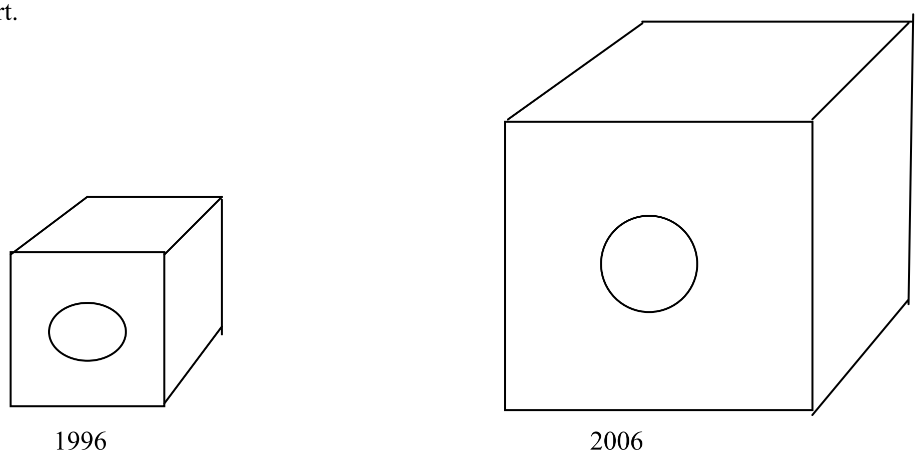

Figure 6 **Sales of Swirly washing machines has doubled.**

All the foregoing examples of' the statistical traps that lie in wait for the unsuspecting user of numerical data illustrate one fundamental principle. In business affairs, it is wise to assume that if anyone can draw the wrong conclusions from a given set of data, then they probably will (a variant of' Murphy's Law).

ANSWERS TO QUESTIONS IN CHAPTER 7

Qu. 7.1 (a) 5. Qu. 7.1 (b) bimodal

Qu. 7.2 18

ANSWERS TO MULTIPLE CHOICE QUESTIONS IN CHAPTER 7

1 (b), 2 (c), 3 (c), 4 (c), 5 (c), 6 (b), 7 (b), 8 (c), 9 (c), 10 (d), 11 (b), 12 (c), 13 (b), 14 (d).

ANSWERS TO WEBSITE QUESTIONS

1 Bar chart or pie chart.
2 You can see the individual values and determine the mode.
3 Two-directional horizontal bar chart. Financial losses; customer decrease compared to previous year.
4 Percentages are 16.25%; 67%; 4.25%; and 12.5%.
7 100
8 7
9 Brand A
10 11

ANSWERS TO MULTIPLE CHOICE QUESTIONS ON WEBSITE

1 (d), 2 (b), 3 (b), 4 (c), 5 (b), 6 (c), 7 (c), 8 (b), 9 (c), 10 (c). 11 (d), 12 (d), 13 (c)