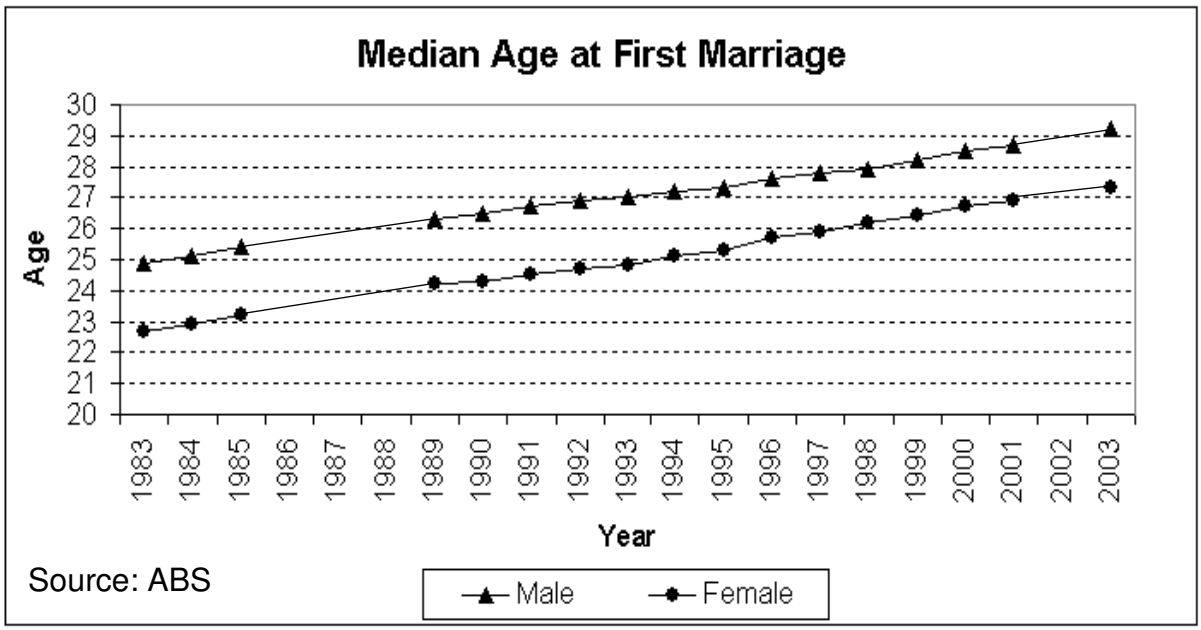
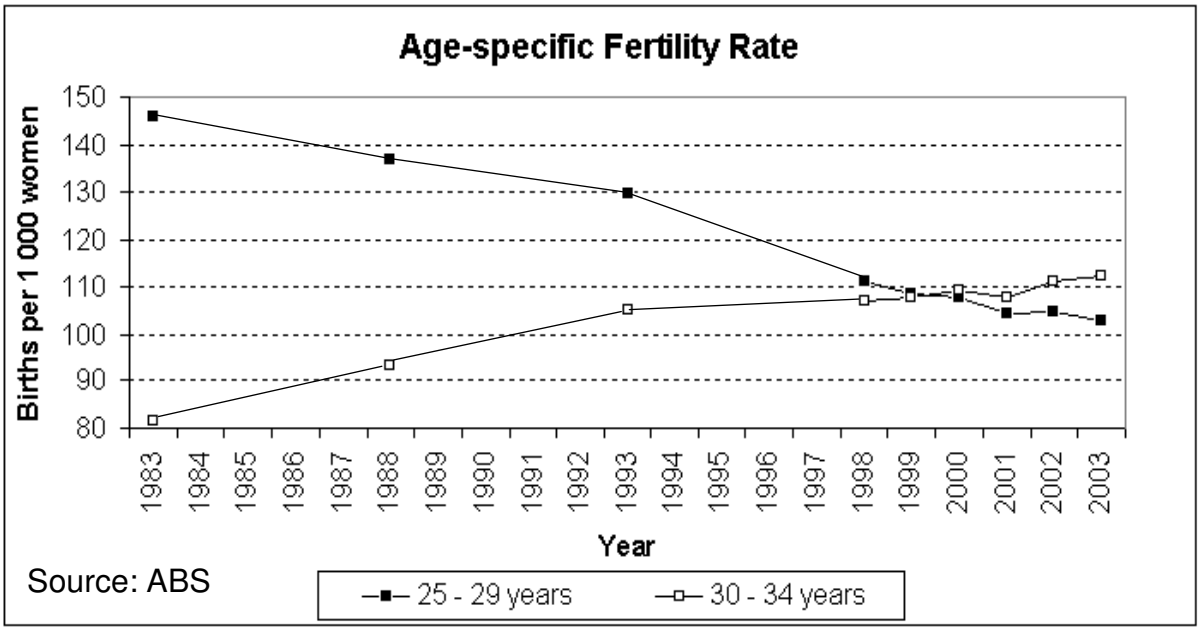
Line graphs of related variables

