PART A

Chapter 5

Descriptive statistics

Contents:

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.

A DESCRIBING DATA

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.

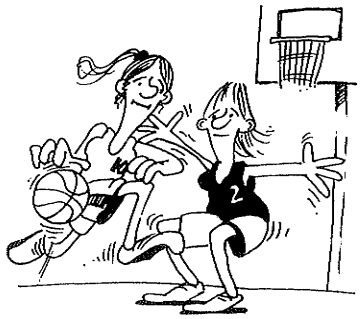
b The variable describes the brands of toothpaste. It is categorical data.

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 5A

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 basketball ball 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 1 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 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.

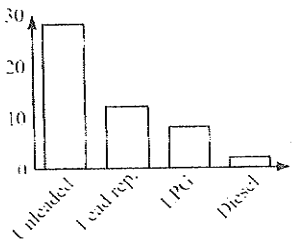
The variable 'type of fuel' is a categorical variable because the information collected for each vehicle can only be one of the four categories: Unleaded, Lead Replacement, LPG or Diesel. The data has been tallied and organised in the given frequency table:

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

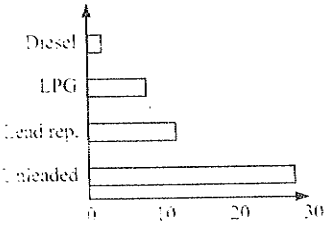
DISPLAYING CATEGORICAL DATA

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

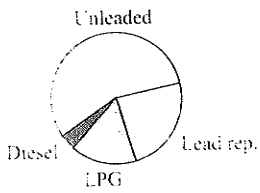
Vertical column graph



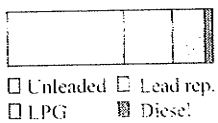
Horizontal bar chart



Pie chart



Segment bar chart



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