

# M1 Maths

## S3-4 Data Types

- numerical, categorical and ordinal data
- univariate, bivariate and multivariate data
- time series

[Summary](#) [Learn](#) [Solve](#) [Revise](#) [Answers](#)

---

---

### Summary

---

---

Numerical data are generally numbers resulting from a count (discrete data) or a measurement (continuous data). Categorical data are not numbers or are numbers with no arithmetic significance (e.g. phone numbers). Ordinal data is data that defines an order.

Univariate data consists of one variable for each item surveyed; bivariate data consists of two; multivariate data consists of more than two.

A time series is an example of bivariate data where the independent variable is time and where the value of the dependent variable is recorded at regular time intervals.

---

---

### Learn

---

---

#### Numerical Data

When we collect data, it can be of different types

**Numerical data** are numbers, like people's heights, the number of pets they have, their ages, their income etc. Numerical data can be discrete or continuous.

**Discrete** data can only take on certain values. The number of pets people own must be whole numbers: no one can own 3.61 pets. Counting things generally produces discrete data.

**Continuous** data, however can take on any of a continuous range of values. A person's height can be 171 cm or 172 cm, but it can also be 171.3 cm or 171.28641 cm. We wouldn't measure a person's height to that degree of precision, but theoretically we could, so the data is continuous. Measuring things generally produces continuous data.

These concepts are basically the same as the concepts of discrete and continuous variables explained in Module A1-4 (Discrete and Continuous Relations). In fact the data can be thought of as a variable.

Now we might consider **money** to be a discrete variable because you can give someone \$4.50 or \$4.55, but not \$4.53875. Statisticians, however treat money as continuous. This is partly because banks can transfer \$4.53875 even though it can't be given in cash. It is also because

when dealing with thousands or millions of dollars, the gap between \$45 824.66 and \$45 824. 67 is so tiny as not to be worth worrying about.

## Categorical Data

Categorical data is usually data that is not numbers. If you asked people what their favourite school subject was, they might answer Maths, History, Phys. Ed. etc. These are the data, but they aren't numbers. They are categorical data.

Other examples might be people's hair colour, the type of pet they own, what they did last Saturday evening and so on.

Now we say 'Categorical data is *usually* data that is not numbers'. If you asked people for their favourite number, that would be categorical data. If someone answers 29, it doesn't mean 29 of anything; they haven't counted 29 of anything, nor have they measured anything. This kind of data is basically the same kind of data as the data we would get if we asked them their favourite letter (E, H, X etc.) or their favourite school subject.

Phone numbers would be another example of data that are numbers, but categorical. The important thing about numerical data is that it is meaningful to do arithmetic on the numbers. For example, if Claudia has 4 pets and Harry had 7, we can add them together to conclude that between them they have 11 pets. But if Claudia's favourite number was 4 and Harry's was 7, adding them together to get 11 doesn't tell us anything about anybody's favourite number. In the same way, if my phone number is 0412 117 441 and yours was 0424 131 228, adding them together to get 0836 248 669 would be a pointless exercise.

## Ordinal Data

Suppose you used a survey with a question like this:

We should build more coal-fired power stations? Circle one number							
Strongly disagree				Strongly agree			
0	1	2	3	4	5	6	

And suppose we get this data: 2, 0, 0, 6, 1, 3, 0, 1, 3

This kind of data is called ordinal data. Ordinal data is not numerical in that the numbers aren't a result of counting or measuring anything. But their order does matter (hence the term 'ordinal'). It would be silly to write the question like this:

We should build more coal-fired power stations? Circle one number							
Strongly disagree				Strongly agree			
4	1	6	0	3	5	2	

It is meaningful to take an average of the responses. The average of 2, 0, 0, 6, 1, 3, 0, 1, 3 is 1.8. This does give a useful summary of the data showing that public feeling is more against building more coal-fired power stations than for it.

## Practice

- Q1 Decide whether each of the following questions would produce numerical data, categorical data or ordinal data. If numerical, say whether it would be discrete or continuous.
- (a) What is your height?
  - (b) How many people live in your house?
  - (c) What is your favourite ice cream flavour?
  - (d) Which party do you want to win the election?
  - (e) How much do you earn per week?
  - (f) Number these election candidates from 1 for your first choice to 5 for your last choice?
    - Margaret Baker
    - Kim Bertie
    - Ronald Singh
    - Vlad Gdanski
    - Maria Jones
  - (g) What colour is your roof?
  - (h) What is the area of your block of land?
  - (i) What is the number on your football jersey?
  - (j) Put these in order from most to least enjoyable
    - Eating lunch
    - Having a tooth pulled out
    - Doing maths
    - Going to a folk festival
    - Doing the washing up

## Univariate Data

Univariate data is data where one thing (variable) is recorded, maybe the temperature, the mass of peaches, the time taken to run 100 m and so on.

When we have a number of univariate data (e.g. temperature at a number of different times, the masses of several peaches or the time taken for everyone in Year 10 to run 100 m), we can summarise the data by calculating the mean, median, mode, range etc. We can also graph it as dot plots, bar graphs, stem plots, box plots etc.

