Background knowledge in statistics

Contents:

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.

TYPES OF DATA

Data are individual observations of a **variable**. A variable is a quantity that can have a value recorded for it or to which we can assign an attribute or quality.

There are two types of variable that we commonly deal with:

CATEGORICAL VARIABLES

A **categorical variable** is one which describes a particular quality or characteristic. It can be divided into **categories**. The information collected is called **categorical data**.

Examples of categorical variables are:

- *Getting to school:* the categories could be train, bus, car and walking.
- *Colour of eyes:* the categories could be blue, brown, hazel, green, grey.
- *Gender:* male and female.

QUANTITATIVE (NUMERICAL) VARIABLES

A **quantitative (numerical) variable** is one which has a numerical value and is often called a numerical variable. The information collected is called **numerical data**.

Quantitative variables can be either discrete or continuous.

A **quantitative discrete variable** takes exact number values and is often a result of **counting**.

Examples of discrete quantitative variables are:

- *The number of people in a household:* the variable could take the values 1, 2, 3, ...
- *The score out of* 30 *on a test:* the variable could take the values 0, 1, 2, 3, ... 30.

A **quantitative continuous variable** takes numerical values within a certain continuous range. It is usually a result of **measuring**.

Examples of quantitative continuous variables are:

- *The weight of newborn babies:* the variable could take any value on the number line but is likely to be in the range 0.5 kg to 8 kg.
- *The heights of Year 8 students:* the variable would be measured in centimetres. A student whose height is recorded as 145 cm could have exact height between 144.5 cm and 145.5 cm.

Example 1

Classify these variables as categorical, quantitative discrete or quantitative continuous:

a the number of heads when 3 coins are tossed
b the brand of toothpaste used by the students in a class
c the heights of a group of 15 year old children.

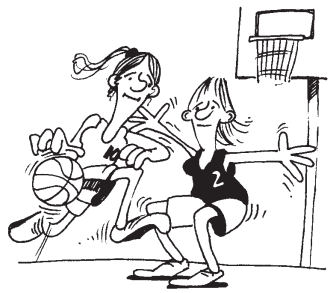
a The values of the variables are obtained by counting the number of heads. The result can only be one of the values 0, 1, 2 or 3. It is quantitative discrete data.

b The variable describes the brands of toothpaste. It is categorical data.

c This is numerical data obtained by measuring. The results can take any value between certain limits determined by the degree of accuracy of the measuring device. It is quantitative continuous data.

EXERCISE A

1 For each of the following possible investigations, classify the variable as categorical, quantitative discrete or quantitative continuous:



- **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 number of loaves of bread bought each week by a family
- **f** the pets owned by students in a year 8 class
- **g** the number of leaves on the stems of plants
- **h** the amount of sunshine in a day
- **i** the number of people who die from cancer each year in the USA
- **j** the amount of rainfall in each month of the year
- **k** the countries of origin of immigrants
- **l** the most popular colours of cars
- **m** the gender of school principals
- **n** the time spent doing homework
- **o** the marks scored in a class test
- **p** the items sold at the school canteen
- **q** the number of matches in a box
- **r** the reasons people use taxis
- **s** the sports played by students in high schools
- **t** the stopping distances of cars doing 60 km/h
- **u** the pulse rates of a group of athletes at rest.

2 **a** For the categorical variables in question **1**, write down two or three possible categories. (In all cases but one, there will be more than three categories possible.) Discuss your answers.

b For each of the quantitative variables (discrete and continuous) identified in question **1**, discuss as a class the range of possible values you would expect.

INVESTIGATION STATISTICS FROM THE INTERNET

In this investigation you will be exploring the web sites of a number of organisations to find out the topics and the types of data that they collect and analyse.

Note that the web addresses given here were operative at the time of writing but there is a chance that they will have changed in the meantime. If the address does not work, try using a search engine to find the site of the organisation.

What to do:

Visit the site of a world organisation such as the United Nations (www.un.org) or the World Health Organisation (www.who.int) and see the available types of data and statistics.