When two or more variables are graphed on the same pair of axes, you can see how one of them varies in relation to the others.

Look at the graph below.
As the median age of females when they first married increased, what happened to the median age of males when they first married?



Look at the graph below.
As the fertility of females aged 25 – 29 years decreased, what happened to the fertility of females aged 30 – 34 years?



ISBN 0-9751382-4-3

If there is a relationship between two variables, the variables are said to be **correlated.** "Co" means "together", so "correlated" means "related together".

If one variable tends to have a low value when the other variable has a low value, and a high value when the other variable has a high value (as in the top graph on the previous page), the variables are said to be **positively correlated**.

If the opposite tends to occur ie. low values of one variable with high values of the other variable (as in the bottom graph on the previous page), then the two variables are said to be **negatively correlated**.

There are 2 dependent variables in each graph on the previous page.

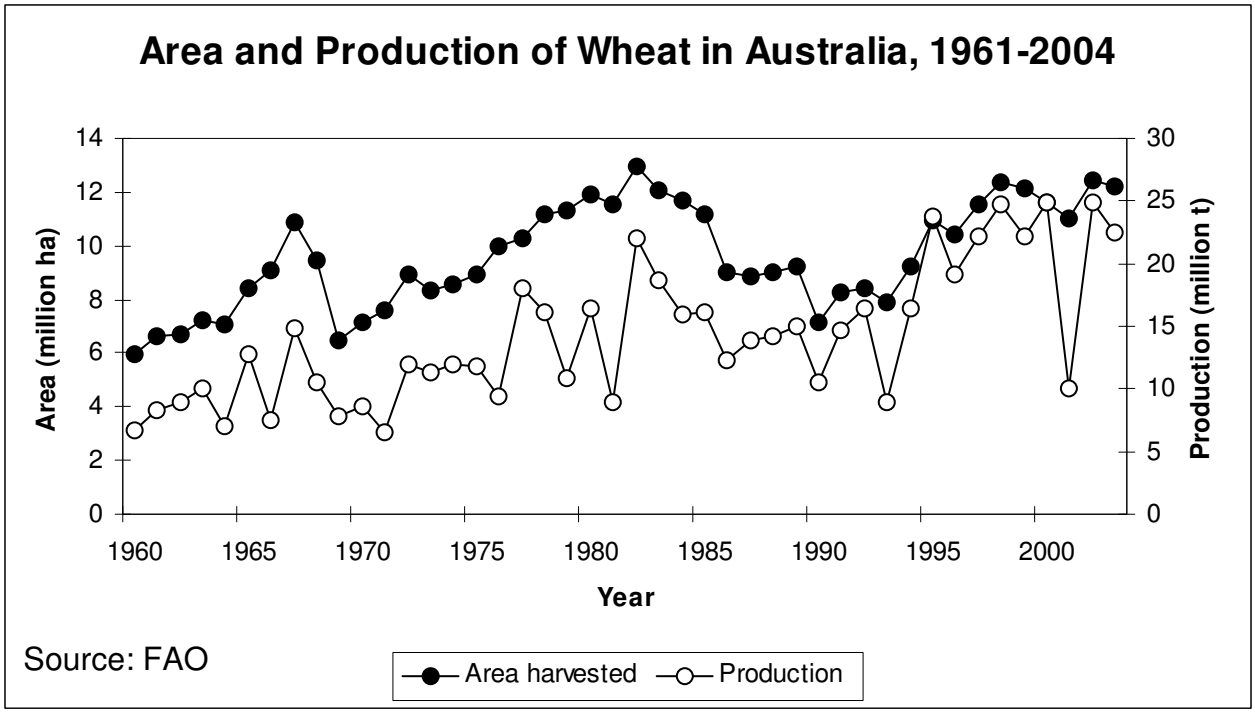
- Which of these pairs of variables are positively corrleated? .....

