



Classifying and displaying categorical data

Year 11 General Maths
Units 1 and 2

www.maffsguru.com

Learning Objectives

By the end of the lesson I would hope that you have an understanding and be able to apply to questions the following concepts:

- Understand that data can be classified as categorical or numerical
- Be able to classify categorical data as nominal or ordinal
- Be able to classify numerical data as discrete or continuous
- Be able to construct frequency and percentage frequency tables for categorical data
- Be able to construct bar charts and percentage bar charts from frequency tables
- Be able to interpret and describe frequency tables and bar charts



Recap

This is the first lesson in this section of the course and it's going to prove to be a very important one. When you move to General Maths in Year 12, this section of work is critical and part of a 9 chapter section of work! So, mastery now will make sure that you are well prepared for next year.

Remember: A good summary book is critical to your success

Data is all around us so let's see what we can do to recognise it and analyse it!



Categorical and Numerical data

There is a lot of data around us. People are paid to help us make sense of all the data. The human brain loves data to be in picture format but, before we do that, we need to know which picture is the best.

If we look at the following table, we can see four columns of data.

Two columns seem to show worded data, and the other two columns number data.

<i>Gender</i>	<i>Coffee size</i>	<i>Number of brothers</i>	<i>Hand span (in cm)</i>
M	Large	0	23.6
F	Small	2	19.6
F	Small	1	20.2
M	Large	1	24.0



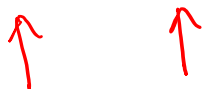
Categorical data

Don't get tricked into thinking that only worded data is categorical. There are lots of examples where numbers are used but are actually categories. This is used to trick people in their VCE exam!

Categorical data is where data has been used to place people into one of several categories.

Looking at the following the first two columns are categorical data.

<i>Gender</i>	<i>Coffee size</i>	<i>Number of brothers</i>	<i>Hand span (in cm)</i>
M	Large	0	23.6
F	Small	2	19.6
F	Small	1	20.2
M	Large	1	24.0



Categorical and nominal data

If we look at the first column we can see that we have used M and F to separate people into categories.

Can we order a gender? Are females more important than males? Are males more important than females? There is **no way to order gender** and so we describe it as **NOMIAL data**.

Nominal

Notice that the first two letter are **NO** meaning NO ORDER.

Gender	Coffee size	Number of brothers	Hand span (in cm)
M	Large	0	23.6
F	Small	2	19.6
F	Small	1	20.2
M	Large	1	24.0



Categorical and ordinal data

If we now look at the second column, we have the categories Small and Large.

Can we order these?

Of course we can. There is an implied order. We have structured the English language to be able to order things like Small, Medium, Large.

Hence, this is an example of Categorical ORDINAL data.

Again, if you look at the first three letters ORD it means there is an ORDER.

ordinal

Gender	Coffee size	Number of brothers	Hand span (in cm)
M	Large	0	23.6
F	Small	2	19.6
F	Small	1	20.2
M	Large	1	24.0



Other examples of Categorical data

I'm not going to get too deep into this at the moment, but the following are examples of categorical data:

- Colours
- Political parties
- Postcodes
- Years
- Running numbers on jerseys (in a race for example)
- Telephone numbers
- Shoe sizes

To be able to decide if something is categorical you can ask yourself:

"Can I work out the mean, median, range or mode of the data and, if I can, will the number make sense."

So, for example, can you have an "average phone number" or an "average postcode"

Numerically you can put the data into your calculator, but the number you get out won't make any sense.



Numerical data

Numerical data comes from something which can be counted or measured.

It also is data which, when you take either the mean, median, mode or range, the output makes sense.

If we look at the following table, we can see that the data in the last two columns can be categorised as numerical.

Again, there are two different ways to describe numerical data.

<i>Gender</i>	<i>Coffee size</i>	<i>Number of brothers</i>	<i>Hand span (in cm)</i>
M	Large	0	23.6
F	Small	2	19.6
F	Small	1	20.2
M	Large	1	24.0



Numerical and Discrete data

When data is said to be discrete it means it can not (generally) be rounded to a number of decimal places. It is generally numerical data which is whole numbers. Again, there are probably exceptions.

