Additional material for projects

Contents:

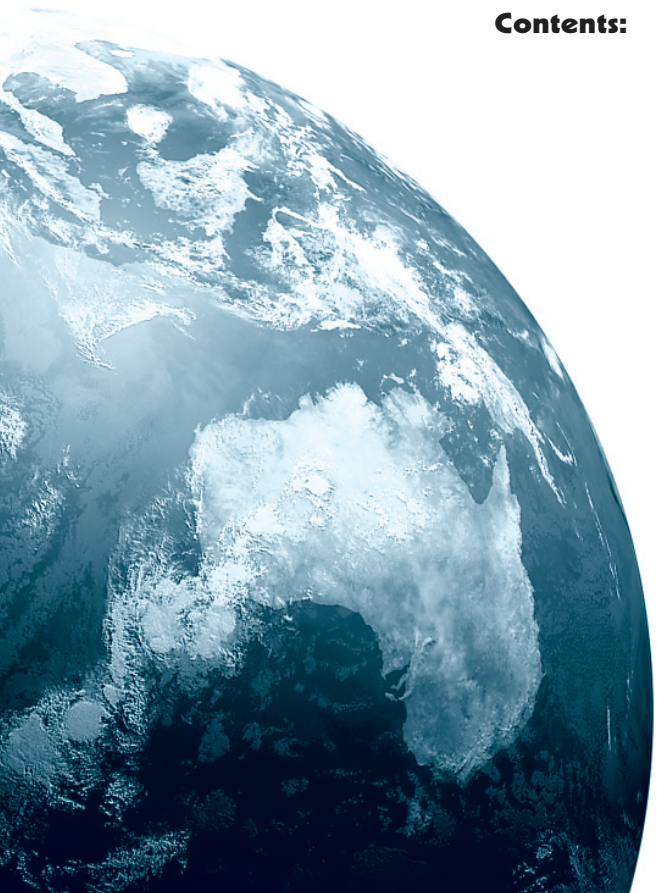
A Statistical enquiries

B Populations and samples

Investigation: How large must a sample be?

C Abuse of statistics

D Projects / Investigations



Facts and figures are part of the flow of information we are bombarded with each day.

For example,
- The median house price has increased by 50% in the last 10 years.
- If an election was held tomorrow the two major parties would get 89% of the vote.
- Yesterday was the coldest August day since August 4th 1897.

Figures like those given above tell us something about the topic being discussed.

Some statements, particularly those about weather, cannot be made without gathering a large amount of information over a long period of time. We also have to know if the statements are reliable.

The facts or pieces of information are called **data**.

Data may be collected by counting, measuring or asking questions.

[One piece of information is known as one piece of *datum* (singular), whereas lots of pieces of information are known as *data* (plural).]

If we collect information about the weights (to the nearest kg) of students in our school, we have a number list such as 53, 57, 69, 63, 48, 56, 56, 43, 57, 57, ...

This number list is called a **data set** and because it is not in organised form it is called **raw data**.

STATISTICS

Statistics is the art of solving problems and answering questions by collecting and analysing data.

Statistics are used by governments, businesses and sports organisations so that they can make informed decisions when they are providing services such as in health, transport and commerce or developing new tactics. They are also interested in using statistics as a means of analysing the effects of certain changes that may have been made, or in predicting what may happen in the future.

Businesses use market research surveys to determine consumer preferences of goods and services.

Statistics are used by manufacturers for quality control and in medical research to test new drugs.

A medical researcher may believe that a newly discovered drug could prolong the life of heart attack patients. To prove this, the drug would have to be given to a group of heart attack sufferers. Statistics which compared their quality of life and life expectancy to another group where the drug was not given would be obtained. If the effect of the drug improved life expectancy and quality of life with no bad side effects it would almost certainly become a legal drug.

In statistical work we use tables, graphs and diagrams showing data collected from scientific experiments, and figures from economics, public opinion polls, census returns and many other situations.

HISTORICAL NOTE

- The **Babylonians** (before 3000 BC) recorded yields for their crops on small clay tablets.
- Pharoahs in ancient **Egypt** recorded their wealth on walls of stone.
- Censuses were conducted by the **Ancient Greeks** so that taxes could be collected.
- The book of Numbers in the Old Testament records the results of two censuses of the **Israelites** taken by **Moses**.
- At the time of the birth of Jesus, **Emperor Augustus** ordered a census to be taken throughout the **Roman Empire**. This is why Mary and Joseph travelled to Bethlehem.
- After **William the Conqueror** invaded and conquered England in 1066, his followers overtook estates previously occupied by Saxons. Confusion reigned over who owned what.

  In 1086, William ordered that a census be conducted to record population, wealth and land ownership. A person's wealth was recorded in terms of land, animals, farm implements and number of peasants on the estate. All this information was collated and has become known as the "Domesday Book". It is regarded as the greatest public record of Medieval Europe.

  The Domesday Book is displayed in the Public Record Office in London.

OPENING PROBLEM

A city school can be easily accessed by train and school bus as well as by walking and riding in a private car. The school is interested in finding out how students get to and go home from school as there are local traffic problems in the area during these times. For safety reasons parents and bus drivers are requesting that the school provides a bus/car pull-in bay.