.....

- Which of these pairs of variables are negatively corrleated? .....

.....

The graph below shows the area of wheat harvested and the production of wheat in Australia between 1961 and 2004.



Source: FAO

Is the area of wheat positively or negatively correlated to the amount of wheat produced? .....

How do you know? .....

.....

.....

ISBN 0-9751382-4-3

Scatter diagrams

Scatter diagrams are used to see whether there is a relationship between 2 quantitative variables, and if so, what the relationship is like.

Plotting data on a scatter diagram is like plotting points on a number plane. The independent variable is usually placed along the horizontal axis and the dependent variable on the vertical axis.

In the table below are 10 years (1995 – 2004) of the wheat area and production data that were graphed as a line graph in Activity 6 - 2.

Year	Area harvested (million ha)	Production (million t)
1995	9.2	16.5
1996	10.9	23.7
1997	10.4	19.2
1998	11.5	22.1
1999	12.3	24.7
2000	12.1	22.1
2001	11.6	24.8
2002	11.0	10.6
2003	12.4	24.9
2004	12.2	22.5

These data are plotted in a scatter diagram on the next page.
In this diagram:

- "year" is not treated as a variable. Instead, each year of data is treated as an individual case.
- "area harvested" is now the independent variable and "production" is the dependent variable.

If two variables are correlated, the symbols (eg. dots or crosses) on a scatter diagram will appear to form something close to a line.

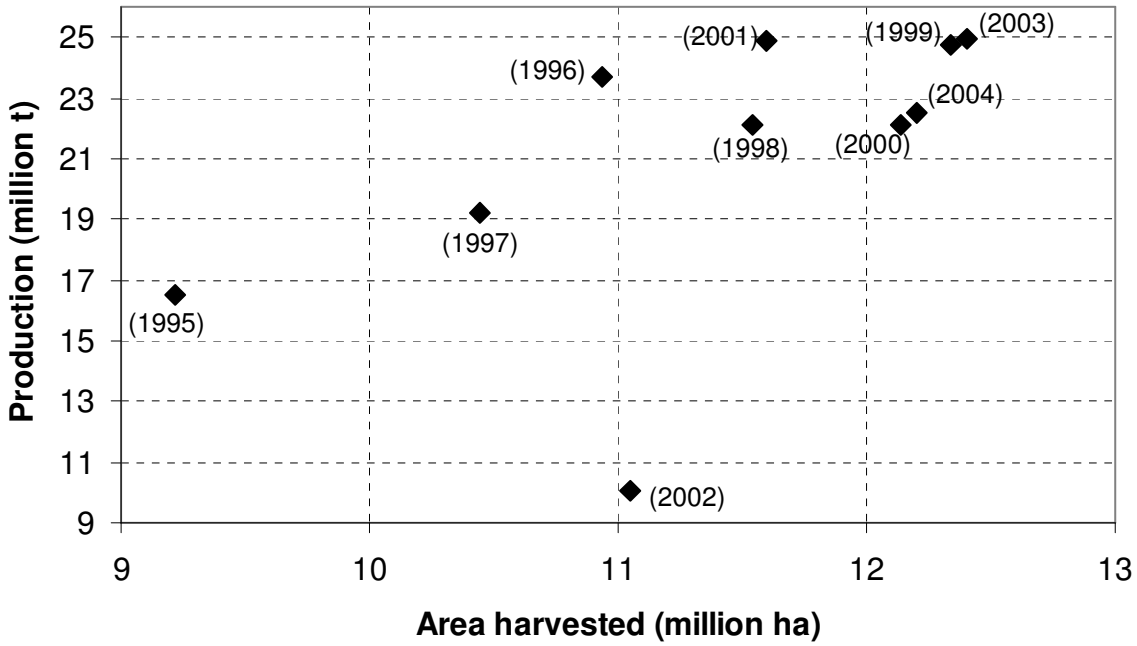
From the scatter diagram on the next page, would you say that "area harvested" and "production" are correlated? .....