Looking at the following table, we can see that the column “Number of brothers” would be **numerical and discrete data**.

discrete
↓

Gender	Coffee size	Number of brothers	Hand span (in cm)
M	Large	0	23.6
F	Small	2	19.6
F	Small	1	20.2
M	Large	1	24.0



Numerical and continuous data

The final column is an example of numerical and continuous data.

This is data which is measured and can take any number of decimal places.

Perfect examples of numerical and continuous data would be time, height, hand span, weight.

Anything which can be measured.

Cont



<i>Gender</i>	<i>Coffee size</i>	<i>Number of brothers</i>	<i>Hand span (in cm)</i>
M	Large	0	23.6
F	Small	2	19.6
F	Small	1	20.2
M	Large	1	24.0



Example: Categorical and numerical variables

Classify the following variables as categorical or numerical.

- Students choose their favourite pet from 'dog', 'cat', 'bird' or 'other'.
- The time, in seconds, taken to solve a puzzle is recorded.

C.

N



Example: Nominal and ordinal variables

Classify the following variables as nominal or ordinal.

- A group of people record their level of happiness as 'very happy', 'happy', 'not too happy' or 'very unhappy'. **Ord**
- Students select their favourite country to visit. **Nom**



Example: Discrete and continuous variables

Classify the following numerical variables as discrete or continuous.

- The number of children in the family is recorded for all the students in a school.
- The birth weight of babies, measured in grams, is recorded at a hospital.

Dis

Cont.



Example: Classifying variables

Classify the following numerical variables as nominal, ordinal, discrete or continuous.

- The number of students in each of 10 classes is counted. **D**
- The time taken for 20 mice to each complete a maze is recorded in seconds. **C**
- Diners at a restaurant were asked to rate how they felt about their meal: 1= Very satisfied, 2 = Satisfied, 3 = Indifferent, 4 = Dissatisfied, 5 = Very dissatisfied. **O**
- Students choose a colour from a list: 1= Blue, 2 = Green, 3 = Red, 4 = Yellow. **N**
- Students' heights were classified as 'less than 160 cm', '160 cm – 180 cm' or 'more than 180 cm'. **O**



Turning words and numbers into pictures

There are lots of ways to representing data but we tend to use certain pictures for certain data types.

A nice table showing this is given in the Cambridge General Maths textbook and would be good in any summary book!

- ▶ **Categorical** data can be summarised in a **frequency table** or a **percentage frequency table**.
- ▶ A frequency table can be displayed in a **bar chart**.
- ▶ A percentage frequency table can be displayed in a **percentage bar chart**.
- ▶ The value of a categorical variable with the highest frequency is called the **mode**.



The frequency table

A frequency table is a table where we work out the frequencies of each of the categories. Frequency means **how many**. When drawing a table we can be asked for either frequencies, percentage frequencies **or both**!

Example:

✓ Thirty children chose a sandwich, a salad or a pie for lunch, as follows:

~~sandwich, salad, salad, pie, sandwich, sandwich, salad, salad, pie, pie, pie, salad, pie, sandwich, salad,~~
~~pie, salad, pie, sandwich, sandwich, pie, salad, salad, pie, pie, pie, salad, pie, sandwich, pie~~

Construct a table for the data showing both frequency and percentage frequency.

	Freq	%freq
sand	7	23.33%
salad	10	33.33
pie	13	43.33
	30	100%

% freq

$$\frac{\text{freq}}{\text{total}} \times 100$$

$$\frac{7}{30} \times 100$$



The bar chart

A bar chart is used to display the information from a frequency table. This is true when there is lots of data and we want to be able to see if there are any patterns.

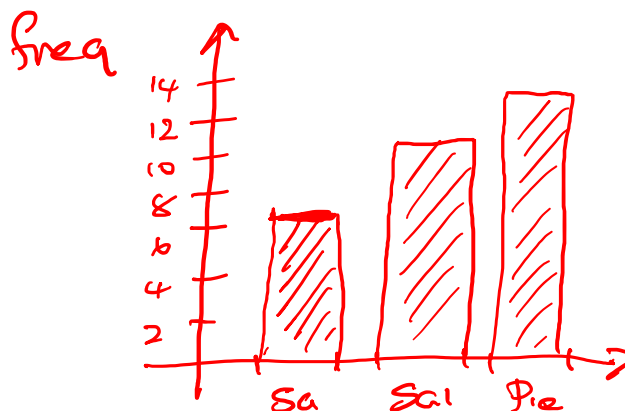
We use bar charts for **categorical data**.