An initial survey of 75 students was carried out and the results in coded form were:

T T B W W	T W C W C	C C B C B	W T W C C	W C C B B
B C B B W	W B T B B	W B C B C	C W B W T	T C B B T
B W T C B	B C C C C	W C W T B	T T T C C	C C C T W

(T = by train, B = by school bus, C = by car, W = by walking)

Questions to consider:

- There are 859 students attending the school. How were the students selected for the survey? Would the survey represent the whole school if the students were only selected from Year 12?
- Is the sample large enough to reflect the method of travel for all students in the school?
- How could we best organise this raw data?
- How could we display this information graphically?
- What calculations could we perform on the organised data to make it more meaningful?
- What likely conclusions could we make and report to the school?

A STATISTICAL ENQUIRIES

Let us consider the steps involved in a statistical investigation.

1 Examining a problem

There is no interest in gathering data just for the sake of it. Which of the following problems may possibly be solved by examining data?

- If I open a shoe store, how many of each size shoe should I keep in stock?
- Is it true that there is global warming and the Earth's temperature is increasing?
- If the maximum speed limit is reduced by 10 km/h will this reduce the number of fatal accidents?
- What is the likelihood that a particular netball team will win next Saturday?

Once we have recognised the problem and written down a question we may proceed.

2 Collection of data (information)

Data for a statistical investigation can be collected from records, from surveys (either face-to-face, telephone, or postal), by direct observation or by measuring or counting. Data can be collected for the whole **population**, which generally means all the people or things that the conclusions of a statistical investigation would apply to. This is called a **census**. Or data can be collected for a **sample** of the population.

Collection of data is the first, and most important task in an investigation, because unless the correct data is collected, valid conclusions cannot be made.

3 Organisation of data

This process involves organising data into tables.

4 Summarising and display of data

We summarise data by counting it in some way and then we display the data with a suitable graph so that some of the features of the data are clearly visible.

For the two types of data that we study in detail (categorical and numerical) some types of graphs are appropriate and some are not.

5 Analysing data and making a conclusion in the form of a conjecture

There are some calculated quantities that are universally used to *describe* a set of data.

Calculating quantities that indicate the centre of the data (mean, median and mode) and the spread of the data (range, interquartile range and standard deviation) gives us a picture of the sample or population under investigation. Using these quantities gives a more satisfactory way of comparing two or more data sets and making a conclusion.

B POPULATIONS AND SAMPLES

When a statistical investigation is to be conducted, there is always a target *population* about which information is required.

The **population** is the total field from which the numerical data is being drawn.

The *population* might be the entire *population* of the country, the entire population of a school, an entire animal species, or the complete output of a machine making a particular item.

CENSUS OR SAMPLE

One of the first decisions to be made when collecting data is from whom, or what, the information is to be collected. There are two ways in which this can be done.
These are: a census or a sample.

A **census** is a method which involves collecting data about every individual in the *whole population.*

The individuals may be people or objects. A census is detailed and accurate but is expensive, time consuming and often impractical.

A **sample** is a method which involves collecting data about a *part of the population* only.

A sample chosen at **random** is one where each member of the population has **an equal chance** of being selected in the sample. Samples are selected in this way in the hope of removing any bias in the selection and so allowing the sample to be truly representative of the population.

A sample is cheaper and quicker than a census but is not as detailed or as accurate. Conclusions drawn from samples always involve some error.

However, the aim is to make the sample an accurate representation of the population so that the features of the sample's distribution are close to the features of the population's distribution. Hence, if the method of sample selection was successful, statistics such as median and range should have similar values for both the sample and the population.

Hence two types of **statistics** exist:

- **sample statistics**
- **population statistics** (more commonly called **population parameters**).

Example 1

Would a census or sample be used to investigate:

a the length of time an electric light globe will last
b the causes of car accidents in a particular state
c the number of people who use White-brite toothpaste?

a *Sample.* It is impractical to test every light globe produced as there would be none left for sale!

b *Census.* An accurate analysis of all accidents would be required.

c *Sample.* It would be very time consuming to interview the whole population to find out who uses, or does not use, White-brite toothpaste.

EXERCISE B.1

1 State whether a census or a sample would be used for each of the possible investigations:

- **a** the number of goals scored each week by a netball team
- **b** the heights of the members of a football team
- **c** the most popular radio station
- **d** the number of children in a Japanese family
- **e** the amount of sunshine in a day
- **f** the countries of origin of immigrants
- **g** the most popular colours of cars
- **h** the gender of school principals
- **i** the time spent doing homework
- **j** the marks scored in a class test
- **k** the items sold at the school canteen
- **l** the number of matches in a box
- **m** the reasons people use taxis

BIAS IN SAMPLING

The most common way of collecting information is by using a sample. The purpose of a sample is to provide an estimate of a particular characteristic of the whole population. Therefore the challenge in selecting a sample is to make it as free from prejudice as possible and large enough to be representative of the whole population.

A **biased sample** is one in which the data has been unfairly influenced by the collection process and is not truly representative of the whole population.

Example 2

Suggest the possible bias in each of the following samples:

- **a** a phone survey during the day
- **b** a survey of people on a train station
- **c** a survey of a football crowd

- **a** The sample would be biased towards people who are at home during the day, i.e., it does not include people who go to work.
- **b** The sample would be biased towards people who catch the train, i.e., it does not include people who use other forms of transport or work at home.
- **c** The sample would be biased towards people who attend football matches. For example, there would probably be more males than females at football matches.