For most years, would you expect "area harvested" and "production" to be correlated? .....
Why or why not? .....

.....

ISBN 0-9751382-4-3

Area of Wheat Harvested vs. Wheat Production in Australia



If the symbol representing the year 2002 were removed from the scatter diagram above, would you say that "area harvested" and "production" are correlated? .....

Because the data for 2002 does not fit the usual pattern, this item of data is called an **outlier**. Outlier means "lies outside."

Sometimes an outlier occurs because a mistake has been made in the collection or recording of the data. An outlier, however, can also occur for other reasons. Thinking about possible reasons can lead to further investigation and discovery.

What could have caused the data point for 2002 to be an outlier? .....

.....

.....

.....

In a scatter diagram:

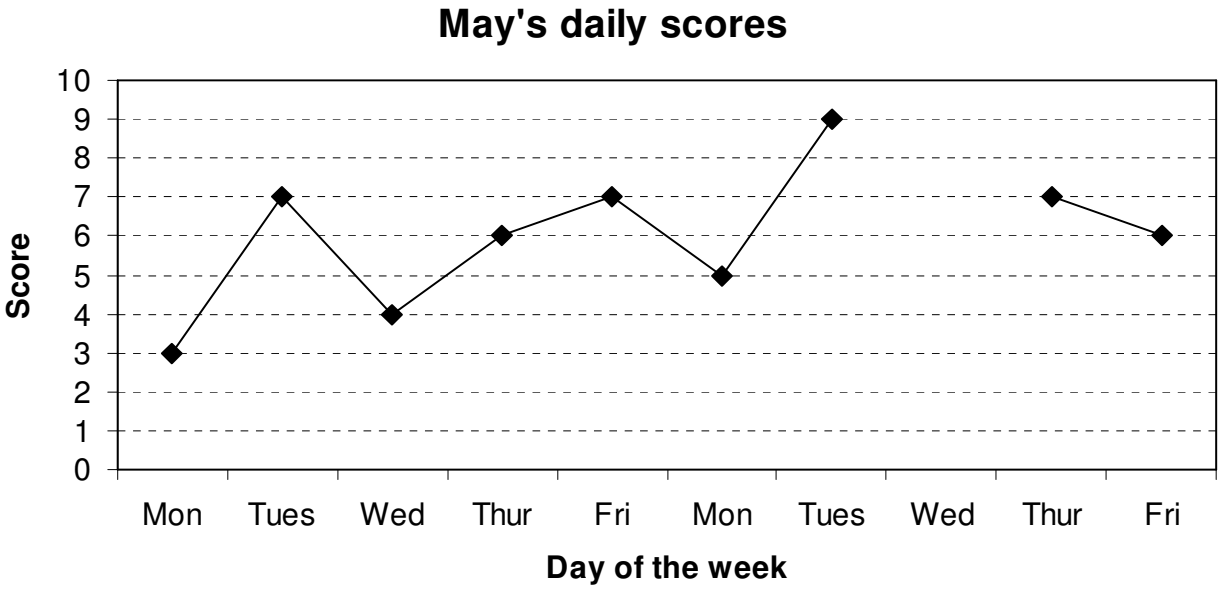
- each symbol (eg. a dot or cross) on the graph represents the value of the two variables for a particular case (eg. for a particular year or for a particular individual)
- lines are not drawn from point to point because the cases are in no particular order
- there may be more than one value of the dependent variable for each value of the independent variable.

ISBN 0-9751382-4-3

Errors in line graphs

A group of girls like playing a game of chance called “Scissors, Paper, Rock”. They hold a tournament at the back of the school bus.
Every morning the girls play 10 games each (not counting drawn games). After 2 weeks, they add up their daily scores to find out who is the tournament winner.

May was absent on the second Wednesday of the tournament.
Her daily scores are graphed below.



The winning score for the tournament was 60.

What was May’s total score for the tournament? .....
May was disappointed with the scoring system. Why? .....

.....

May drew a line on the graph above to fill in the gap between her Tuesday and Thursday scores. Draw this line on the graph.

From the graph, what did May claim that her score would have been if she had played on the Wednesday? .....

Is this a reasonable way of estimating her missing score? .....
Why or why not? .....

.....

Describe another way of estimating her missing score, then estimate it.

.....

.....