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* 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 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 2

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 in question **1** on page **6**.

2 Give three examples of data which would be collected by using a:
a census **b** sample

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 3

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.

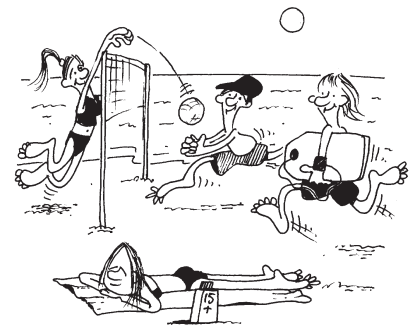
c The sample would be biased towards people who attend football matches. For example, there would probably be more males than females at football matches.

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

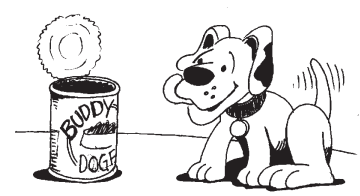


1 Explain and discuss any possible bias in the following samples:

- **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."



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.

C PRESENTING AND INTERPRETING DATA

ORGANISING CATEGORICAL DATA

A **tally and frequency table** can be used to organise categorical data.

For example, a survey was conducted on the type of fuel used by 50 randomly selected vehicles.

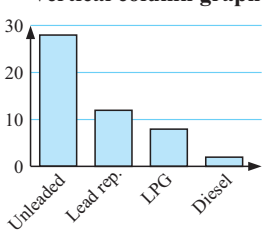
The variable 'type of fuel' is a categorical variable because the information collected for each vehicle can only be one of the four categories: Unleaded, Lead Replacement, LPG or Diesel. The data has been tallied and organised in the given frequency table:

Fuel type	*Tally*	*Freq.*
Unleaded	𝍸 𝍸 𝍸 𝍸 𝍸 |||	28
Lead Rep	𝍸 𝍸 ||	12
LPG	𝍸 |||	8
Diesel	||	2
	Total	50

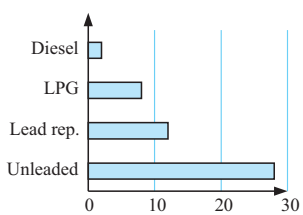
DISPLAYING CATEGORICAL DATA

Acceptable graphs to display the 'type of fuel' categorical data are:

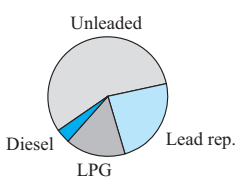
Vertical column graph



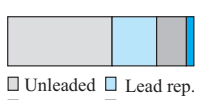
Horizontal bar chart



Pie chart



Segment bar chart



For categorical data, the **mode** is the category which occurs most frequently.

ORGANISING DISCRETE NUMERICAL DATA

Discrete numerical data can be organised:

- in a **tally and frequency table**
- using a **dot plot**
- using a **stem-and-leaf plot** (also called a **stemplot**).

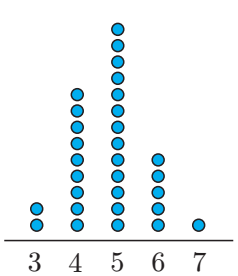
Stemplots are used when there are many possible data values. The stemplot is a form of grouping of the data which displays frequencies but retains the actual data values.

Examples:

- **frequency table**

Number	*Tally*	*Freq.*
3	||	2
4	𝍸 ||||	9
5	𝍸 𝍸 |||	13
6	𝍸	5
7	|	1

- **dot plot**



- **stemplot**

Example:

Stem	*Leaf*
0	9
1	7 1
2	8 3 6 7 6 4
3	9 3 5 5 6 8 2 1
4	7 9 3 4 2
5	1

As data is collected it can be entered directly into a carefully set up tally table, dot plot or stemplot blank sheet.