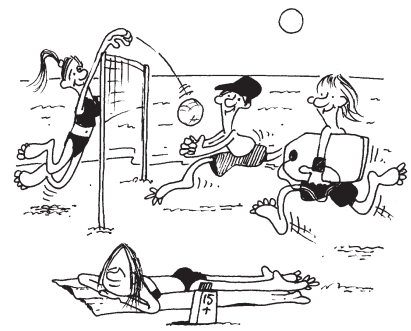
Sometimes people use **biased samples** to enhance their claims for their products or to support a particular point of view.

For example, a person wanted the local council to upgrade its swimming pool, and sampled the views of swimmers who used the pool.

In this case you would expect the people who use the pool to be biased very favourably towards this proposal, so the person taking the sample could be accused of producing an unfair report.

EXERCISE B.2

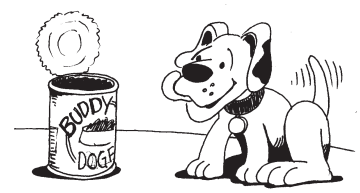
1 Explain and discuss any possible bias in the following samples:



- **a** a phone survey on a Saturday night
- **b** a survey of the people at a bus stop
- **c** a survey of the people in a supermarket carpark
- **d** a survey of people at the beach
- **e** a survey of people in your street
- **f** people selected from the electoral roll

2 Comment on any possible bias in the following situations:

- **a** Year 12 students are interviewed about changes to the school uniform.
- **b** Motorists stopped in peak hour are interviewed about traffic problems.
- **c** Real estate agents are interviewed about the prices of houses.
- **d** Politicians are interviewed about the state of the country's economy.
- **e** People are asked to phone in to register their vote on an issue.
- **f** An opinion poll is conducted by posting a questionnaire to people.
- **g** An advertisement claims that "Dog breeders recommend Buddy dog food."



3 A survey on the number of people unemployed in a certain city, is conducted by door knocking in a certain suburb each morning for a week.

- **a** Will the survey produce a fair or a biased result? Why?
- **b** Suggest a better way to conduct the survey.

4 A TV station conducts a survey on a current political issue by asking the viewers of a current affairs programme to telephone the station with their views. Would this produce a fair reflection of societies' attitude to this political issue? Give reasons for your answer.

5 A research company is asking peoples' opinions on whether smoking should be banned in all public places.

They ask people standing outside buildings in the city during office hours. Explain why the data collected is likely to be biased.

6 The residents of an electorate are surveyed to find how they will vote in the coming election. From the data collected they are to predict the election result in that electorate. Explain why each of the following situations would produce a biased sample:

- **a** A random selection of people in the local large shopping complex is surveyed between 1 pm and 3 pm on a weekday.
- **b** All members of the local golf club are surveyed.
- **c** A random sample of people on a train station between 7 am and 9 am are surveyed.
- **d** A doorknock is undertaken, surveying every voter in a particular street.

7 A school has 820 students. An investigation concerning the school uniform is being conducted.

a 40 students from the school are randomly selected to complete the survey on their school uniform. In this situation:

i what is the population size **ii** what is the size of the sample?

b Explain why data collected in the following situations would not produce a sample that is representative of the population.

i The surveyor's ten best friends are asked to complete the survey.

ii All the students in one class are surveyed.

iii Volunteers are asked to complete the survey.

A NOTE ON VARIABLES

Two variables under consideration are usually linked by the fact that one of them is *dependent* on the other.

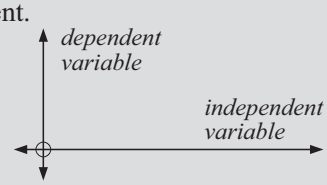
For example, the *total cost of a dinner* depends on *the number of guests present.*

We say that *the total cost of a dinner* is the **dependent variable**, and *the number of guests present* is the **independent variable**.

DISCUSSION

- Discuss the following sentences and find the dependent and independent variables:
  'The number of hours worked by a plumber affects the total charge.'
  'The amount received by each person in a Lottery syndicate is linked to the number of people in the syndicate.'
  'The diameter of a circular table top determines its area.'
- Discuss and write down *two* sentences which contain variables, and identify which of the variables is the dependent and which is independent.

Generally, when drawing **graphs** involving two variables, the *independent variable* is on the **horizontal axis** and the *dependent variable* is on the **vertical axis**.
An exception to this is when we draw a horizontal bar chart.



SAMPLING AND SAMPLE SIZE

When taking a sample it is hoped that the information gathered is **representative of the entire population**. We must take certain steps to ensure that this is so. If the sample we choose is too small, the data obtained is likely to be *less reliable* than that obtained from larger samples.

For accurate information when sampling, it is essential that:

- the individuals involved in the survey are **randomly chosen** from the population
- the number of individuals in the sample is **large enough**.

Different characteristics of the population need to be taken into account when choosing the sample. For example, a survey of society's opinion of the medical system would be pointless unless it included all groups within the society and not just one group such as pensioners, doctors, women, etc.

INVESTIGATION HOW LARGE MUST A SAMPLE BE?

Click on the icon to view a population of known mean $\overline{x}$.

What to do:

1. Select a sample of size $n = 2$ and find its mean $\overline{x}$.
2. Repeat several times. Comment on how $\overline{x}$ compares with the population mean.
3. Now select samples of size $n = 10$ and in each case find $\overline{x}$. Comment on how these $\overline{x}$'s compare with the true population mean.
4. Repeat for samples of size $n = 100$.
5. Write a brief report on your findings.

From the investigation you should have observed that:

> The larger the sample size, the closer the mean of the sample reflects the mean of the population.

This is true for other population characteristics, for example, the standard deviation. We examine the mean and standard deviation in greater detail later.