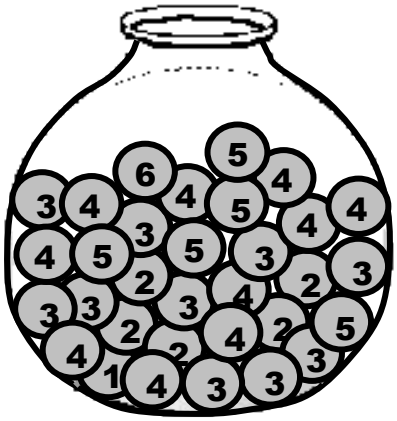
Do you think May should have won the tournament? .....

ISBN 0-9751382-5-1

Choosing an average

Resources required:
a calculator per student.

An average is a value that helps to describe a set of data. It tells you what you would expect an individual from the population to be like.

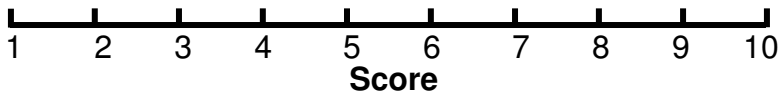


This bowl contains a population of 32 balls. Each ball has a score written on it.

The bowl is shaken up and one ball is randomly chosen from the population.

The average of the population (ie. the mean, median and/or mode) should give you a reasonable expectation of what this ball's score might be.

Make a dot plot of the population of balls using the number line below.



Describe the shape of this distribution (ie. whether it is even, symmetrical or skewed and whether it has one or more clusters or outliers).....

.....

Use your dot plot (above) and a calculator to find the mean, median and mode of the population.

Mean:..... Median:..... Mode:.....

Which of these three averages would you say is the most representative of the population? .....

A ball with a score of 33 is added to the bowl. Find the new averages.

Mean:..... Median:..... Mode:.....

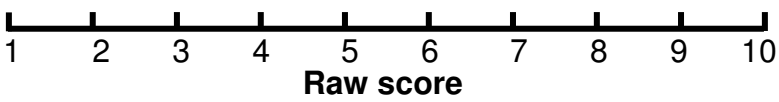
Which average is the most sensitive to an outlier?.....

ISBN 0-9751382-5-1

Norman played 18 holes of putt-putt golf. His scorecard is shown below.
His raw score is the actual number of strokes he needed.
His score was adjusted because in the rules, no more than 5 strokes could be awarded for any one hole.

Hole number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Raw score	3	3	2	4	1	5	6	2	3	7	3	6	8	2	4	3	9	5
Adjusted score	3	3	2	4	1	5	5	2	3	5	3	5	5	2	4	3	5	5

Make a dot plot of his raw scores.



Describe the shape of this distribution. .....

Find the mean, median and mode of Norman's raw scores.

Mean:..... Median:..... Mode:.....

Which average would you say is the most representative of Norman's scores?.....

Find the mean, median and mode of the Norman's adjusted scores.

Mean:..... Median:..... Mode:.....

From the distribution of his adjusted scores, which average do you think best indicates Norman's performance?.....
Why?.....

.....

With a partner, decide what averages you would use if you had:

1) a roughly symmetrical distribution?.....
2) a skewed distribution?.....
3) a distribution with more than one peak (cluster)?.....
4) a distribution with an outlier?.....
5) a distribution of ordered categorical data?.....

ISBN 0-9751382-5-1

Random sampling

Resources required:
a calculator per student.

Random numbers can be used to select a **random sample** from the population. You simply number the individuals and generate some random numbers in that range.
A random sample is unbiased. Why?.....

.....

In this activity, two random samples will be selected from a herd of goats, then used to estimate the number of goats with a coloured coat.

Sample 2 will be bigger than Sample 1. Which sample do you expect to represent the population better?.....

A herd of 100 goats is drawn on the next page. Each goat is identified by its ear tag number. These numbers (1 to 100) are written on the goats. Goats with coloured coats are shaded grey.

Sample 1. Generate 5 different random two-digit numbers and write them here:.....

Circle the goats that have these numbers on their tags.
(If you have the random number 00, this is goat number 100.)
How many goats in your sample have coloured coats?.....
The herd is how many times larger than your sample?.....
If you only had this sample, what would you estimate to be the number of coloured goats in the herd?.....

Sample 2. Generate 10 different random two-digit numbers and write them here:.....