THE PEA PROBLEM

A farmer wishes to investigate the effect of a new organic fertiliser on his crops of peas. He is hoping to improve the crop yield by using the fertiliser. He set up a small garden which was subdivided into two equal plots and planted many peas. Both plots were treated the same except for the use of the fertiliser on one, but not the other. All other factors such as watering were as normal.

A random sample of 150 pods was harvested from each plot at the same time and the number of peas in each pod counted. The results were:

Without fertiliser

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

With fertiliser

6 7 7 4 9 5 5 5 8 9 8 9 7 7 5 8 7 6 6 7 9 7 7 7 8 9 3 7 4 8 5 10 8 6 7 6 7 5 6 8
7 9 4 4 9 6 8 5 8 7 7 4 7 8 10 6 10 7 7 7 9 7 7 8 6 8 6 8 7 4 8 6 8 7 3 8 7 6 9 7
6 9 7 6 8 3 9 5 7 6 8 7 9 7 8 4 8 7 7 7 6 6 8 6 3 8 5 8 7 6 7 4 9 6 6 6 8 4 7 8
9 7 7 4 7 5 7 4 7 6 4 6 7 7 6 7 8 7 6 6 7 8 6 7 10 5 13 4 7 7

For you to consider:

- Can you state clearly the problem that the farmer wants to solve?
- How has the farmer tried to make a fair comparison?
- How could the farmer make sure that his selection is at random?
- What is the best way of organising this data?
- What are suitable methods of display?
- Are there any abnormally high or low results and how should they be treated?
- How can we best indicate the most typical pod size?
- How can we best indicate the spread of possible pod sizes?
- What is the best way to show 'typical pod size' and the spread?
- Can a satisfactory conclusion be made?

ORGANISATION AND DISPLAY OF DISCRETE DATA

In **The Pea Problem**, the **discrete quantitative variable is**: *The number of peas in a pod.*

To organise the data a tally/frequency table could be used.

We count the data systematically and use a '|' to indicate each data value.

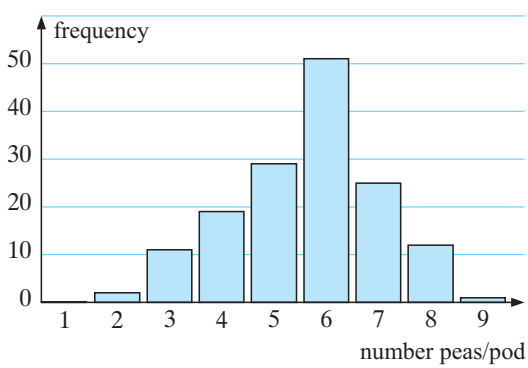
Remember that 𝍸 represents 5.

Below is the table for *Without fertiliser*:

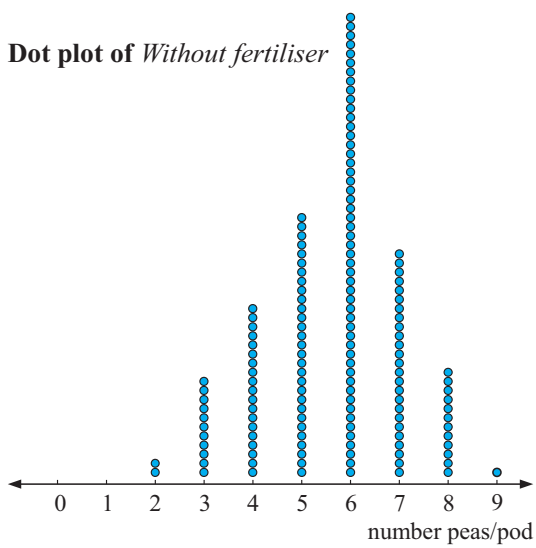
Number of peas/pod	*Tally*	*Frequency*
1		0
2	||	2
3	𝍸 𝍸 |	11
4	𝍸 𝍸 𝍸 ||||	19
5	𝍸 𝍸 𝍸 𝍸 𝍸 ||||	29
6	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 |	51
7	𝍸 𝍸 𝍸 𝍸 𝍸	25
8	𝍸 𝍸 ||	12
9	|	1

A **dot plot** could be used to organise and display the results, or a **column graph** could be used to display the results.