It is true to say that: "The greater the sample size, the more reliable will be our findings".

However, we must strike a balance between the confidence in the reliability of our results and the expense of carrying out a large sampling procedure.

SAMPLE SELECTION

For a sample to have the best chance of being truly representative of the population it should be chosen at random.

One way in which a sample can be randomly selected is to use a **random number table**. This is a series of numbers randomly generated by a computer.

In order to randomly select a sample, each member of the population is assigned a number. The numbers selected for the sample are then found by starting at any point in the random number table and listing the numbers as they appear.

Example 3

An insurance company wishes to survey its 8000 policy holders by choosing a random sample of 25. All the policy holders are given a number from 0001 to 8000. Use the random numbers shown below to randomly select the sample.

We need a starting point which can be anywhere on the table.
Let us choose row 02, column 03 of the random numbers listed below, i.e., the number 26255.

Since we require four digit numbers, we will use the first four digits 2625.
(We could equally have chosen the last four or any other random selection of four of them.)

We could continue across the rows or down the columns. Let us continue across the rows.

	00	01	02	03	04	05	06	07	08	09	10	11	12
00	15918	56193	90740	81701	07513	05148	80668	59006	15279	76364	95419	27400	21306
01	66847	29692	24172	13427	87722	34399	54242	33544	55618	42460	20254	37761	73373
02	15312	97320	01059	26255	73679	25899	73891	01513	78588	37745	16621	59996	38146
03	56132	97211	33124	33054	46190	55251	43718	73276	67869	65301	40166	93619	57456
04	92269	97087	08335	76911	65756	15502	67100	10384	00696	58199	95377	21668	21354

Thus the policy holders with the following numbers will be selected for the sample:

2625 7367 2589 7389 0151 7858 3774 1662 5999 3814
5613 3312 3305 4619 5525 4371 7327 6786 6530 4016
5745 0833 7691 6575 1550

Note: (1) The random numbers crossed out are those that exceed the size of the population.

(2) The sample represented 0.31% of the population $\left(\frac{25}{8000} \times 100\%\right)$.

EXERCISE B.3

Generate **random numbers** on your calculator and use them to randomly select samples in the following cases:

1 A survey of 851 students from a school is to be undertaken. 25 students are to be randomly chosen for a sample. Select the student numbers for your sample by moving from left to right and using only the first three digits in each number.

2 A company wishes to test the products from one of its machines which produces 7500 units each week. A sample of 20 units is considered adequate for testing purposes and these are to be selected randomly using random number techniques. Select the 20 numbers from the sample by using only the last 4 digits in each number.

3 A local council, population 39 895, wants to canvass the residents' attitude towards a development project under consideration. The council decides on a random sample of 50 people to gauge the reaction of the residents to this project. All residents are given numbers from 00001 to 39895 and then selected from the electoral rolls using random numbers. Select the numbers to use in the sample.

4 What percentage of the population is used in each of the sample selections above (to two decimal places)?

METHODS OF SAMPLING

If the population from which the sample is drawn is a large one, it would be a costly and time consuming exercise to assign numbers to each member of the population and then select numbers randomly from tables. The random process can be modified. These modifications are detailed below.

Multi-Stage Sampling

Multi-stage sampling is used when the population is very large (for example, a survey on the population of a country). To make the sample selection easier, the geographical area which contains the population is divided into smaller regions. These regions may then be further subdivided into smaller areas or zones. The sample would then be selected from these zones. The zones themselves may be randomly selected, and the sample taken from these zones likewise randomly selected.

A good example where this method is used is the **unemployment statistics** as compiled by the Australian Bureau of Statistics (ABS). The ABS divides the geographical region of Australia into smaller zones (for example, cities/towns) and then further subdivides these zones into smaller ones (for example, suburbs). The suburbs to be surveyed are then randomly chosen and the people to be used in the surveys are randomly selected from these suburbs.

Cluster Sampling

This method is very similar to multi-stage sampling in that the geographical region is again divided into smaller zones. However, when the sample is chosen it is drawn from members of the population who have **a certain characteristic** (for example two-income families). An example where this method would be used would be a Canada wide survey involving Non-English migrants. For this survey Canada would be divided into zones and probably divided into smaller zones. A sample of Non-English migrants would then be randomly selected from these smaller zones.

Systematic Sampling

This is similar to pure random selection in that the population to be surveyed is arranged in some order (for example, alphabetic, numeric, etc) and the first sample selection is then made at random. This could be done using random number tables. However, subsequent sample selections are made in some systematic fashion (for example, every 25th number after the first number is chosen). The advantage of this method over pure random sampling is that it is much simpler to operate and a lot less costly.

An example of where this method might be used is a sample taken of the members of a club or some organisation. For instance, the local Automobile Association may want to survey their 30 000 members about some traffic issue and decide that a sample of 1% (i.e., 300 members) would be large enough for their purposes. To select the sample, they would arrange their members in some order (such as membership number perhaps) and make their first selection randomly. Every 100th member after this point would then be chosen until all 300 were selected.

Stratified Sampling

If the population to be surveyed contains **distinct groups** whose characteristics are quite different, then it would be desirable if these groups were included in any sample taken. The sample then is more likely to be representative of the population.

For example, if 75% of the doctors are male, then a survey of doctors' opinions on alternative medicine should contain a sample that has 75% male and 25% female doctors. In this case, the population consisting of all doctors in Japan is divided into two groups (male doctors, female doctors). The sample for the survey would then be randomly selected from these two groups with 75% of the sample being selected from the male group and 25% from the female group.