Put a cross through the goats that have these numbers on their tags.
How many goats in your sample have coloured coats?.....
The herd is how many times larger than your sample?.....
If you only had this sample, what do you estimate to be the number of coloured goats in the herd?.....

Which of the two samples gave you the best estimate?.....

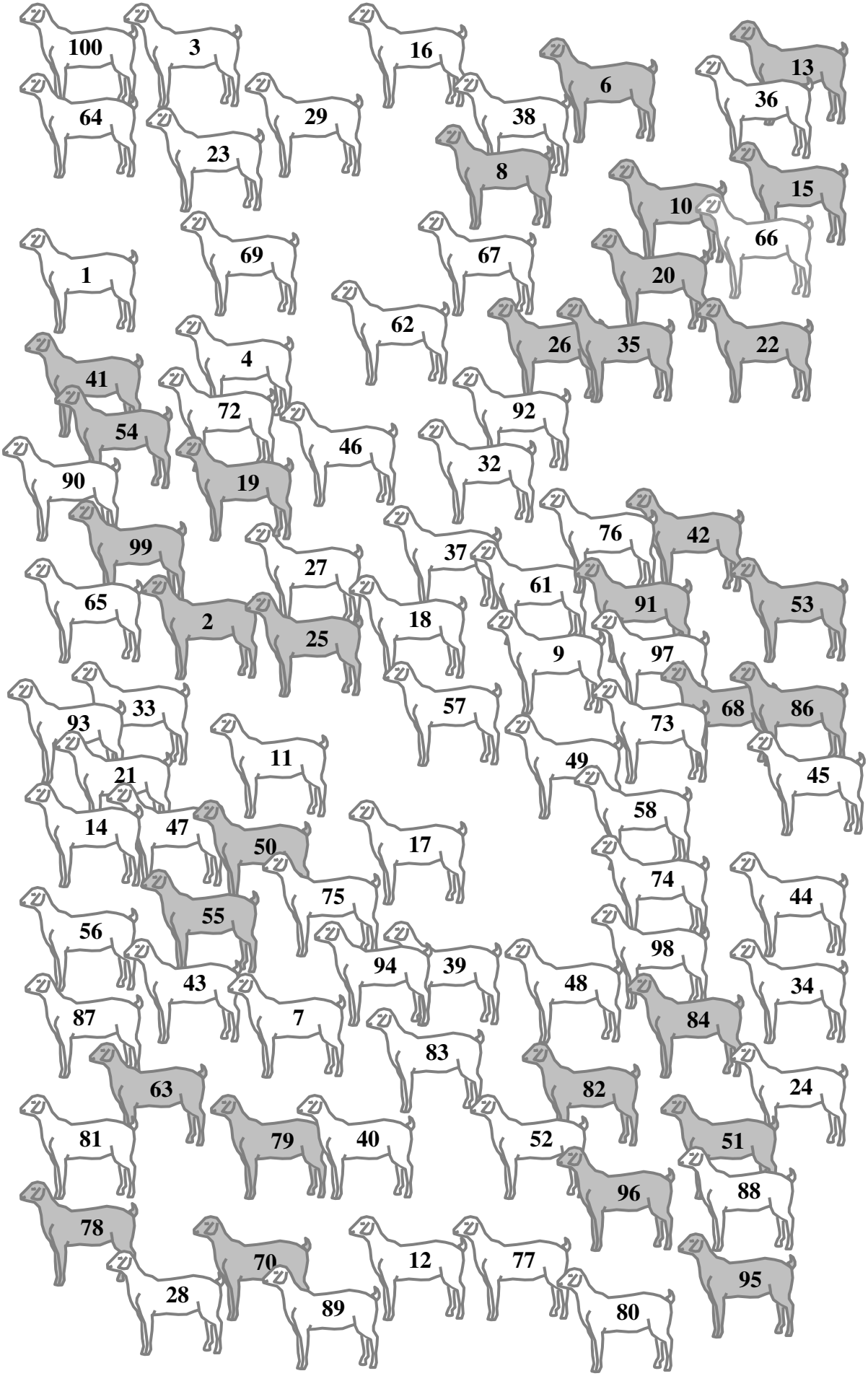
List the number of coloured goats that others found in their sample of 10:

.....