Column graph of *Without fertiliser*



Dot plot of *Without fertiliser*



DISCUSSION

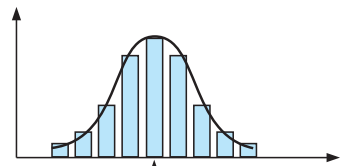
Are there any advantages/disadvantages in using a dot plot rather than a column graph?

From both graphs we can make observations and calculations such as:

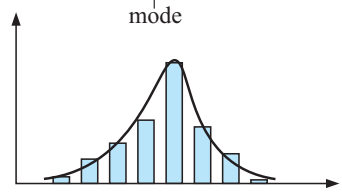
- 6 peas per pod is the mode of the *Without fertiliser* data.
- 8.7% of the pods had fewer than 4 peas in them, etc.

DESCRIBING THE DISTRIBUTION OF THE DATA SET

Many data sets show **symmetry** or **partial symmetry** about the mode.

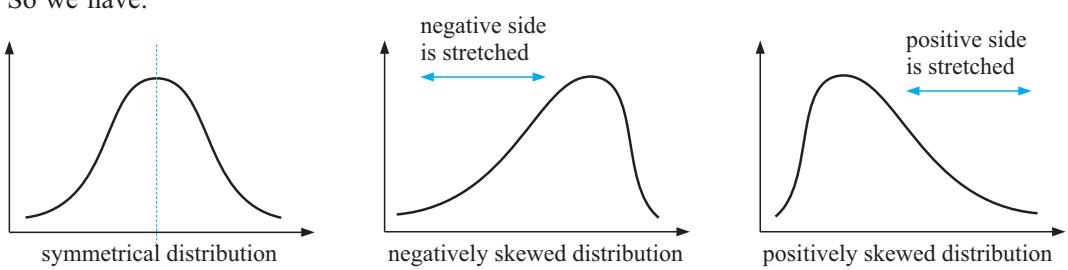


If we place a curve over the column graph (or dot plot) we see that this curve shows symmetry and we say that we have a **symmetrical distribution** of the data.



For the *Without fertiliser* data we have: This distribution is said to be **negatively skewed** as if we compare it with the symmetrical distribution it has been 'stretched' on the left (or negative) side of the mode.

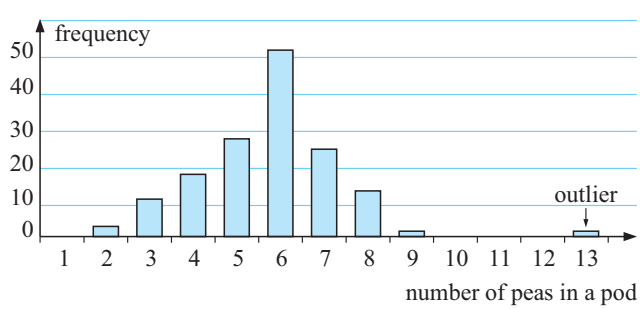
So we have:



OUTLIERS

Outliers are data values that are either much larger or much smaller than the general body of data. Outliers appear separated from the body of data on a frequency graph.

For example, if the farmer in **The Pea Problem** (page **11**) found one pod in the *Without fertiliser* sample contained 13 peas, then the data value 13 would be considered an outlier. It is much larger than the other data in the sample. On the column graph it would appear separated.