Example 4

A survey about students' attitudes towards school uniforms it to be done at a school with the following population:

A sample of 50 students is to be surveyed. How many students from each year group should be included?

Year group	*Number*
8	212
9	175
10	170
11	125
12	112
Total	794

The sample of 50 should reflect the same proportions as the school population on a year to year basis.

212 out of 794 are year 8 students

$\therefore$ $\frac{212}{794}$ of 50 year 8s are needed

$= \frac{212}{794} \times 50$

$= 13.35$ $\therefore$ need 13 year 8s.

Year	*fraction (proportion)*	*Number*	*Selection*
8	$\frac{212}{794}$	$\frac{212}{794} \times 50 = 13.35$	13
9	$\frac{175}{794}$	$\frac{175}{794} \times 50 = 11.02$	11
10	$\frac{170}{794}$	$\frac{170}{794} \times 50 = 10.70$	11
11	$\frac{125}{794}$	$\frac{125}{794} \times 50 = 7.87$	8
12	$\frac{112}{794}$	$\frac{112}{794} \times 50 = 7.05$	7
		Total	50

The actual students selected from each year group would then be **randomly** selected, 13 year 8s, 11 year 9s, 11 year 10s, 8 year 11s, 7 year 12s.

QUOTA SAMPLING

In this method of sampling the sample size is set beforehand at a certain number, or quota, of items that have a **common characteristic** (for example, male or female, etc). The sample selections can be made at random but only from those members of the population that have this common characteristic. It is similar to a stratified sample.

An example of where this method could be used, would be a survey conducted by the management of a shopping centre on the difficulties encountered by disabled people when shopping. The people conducting the survey might then be given a quota, or set number, of shoppers to ask. Of course, the shoppers asked would have to be disabled people.

EXERCISE B.4

1 The auditor of a company wants to check a company's invoice system. The invoices are all prenumbered and are recorded numerically in the sales journal. Last year there were 5000 invoices processed. The auditor wants to check 1% of them, to see whether all the proper procedures have been carried out. Suggest two ways in which the auditor could select the sample.

2 A company employs 420 men and 530 women and conducts a survey on employees' attitudes to changes to the superannuation fund. If the survey is conducted over 40 employees and a stratified sampling method was used, how many men and women should be in the sample?

3 Unemployment statistics reveal the following details:

Age	*Male*	*Female*
15 - 19	20%	18%
20 - 26	12%	10%
27 - 35	8%	7%
36 - 45	10%	5%
over 45	5%	5%

A survey is conducted amongst unemployed people. In a stratified sample of 2960 people, how many should be:

a male **b** female

c male and between the ages of 27 - 35

d female and between the ages of 15 - 19?

4 The population of Australia (in 2003) is shown alongside:

State/Territory	*Population*
NSW	6 700000
Vic	4 900 000
Qld	3 800 000
WA	2 000 000
SA	1 500 000
Tas	480 000
NT	200 000
ACT	320 000
Total	19 900 000

An Australia-wide survey is to be carried out on the communities' feelings regarding the protection of the environment. The sample size decided on is 3180. How many from each state/territory should be included in the survey?

SURVEY DESIGN

Surveys are usually conducted in the form of a questionnaire. When designing a survey questionnaire, the following points would need to be considered.

- The scope of the survey. What information is required and why?
- Clear numbering of questions and answers, as this will help in sorting out the survey responses.
- Clear wording of questions so that a definite answer has to be given.
- Simple and easy to understand questions.
- Logical development of questions with the hardest questions at the end.
- Each question should only cover one point.
- Questions should not be too personal as they may not be answered.
- If an opinion is being sought, be clear about whether detailed information or a reaction is required.

EXERCISE B.5

1 Select a topic for a survey of your school and then:

a design your own questionnaire

b decide on your sample size and method of selection

c collect the data

d process and analyse the results

e compile a written report outlining any observations you made as a result of your survey. State any conclusions you deduced from your results.

C ABUSE OF STATISTICS

A famous English Prime Minister of the 19th Century by the name of Disraeli once said:

"There are lies, there are damn lies, and then there are statistics!".

What he was implying is that statistics are often manipulated and used to distort and mislead. It is important to always be aware of what statistics or figures are being quoted and how they are being used. Whenever statistics are presented to you, the following questions should be kept in mind before you allow the figures to sway you.

SAMPLES

Statistical conclusions are based on the results from a sample of the population, not the whole population. Thus, before accepting their conclusions we should be certain that the sample was taken at **random** and was **representative** of the population. We should also check the **size** of the sample. Too small a sample can product distorted results and misleading statistics.

Discussion:

(1) A community wishes to find out whether or not it is feasible to erect a Hall. They require projections into the future, ability to finance the project, etc. The committee, who are generally supportive of the project, decide to sample 10% of the population by way of a questionnaire. When the results of the questionnaire are examined, they are not as favourable as the committee would have liked. Consequently, the committee survey another group of the population, pick out the results which are favourable, delete the unfavourable results from the previous survey and finally come up with some "pretty good results". Comment.