How could you use the other samples to get a better estimate of the number of coloured goats in the herd?.....

.....

ISBN 0-9751382-5-1



ISBN 0-9751382-5-1

Sample size and variability

Resources required:
a small paper plate per student
a small box of smarties per student
a calculator per student.
(Note: There are 14 small boxes in a 180g packet of boxed smarties.)

Georgia's soccer team has won the grand final.

The team decides to celebrate by baking a round cake and putting two rings of smarties around the edge.

The team colours are blue and yellow, so they would like to decorate the cake with yellow icing and blue smarties.

They need 100 smarties.

The girls buy smarties by walking to the corner shop.
This shop only sells smarties in small boxes.

The girls think about how much they are prepared to spend.
They have two questions:

- What is the least number of boxes they need to have a good chance of getting 100 smarties?
- What is the least number of boxes they need to have a good chance of getting 100 blue smarties?

Your whole class will carry out an investigation to help the girls answer these questions. In this investigation, what is the population that the girls are interested in? .....

The girls are interested in two variables:

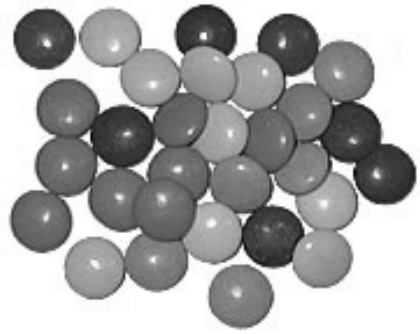
- the total number of smarties in a box
- the number of blue smarties in a box.

To find out about these variables in the population, you and your class will examine a sample from it.
Each person will look at the smarties inside one box.

Open your box of smarties and empty it into a plate.
Count the number of smarties of each colour.
Write the numbers in the table on the next page.

ISBN 0-9751382-5-1

Colour of Smartie	Frequency
Red	
Orange	
Yellow	
Green	
Blue	
Purple	
Pink	
Brown	
Total:	



If all the boxes of smarties were the same as yours,

- calculate how many boxes of smarties the girls would need to buy so that they have at least 100 smarties

- calculate how many boxes of smarties the girls would need to buy so that they have at least 100 blue smarties.

The total number of smarties in a box

Find out the total number of smarties that each person in the class found in their boxes. List these numbers here.

.....

.....

Did everyone find the same number of smarties in a box? .....

If the number of smarties in a box is not always the same, that number is said to vary. You can use the range to measure the variability (ie. the amount of variation).

In the sample that your class investigated, what was the range of the total number of smarties in a box?.....

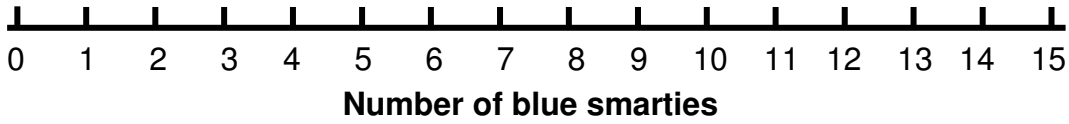
What was the mean number of smarties in a box?.....

About how many boxes of smarties do you recommend buying so there is a good chance of getting at least 100 smarties? .....

ISBN 0-9751382-5-1

The number of blue smarties in a box

Find out the number of blue smarties that each person in the class found in their boxes. Show them here as a dot plot.



What was the range of the number of blue smarties in a box? .....

What was the mean number of smarties in a box? .....

About how many boxes of smarties do you recommend buying so there is a good chance of getting at least 100 blue smarties? .....

If the sample size was larger (ie. if the class had sampled more boxes), would you be more confident in making this recommendation? .....
Why or why not? .....

.....

In the sample of boxes that the class investigated, which had the greater variation: the number of smarties in a box or the number of blue smarties in a box? .....

Which of your two recommendations (the number of boxes to get 100 smarties or the number of boxes to get 100 blue smarties) do you feel most confident about? .....
Why? .....

.....

.....

You can increase your confidence in the location of the mean by looking at a larger sample of boxes from the population.

Circle the correct words in the sentences below.

If the variable you are interested in has a large amount of variability (ie. if there are large differences between individuals) a larger/smaller sample needs to be taken to compensate for this. If it has little variability, a larger/smaller sample can be taken.

ISBN 0-9751382-5-1