EXERCISE C.1

1 State whether the following quantitative (or numerical) variables are discrete or continuous:

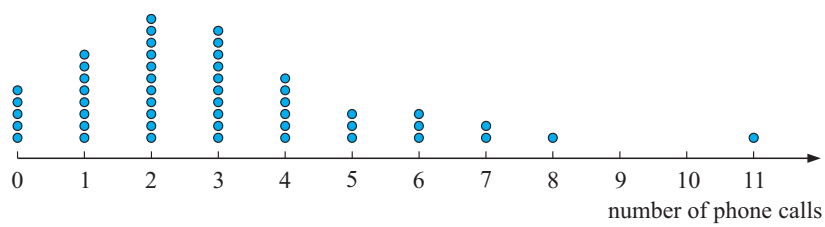
- **a** the time taken to run 100 metres
- **b** the maximum temperature reached on a January day
- **c** the number of matches in a box
- **d** the weight of luggage taken on an aircraft
- **e** the time taken for a battery to run down
- **f** the number of bricks needed to build a house
- **g** the number of passengers on a bus
- **h** the number of minutes spent on the internet per day

2 A class of 20 students was asked "How many pets do you have in your household?" and the following data was collected:

0 1 2 2 1 3 4 3 1 2 0 0 1 0 2 1 0 1 0 1

- **a** What is the variable in this investigation?
- **b** Is the data discrete or continuous? Why?
- **c** Construct a dotplot to display the data. Use a heading for the graph, and scale and label the axes.
- **d** How would you describe the distribution of the data? (Is it symmetrical, positively skewed or negatively skewed? Are there any outliers?)
- **e** What percentage of the households had no pets?
- **f** What percentage of the households had three or more pets?

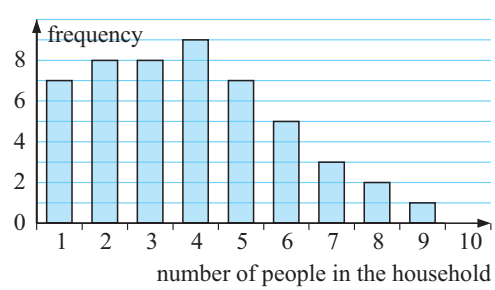
3 For an investigation into the number of phone calls made by teenagers, a sample of 50 fifteen-year-olds were asked the question "How many phone calls did you make yesterday?" The following dotplot was constructed from the data:



- **a** What is the variable in this investigation?
- **b** Explain why the data is discrete numerical data.
- **c** What percentage of the fifteen-year-olds did not make any phone calls?
- **d** What percentage of the fifteen-year-olds made 5 or more phone calls?
- **e** Copy and complete:
  "The most frequent number of phone calls made was ....."
- **f** Describe the distribution of the data.
- **g** How would you describe the data value '11'?

4 A randomly selected sample of households has been asked, 'How many people live in your household?' A column graph has been constructed for the results.

- **a** How many households gave data in the survey?
- **b** How many of the households had only one or two occupants?
- **c** What percentage of the households had five or more occupants?
- **d** Describe the distribution of the data.



5 The number of matches in a box is stated as 50 but the actual number of matches has been found to vary. To investigate this, the number of matches in a box has been counted for a sample of 60 boxes:

51 50 50 51 52 49 50 48 51 50 47 50 52 48 50 49 51 50 50 52
52 51 50 50 52 50 53 48 50 51 50 50 49 48 51 49 52 50 49 50
50 52 50 51 49 52 52 50 49 50 49 51 50 50 51 50 53 48 49 49

a What is the variable in this investigation?
b Is the data continuous or discrete numerical data?
c Construct a frequency table for this data.
d Display the data using a bar chart.
e Describe the distribution of the data.
f What percentage of the boxes contained exactly 50 matches?

6 **Revisiting The Pea Problem**

For the *With fertiliser* data on page **11**:

a Organise the data in a tally-frequency table.
b Draw a column graph of the data.
c Are there any outliers?
d Is the data skewed?
e What evidence is there that the fertiliser 'increases the number of peas in a pod'?
f Can it be said that the fertiliser will increase the farmer's pea crop and therefore his profits?

GROUPED DISCRETE DATA

It is not sensible to organise some discrete data by using a frequency table. Also graphing by dot plot or column graph is not appropriate.

For example, a local kindergarten is concerned about the number of vehicles passing by between 8.45 am and 9.00 am.

Over 30 consecutive week days they recorded data.