(2) Discuss the following quotation: *"He uses statistics like a drunk uses a lamp-post; for support, rather than illumination."*

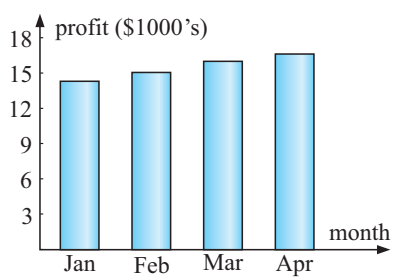
PRESENTATION

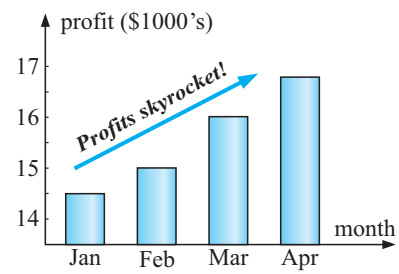
Statistical data can also be presented in such as way that a **misleading impression** is given.

- A common way of doing this is by manipulating the scales on the axes of a line graph.

For example, consider the graph shown.

The vertical scale does not start at zero. So the increase in profits looks larger than it really is. The break of scale on the vertical axis should have been indicated by ≷.



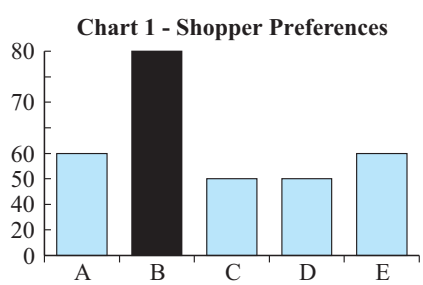


The graph should look like that shown alongside.

This graph shows the true picture of the profit increases and probably should be labelled '*A modest but steady increase in profits*'.

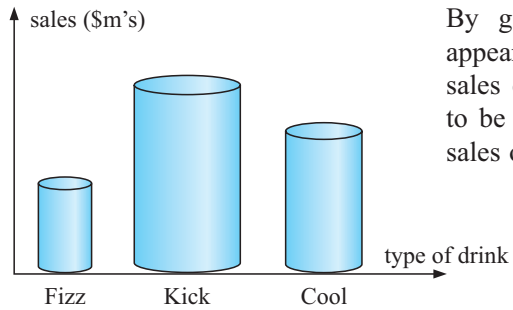
- These two charts show the results of a survey of shoppers' preferences for different brands of soap. Both charts begin their vertical scales at zero, but chart 1 does not use a uniform scale along the vertical axis. The scale is compressed at the lower end and enlarged at the upper end.

  This has the effect of **exaggerating the difference** between the bars on the chart. The bar for brand 'B', the most preferred brand, has also been darkened so that it stands out more than the other bars. Chart 2 has used a uniform scale and has treated all the bars in the same way. Chart 2 gives a more accurate picture of the survey results.

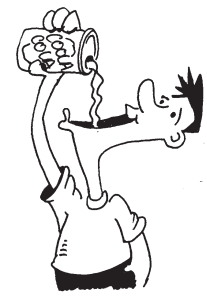


- The 'bars' on a bar chart (or column graph) are given a larger appearance by adding area or the appearance of volume. The height of the bar represents frequency.

  For example, consider the graph comparing sales of four different types of soft drink.

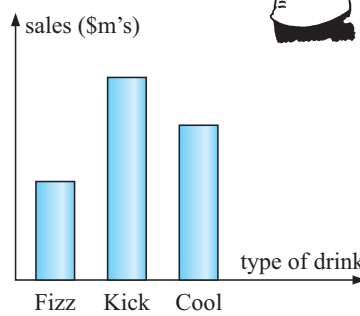


By giving the 'bars' the appearance of volume the sales of 'Kick' drinks look to be about eight times the sales of 'Fizz' drink.



On a bar chart, frequency (sales in this case) is proportional to the height of the bar only. The graph should look like this:

It can be seen from the bar chart that the sales of Kick are just over twice the sales of Fizz.

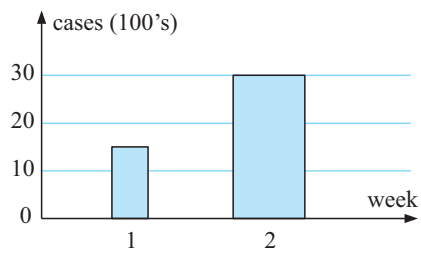


There are many different ways in which data can be presented so as to give a misleading impression of the figures. The people who use these graphs, charts, etc., need to be careful and to look closely at what they are being shown before they allow the picture to "tell a thousand words".