## Bivariate Data

Bivariate data is data where we record two things (variables) about each person or thing. For instance we might record people's height and the time they take to run 5 km, or we might

record their score on the English test and their score on the maths test, and so on.

We can calculate the mean, median etc. for each variable and we can draw dot plots, bar graphs etc. for each variable, just as we can for univariate data. But the value of bivariate data lies in being able to draw a scatter plot and see relations might exist between the two variables.

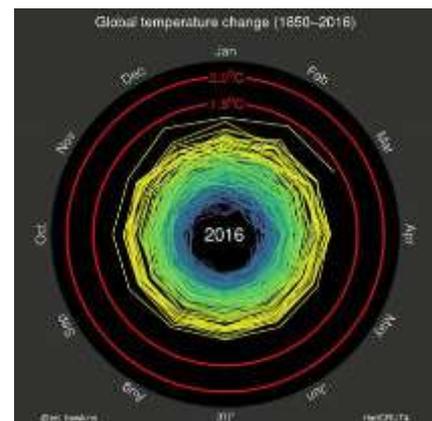
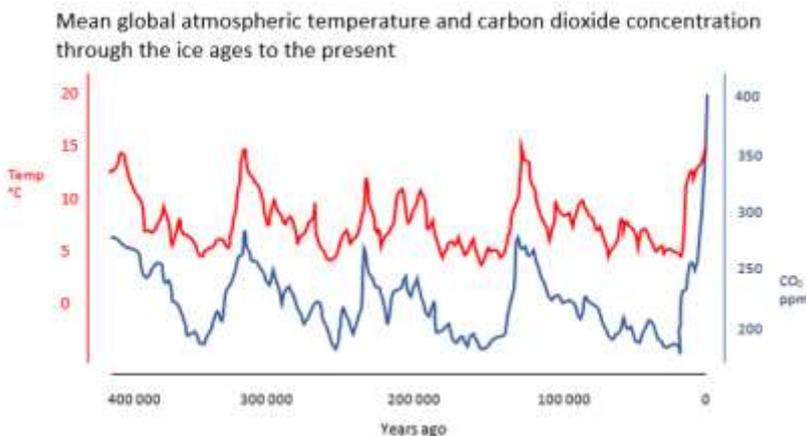
## Multi-variate Data

Multivariate data is like bivariate data except that more than two variables are recorded. There are ways of determining relationships between these variables, but they are beyond the scope of high-school maths.

## Time Series

A time series is an example of bivariate data where the independent variable is time and where the value of the dependent variable is recorded at regular time intervals. Here are three examples:

Time (24h)	0000	0020	0040	0060	0080	1000	1200	1400	1600	1800	2000	2200
Temperature	14.3	13.1	11.8	10.9	12.4	16.5	18.8	19.2	17.3	15.1	14.4	13.0



Used with permission from Ed Hawkins (@ed\_hawkins)

## Solve

- Q51 When Genghis moved into his new home on the first day of the month, he put \$80 000 under his mattress. On the last day of each month, he took 3% of what was there. Each year, on the anniversary of his moving in, his mother, who worried about him, added another \$10 000 to his hoard. Use a spreadsheet to produce a time series showing the amount there on the second of each month for the first 50 months that he lived there.

---

---

## Revise

---

---

### Revision Set 1

Q61 Decide whether each of the following questions would produce numerical data, categorical data or ordinal data. If numerical, say whether it would be discrete or continuous.

- (a) Who will you vote for on the council election?
- (b) How many nights did you spend away from home last year?
- (c) List your school subjects from favourite to least favourite
- (d) How long did you spend on homework last night?
- (e) What is the price of a ham sandwich at this cafe?

Q62 (a) What is bivariate data?

- (b) What commonly used graph of bivariate data cannot be used on univariate data?

---

---

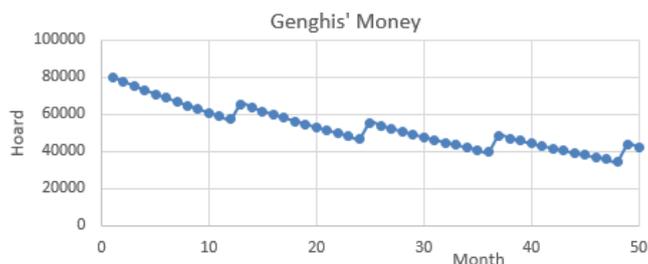
## Answers

---

---

- Q1 (a) numerical - continuous      (b) numerical - discrete      (c) categorical  
(d) categorical      (e) numerical - continuous      (f) ordinal  
(g) Categorical      (h) numerical - continuous      (i) categorical  
(j) ordinal

Q51



- Q61 (a) categorical      (b) numerical - discrete      (c) ordinal  
(d) numerical - continuous      (e) numerical - continuous
- Q62 (a) Data in which two variables are recorded for each subject  
(b) Scatter plot