The results were: 27, 30, 17, 13, 46, 23, 40, 28, 38, 24, 23, 22, 18, 29, 16,
35, 24, 18, 24, 44, 32, 52, 31, 39, 32, 9, 41, 38, 24, 32

In situations like this grouping the data into **class intervals** is appropriate.

It seems sensible to use class intervals of length 10 in this case.

The tally/frequency table is:

Number of cars	*Tally*	*Frequency*
0 to 9	|	1
10 to 19	卌	5
20 to 29	卌 卌	10
30 to 39	卌 ||||	9
40 to 49	||||	4
50 to 59	|	1
	Total	30

EXERCISE C.2

1 The data set below is the test scores (out of 100) for a Maths test for 50 students.

56	29	78	67	68	69	80	89	92	71
58	66	56	88	81	70	73	63	74	38
67	64	62	55	56	75	90	92	47	44
59	64	89	62	51	87	89	76	59	88
72	80	95	68	80	64	53	43	61	39

a Construct a tally and frequency table for this data using class intervals 0 - 9, 10 - 19, 20 - 29, ..., 90 - 100.

b What percentage of the students scored 80 or more for the test?

c What percentage of students scored less than 50 for the test?

d Copy and complete the following:
More students had a test score in the interval ..... than in any other interval.

2 The number of points scored by the eight winning teams in the first five rounds of the Year 2001 AFL season are to be investigated. The variable 'the number of points scored by a winning team' is a discrete numerical variable that can theoretically take the values 1, 2, 3, ..., 101, 102, 103, ..., 201, ... etc.

The winning scores are:

94	156	154	131	129	134	152	140	124	162
103	139	82	170	110	111	116	160	104	110
98	106	187	149	165	88	118	123	137	128
113	130	145	139	125	154	126	141	122	106

a Construct a frequency table for this data using class intervals 80-89, 90-99, 100-109, ..., 180-189.

b Which class interval has the most number of winning scores?

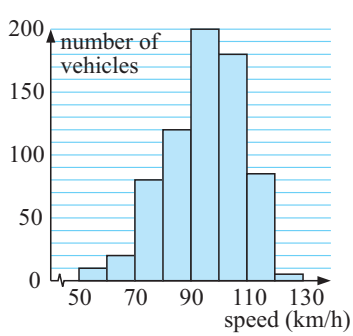
c How many matches were won with a winning score of 99 points or less?

d What percentage of the matches had winning scores of 150 points or more?

e Describe the distribution of the data.

f Are there any outliers in this data set?

3 The speed of vehicles travelling along a section of highway has been recorded and displayed using the histogram shown:



a How many vehicles were included in this survey?

b What percentage of the vehicles were travelling at a speed from 100 up to 110 km/h?

c What percentage of the vehicles were travelling at speeds equal to or greater than 100 km/h?

d What percentage of the vehicles were travelling at a speed less than 80 km/h?

e If the owners of the vehicles travelling at 110 km/h or more were fined $165 each, what amount would be collected in fines?

EXERCISE A

1 **a** quantitative discrete **b** quantitative continuous
c categorical **d** quantitative discrete
e quantitative discrete **f** categorical
g quantitative discrete **h** quantitative continuous
i quantitative discrete **j** quantitative continuous
k categorical **l** categorical **m** categorical
n quantitative continuous **o** quantitative discrete
p categorical **q** quantitative discrete **r** categorical
s categorical **t** quantitative continuous
u quantitative discrete

2 Answers will vary

EXERCISE B.1

1 **a** census **b** census **c** sample **d** sample or census
e sample **f** census **g** sample **h** census **i** census
j census **k** census **l** sample **m** sample **n** sample
o census **p** census **q** sample **r** sample **s** sample
t sample **u** census

2 **a** age structure of Australia's population,
marks in a class test,
the uniform preferences of students at a high school
b favourite television program,
favourite football team,
favourite model of motor car

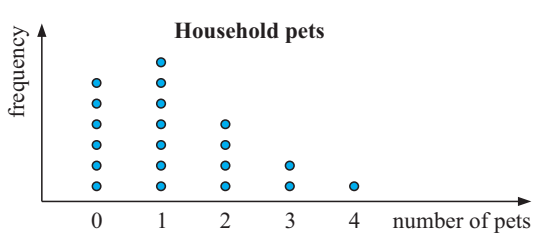
EXERCISE B.2

This exercise is for discussion. Answers will vary. None are given.

