Surveying and sampling

Monday, 11 February 2019 8:13 pm

By the end of the lesson I would hope that you have the knowledge and understanding for the following points:

- Understand the language which is used with Surveying and Sampling
  - Population
  - Sample
  - Survey
  - Census
  - Sample size
  - Biased
  - Outliers
  - Symmetric, Skewed, bi-modal
  - Proportion
- Know how to use the above terms to describe some data.

RECAP:

Much of what we do, in real life, is based on finding information (data) and doing something with it.
Computers help us make sense of the data, but we generally have to spend some time gathering the data.
We can do this in lots of ways, but today we're going to look at surveys and samples.

There is a lot of language used in Mathematics and, as such, we need to make sure we completely understand what each of the terms mean.

Barry is at it again!

Let's look at some of the language first!

Population: All the members of a group which is going to be studied. Example: All Year 12 students in Victoria

Sample: A small group which we will select from the population. Example: Year 12 students in Melbourne.

Survey: A means of collecting information from a sample.

Census: A means of collecting information from a population. Example: The government conducts a census every 5 years.

Sample size: The number of people who have been selected from the population

Biased: When a question or answer is phrased in a way where a particular answer might be more popular. Example: "I think fox hunting is a horrible sport. Do you agree?"

Outliers: Data points which (obviously) fit outside the data we were expecting. Example: When someone gives their age as -6.

Proportion: A fraction. Example: 10 out of 100 people is my sample size. Hence the proportion is $\frac{10}{100} = \frac{1}{10}$

$\frac{10}{100} = \frac{1}{10}$

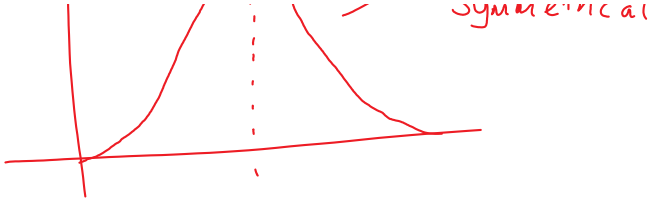
There is something else we need to know about ...

In Year 12 there is an awesome subject called "Further Mathematics".
Part of the course looks at how we can describe the shape of data.

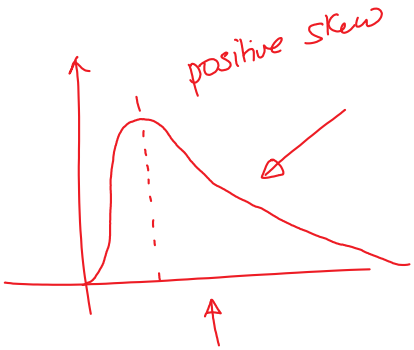
When we draw bar charts, we can see that the data might show different patterns.

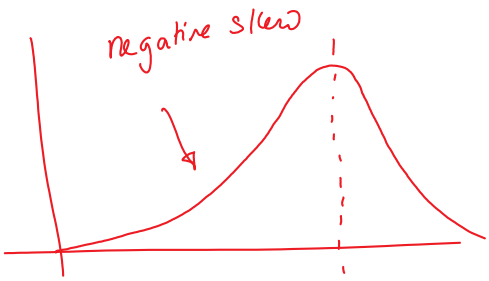
We can describe some of these patterns in the following way:

Symmetric
This is when the data (or graph) can be split into two and looks (roughly) the same either side of the central point.

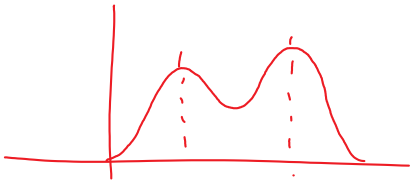


Skewed (Positively and Negatively)
This is when the graph is not evenly distributed either side of a central point.
We have can positively skewed and negatively skewed data.





Bi-Modal
Remember: We have met the MODE before. This is the most common vale.
When some data has two modes, we call the data bi-modal.
We can see how this looks on a diagram.



Collecting Data

There are lots of ways to collect data.
Standing outside of a supermarket and asking people might be one of them.
Other ways are posting surveys.
Doing online surveys.

How would I find out the average word length in a book?

Working out the Mean, Median and Mode of data is only part of the story.

Let's look at an example of data and how we might be asked to read it.
Example extracted from the *Cambridge Essentials Textbook Series*.

A survey is conducted asking 100 randomly selected adults how many children they have. The results are shown in this histogram.

a Assume that this sample is representative of the population.

i What proportion of the adult population has two or more children?

ii In a group of 9000 adults, how many would you expect to have 4 children?

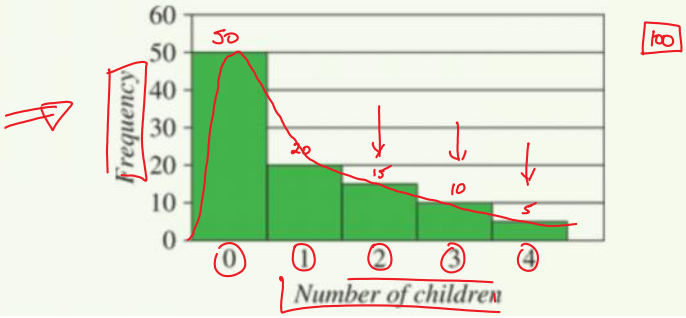
b Is this distribution symmetric, skewed or bi-modal?

c Which of the following methods of conducting the survey could lead to bias?

Method 1 Asking people waiting outside a childcare centre

Method 2 Randomly selecting people at a night club

Method 3 Choosing 100 adults at random from the national census and noting how many children they claimed to have



Accuracy of a survey

We need to be careful when doing surveys that the results are as accurate as possible and not affected by **bias**.

For example: If I wanted to find out if all Libraries should be closed, I wouldn't conduct a survey outside of a library as, the people coming out of a library will be the people who wish it to stay open. Hence, most people will say "No" which is biasing my results.

Can you think of a better place to ask the question?

If I wanted to know if the local town needed a Skate Park, would I stand outside of Coles at 11:00 a.m. on a school day?

Are there any outliers which might affect my final answer? What do I do with them?