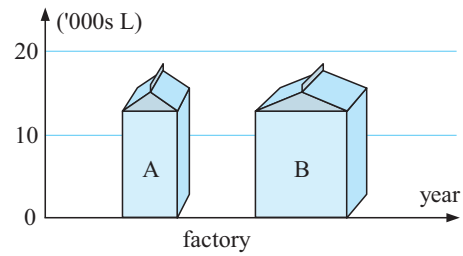
EXERCISE C

1 Describe the misleading or poor features of each of the following graphs:

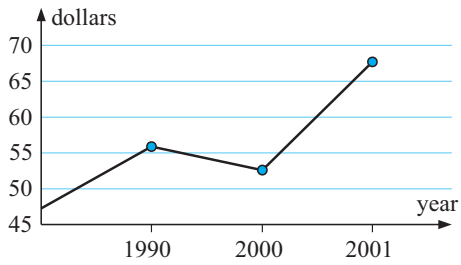
a **Fish sold at markets**



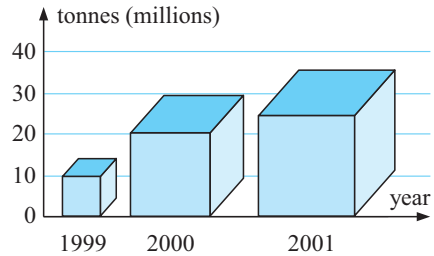
b **Milk production**



c **Interstate bus fares**

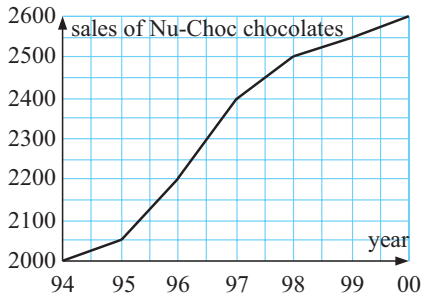


d **Exports**

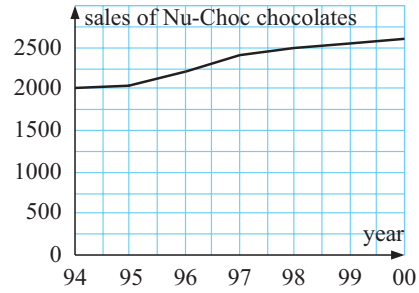


2

Graph A



Graph B



a Which graph appears to show rapidly increasing sales?

b Have sales in fact rapidly increased over this 7 year period?

c According to graph A the sales for 1997 appear to be double those of 1996. Is this true?

MEASUREMENT

Statistical measurements such as the mean, mode and median are quoted in order to give some meaning to the data. However, the **inappropriate use** of one of these measures can convey a misleading impression about the data. For instance, if the data under question is badly skewed (i.e., has extremes of values) then the mean would not be as good as, say, the median to represent the middle of the data. Consider the example of the selling prices of five houses: $80 000, $81 000, $82 500, $85 000, $190 000.

The mean selling price of $103 700 has been dragged up by the extreme value $190 000. In this case the median value of $82 500 seems more representative. A real estate agent wanting to encourage sellers may quote the mean to give the impression of higher selling prices.

So, you need to be aware that the figure being quoted is the appropriate one for the data under consideration.

BIAS

If the author of the statistics being presented has a particular point of view to push, they may present or interpret the figures in a way that supports their case or point of view. Being aware of the author's bias means you are less likely to blindly accept the author's interpretation and conclusions.

This bias will sometimes mean that **important information is ignored** because it does not support the author's argument or point of view. The author's case is thus strengthened if this information is ignored. You should, as far as possible, check to see that **all the facts** have been presented and not just those that support a particular viewpoint.

Being aware of any bias is vitally important when figures are based on **subjective** rather than objective information. The more subjective the information, the more chance there is for the author's prejudices to influence the results and their interpretation.

COMPARISONS

Statistics can be used to compare things and there are often many ways in which comparisons are unfair and inappropriate and thus misleading. An example of this is comparing the standard of living for countries using only the per capita income as a comparison (that is,

the total income of the country divided by the number of people in the country). Using this figure alone can give a misleading impression, because it does not take into account the cost of living, distribution of incomes, size of families, etc, within the country.

For example, country A's per capita income is $10 000 whereas country B's is $15 000. However the cost of living in B is twice that of A and the incomes in B are much less evenly distributed than they are in A. Country A, therefore may have a much higher standard of living than B even though their per capita income is less.

Another example of this is comparing monetary figures over time. To compare a sum of money at one point in time with another sum of money at a later date is misleading unless the **effects of inflation** have been taken into consideration.

Inflation reduces the purchasing power of money because as prices rise your money will not buy as much as it used to. $100 will not be worth as much in five years' time because it will not buy as much then as it could today. Therefore, when comparing money over a period of time make sure that the figures are in **real terms**, that is, adjusted to allow for the effect of inflation.

Comparisons over **different periods** can also be misleading unless you are aware of what is being compared. If a normal year is compared with an abnormal one then a distorted impression may be conveyed. For instance, the first year of a business's operation is not likely to be widely successful. It will take time for the business to become established and known and to build a reputation. The second year might be more successful as the business reaps the rewards of the first year's efforts. To compare the success of the two years would be most unwise.

Another pitfall to be aware of when comparisons are made is whether the **basis upon which the figures have been derived has changed over time**. It is not fair to compare figures unless they have been compiled in the same way. A good example of this is the inflation rate as measured by the Consumer Price Index (CPI). If the basis on which the CPI is calculated is changed from one year to the next, valid comparisons cannot be made. Another example could be unemployment figures collected by the government. If the definition of unemployed changes from one year to the next, this would alter the number of people counted as being unemployed. An unfair comparison would then result.

Finally, you should be careful of how **percentages** are used when comparing data. Inappropriate calculation or use of percentages can radically alter conclusions drawn from the data. For instance, the percentage of the government's budget spent on defence may increase from 10% to 15%. The government, wanting to 'minimise' the increase may claim that only a 5% increase over the last year had occurred (10% to 15%). However the truth is that a 50% increase has occurred. It is important to distinguish between absolute increases (5%) and relative increases (50%). How percentages are quoted may depend upon what point of view is being pushed, so you need to keep your wits about you.

CONCLUSIONS

Be aware that conclusions drawn from statistics are not necessarily correct. Make sure that the conclusions make sense and are **compatible** with the statistics they are based on.

For example, a wine company might show that wine consumption has increased and so has the average life expectancy of the population. However, it would be false to conclude that consuming wine increases a person's life expectancy. Just because two variables increase over time does not mean that they are related.