EXERCISE C.1

1 **a** continuous **b** continuous **c** discrete **d** continuous
e continuous **f** discrete **g** discrete **h** continuous

2 **a** number of pets
b Discrete since you can't have part of a pet.
c



d positively skewed, no outliers **e** 30% **f** 15%

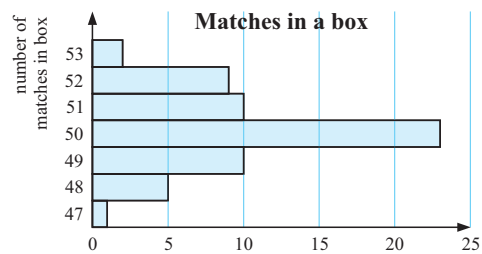
3 **a** the number of phone calls made in a day
b You can only make whole phone calls. **c** 10%
d 20% **e** ..... two calls per day
f positively skewed with an outlier
g Data value 11 is an outlier.

4 **a** 50 households **b** 15 households **c** 36%
d positively skewed, no outliers

5 **a** number of matches in a box **b** discrete
c

No. matches	*Tally*	*Freq.*
47	|	1
48	卌	5
49	卌 卌	10
50	卌 卌 卌 卌 |||	23
51	卌 卌	10
52	卌 ||||	9
53	||	2

d

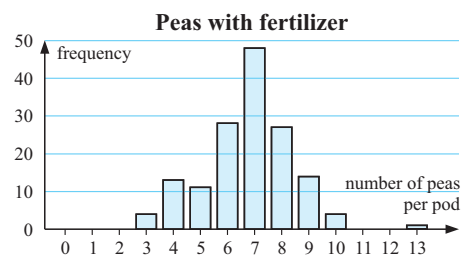


e approximately symmetrical **f** 38.3%

6 **a**

No. peas	*Tally*	*Freq.*
3	||||	4
4	卌 卌 |||	13
5	卌 卌 |	11
6	卌 卌 卌 卌 卌 |||	28
7	卌 卌 卌 卌 卌 卌 卌 卌 卌 |||	48
8	卌 卌 卌 卌 卌 ||	27
9	卌 卌 ||||	14
10	||||	4
11		0
12		0
13	|	1
	Total	150

b



c Yes, data value 13 is an outlier. **d** negatively skewed
e On average the number of peas is higher in the "with fertiliser" group. The mode has increased from 6 to 7.
f Yes, assuming the fertiliser is not too expensive and the peas are as big as they were previously.

EXERCISE C.2

1 **a**

Test Score	*Tally*	*Freq.*
0 - 9		0
10 - 19		0
20 - 29	|	1
30 - 39	||	2
40 - 49	|||	3
50 - 59	卌 ||||	9
60 - 69	卌 卌 |||	13
70 - 79	卌 |||	8
80 - 89	卌 卌	10
90 - 100	||||	4
	Total	50

b 28% **c** 12%
d More students had a test score in the interval 60 - 69 than in any other interval.

2 **a**

Winning Score	*Frequency*
80 - 89	2
90 - 99	2
100 - 109	4
110 - 119	6
120 - 129	7
130 - 139	6
140 - 149	4
150 - 159	4
160 - 169	3
170 - 179	1
180 - 189	1

b 120 - 129
c 4
d 22.5%
e approximately symmetric distribution
f yes, data value 187 is an outlier

3 **a** 700 **b** $\doteqdot 25.7\%$ **c** $\doteqdot 38.6\%$ **d** $\doteqdot 15.7\%$
e \$14 850