It's important to know:

In a bar chart:

- frequency or percentage frequency is shown on the vertical axis
- the variable being displayed is plotted on the horizontal axis
- the height of the bar (column) gives the frequency (or percentage)
- the bars are drawn with gaps to indicate that each value is a separate category
- there is one bar for each category.

Lunch choice	Frequency	
	Number	%
Sandwich	7	23.3
Salad	10	33.3
Pie	13	43.3
Total	30	99.9



Example: Constructing a bar chart

Use the frequency table for Lunch choice from Example 5 to construct:

- a bar chart
- a percentage bar chart.

Lunch choice	Frequency	
	Number	%
Sandwich	7	23.3
Salad	10	33.3
Pie	13	43.3
Total	30	99.9

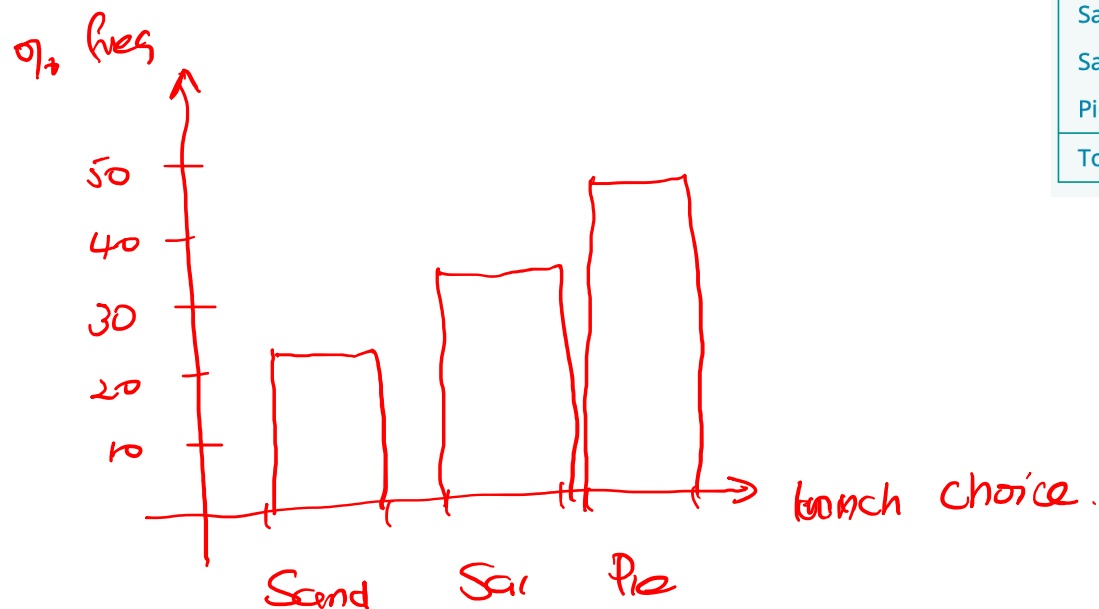


Example: Constructing a bar chart

Use the frequency table for Lunch choice from Example 5 to construct:

- a bar chart
- **a percentage bar chart.**

Lunch choice	Frequency	
	Number	%
Sandwich	7	23.3
Salad	10	33.3
Pie	13	43.3
Total	30	99.9



The mode or modal category

Mode means MOST.

When we are asked for the modal category we are never been asked for the value of the frequency. We are being asked to find which category has the highest number.

Lots of people make the common mistake of giving the highest number rather than the category.

So, from the table on the right, we can see the modal lunch choice would be Pie as it has the highest frequency (and percentage frequency).

We would NOT say 13.

Lunch choice	Frequency	
	Number	%
Sandwich	7	23.3
Salad	10	33.3
Pie	13	43.3
Total	30	99.9

modal cat = Pie



Why do we use the mode

In later lessons we will be able to compare the mode and the median and make some judgements based on how close they are to each other.

For now, it's important to note that we can have one or two modes in data. Anything more than this doesn't make a lot of sense!

We can use the mode to help us answer questions like, next week, do we order more pies or salads?

Lunch choice	Frequency	
	Number	%
Sandwich	7	23.3
Salad	10	33.3
Pie	13	43.3
Total	30	99.9