So, be careful with conclusions as they are often affected by the bias or prejudices of their authors. Try to reach your own conclusions rather than rely on someone else's. There are many tricks that can be used when dealing with statistics. The best defence you have against them is to be very careful and to always be aware of the dangers. Be aware of the biases of the person quoting the figures, and remember:

"lies, damn lies, and statistics!"

D PROJECTS/INVESTIGATIONS

1 There are two different types of data collected from surveys.

Primary data: This is when the person doing the survey goes out and collects the data from the sample of the population being surveyed.

Secondary data: Here the data has already been collected and the person starts the survey from this point.

Conduct a survey of your own either from the start (i.e., design a questionnaire, choose your sample, organise and analyse your data) or by making use of supplied data (for example, desired changes to the school canteen, desirability of holding a senior social; share prices and volume of sales of a company listed on the stock exchange; location of caravan parks).

Prepare a complete report on your survey using graphs and/or diagrams to present your data. Include statistical calculations (mean, standard deviation) etc., where appropriate, and any conclusions made as a result of your survey.

2 Find out about:

a the census **b** year books for your country.

Write a report of your findings.

3 Companies often use statistics in their annual report to help shareholders understand how the company is performing. Look at an annual report and write about the use of statistics that you find in it.

4 We have discussed the wide use that is made of Statistics in real life. In the newspapers and magazines that you read, look for examples of the use of statistics either in articles or advertisements. Examine the statistics closely, especially looking for bias and misleading use. Write briefly of your findings, including a copy of the statistics, if possible.

5 Conduct your own survey of some boxes of matches or sweets or other items where either the number or weight is given. Choose suitable statistical measures and charts to help you analyse and report the results of your findings.

6 Countries use their own share price index. Find out the name of the share price index used in Japan, the United States, England and your own country and how it is calculated. Prepare a report of your findings.

7 Sporting competitions use statistics, for example, batting averages, premiership tables etc. Choose a sport and investigate what statistics are collected and used. Give examples.

8 Use the results of the last census to investigate your local government area, for example, household incomes, number of cars per household, number of children per family, age distribution etc. Present the information you find in newspaper article form. Include tables and graphs.

9 Find out how the population of your country (or a state, city or town of your country) has grown (or changed). Present your findings on a suitable graph.

10 Graph the daily share price of a listed company for ten weeks. Find out what the term 'moving average' means and how it is calculated. Calculate a five day moving average for your share price data and graph your results on the daily share price graph. Comment on the difference between the two graphs and suggest a use for the moving average share price graph.

11 Statistics are useful to small business. Choose a particular type of small business, for example, delicatessen, lawnmowing round, etc., and investigate what statistics would be used (besides financial records) and what statistics of the local government area might be relevant.

12 A very newsworthy statistics item deals with the number of road accidents and the number of deaths on roads. Contact your appropriate source of information such as the Department of Road Transport to get copies of reports they produce. Prepare a poster for your school or class noticeboard.

13 Not all statistics collected are number-based. By collecting names from the birth notices in newspapers it is possible to find out the most popular names. Carry out such an investigation and present your findings in newspaper article form.

Alternatively, survey your school to find the most popular first names for students at your school. Write up your findings for the school newsletter or magazine.

14 Statistics are used to forecast or predict what might happen in the future. Choose a set of statistics, for example, school population number, rainfall for the local area etc. and apply appropriate statistical measures to your set of statistics to predict what might happen in the future.

15 Each day the Weather Bureau predicts the maximum (and minimum) temperatures for the next day as well as reporting the temperatures of the current day. Collect data and apply a statistical investigation comparing actual and predicted values. Present a report on your findings of how good the predictions are.

16 Critically examine how a TV/Radio/Magazine survey or poll is conducted. You should investigate the population, sampling methods, questioning methods and presentation of results. Discuss the fairness and value of the survey.

17 TV and radio stations are very concerned about "ratings". Find out what they are and how information is collected. Write a report of your findings.

EXERCISE B.1

1 **a** census **b** census **c** sample **d** sample or census
e census **f** census **g** sample **h** sample **i** sample
j census **k** census **l** sample **m** sample

EXERCISE B.2

5 Those standing outside the office building are more likely to be smokers since smoking is banned in the workplace.

6 **a** Only people who do not work between 1 pm and 3 pm would be selected.

b People who play golf are not likely to be representative of the whole population.

c These people are likely to be students or the employed, not the unemployed, senior citizens etc.

d Only those people home at the time will be surveyed, and the people in that street may not have opinions indicative of the general population (street in a poor area, rich area etc.).

7 **a** **i** 820 **ii** 40

b **i** The best friends of the surveyor would probably have opinions like his.

ii The students in one class will not represent students of all ages within the school.

iii People who would volunteer to fill in the survey would be likely to have strong opinions and so not represent the population.

EXERCISE B.3

4 **a** 2.94% **b** 0.27% **c** 0.12%

EXERCISE B.4

2 18 men, 22 women

3 **a** 1628 **b** 1332 **c** 237 **d** 533

4 NSW 1071, VIC 783, QLD 607, WA 320, SA 240, TAS 77, NT 32, ACT 51

EXERCISE C

1 **a** The fish sold at Market 2 looks to be 4 times as much as at Market 1 whereas it is actually only double.

b A and B produce the same quantity of milk, but B looks more because a bigger carton is shown.

c Because the y-axis does not begin at zero, it looks as though the increases are far greater than they really are.

d The width of the boxes is increased, not just the height, so the increase in exports looks greater than it really is.

2 **a** A **b** no, only about 30% **c** no