

AVERAGES

Summer Term

Averages

Career Links

Being able to confidently work with data is a great skill to have and has lots of links with a number of careers such as:

- Statistician
- Business Analyst
- Biostatistician
- Healthcare

Key Terms:

Mean - Add up the values you are given and divide by the number of values you have.

Median - The median is the middle value, when your data is in order.

Mode - It is the value or item there is the most of.

Range - This is the difference between the largest and smallest values.

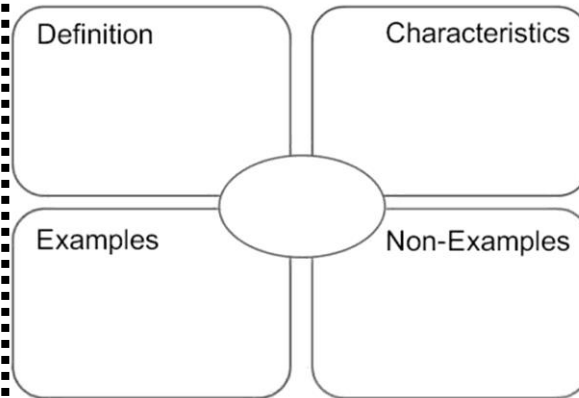
Frequency - the number of pieces of data we have.

Grouped Data - If we have a large spread of data, we put it into categories (classes) to make the data easier to display or analyse



What do I need to be able to do?

- Work out the mean, mode and median of a set of data
- Work out averages from a table



Topics

- Averages from a table
- Averages
 - Mean
 - Mode
 - Median

Key: 16|2 means 162

Stem	Leaf
15	3 4 4
16	2 5 9
17	2 6 6 6 7
18	0 2 6 7
19	0

Each number is split into two parts.

- The first digit(s) form the stem.
- The last digit forms the leaf.

Here is a discrete data set, calculate the mean, mode, median and range for this data.

2 5 3 9 7 7

Mode: 7

Range: $9 - 2 = 7$

Median: 2 3 5 7 7 9

Mean: $\frac{2+3+5+7+7+9}{6} = 5.5$

Put them in order first
Two numbers in the middle – add them together and divide by two

$$\frac{5+7}{2} = 6$$

There are six data values so divide by 6.

Average	Advantage	Disadvantage
Mode	Can be used for qualitative data Easy to obtain	There can be more than one mode or even no mode
Median	Not affected by very large or very small values	Can be time consuming when there is a lot of data
Mean	Takes into account all of the data	Very small or very large values affects the mean

Averages from Frequency Tables

a) Find the mean of this data

Goals Scored (x)	Frequency (f)	f × x
0	2	$0 \times 2 = 0$
1	2	$1 \times 2 = 2$
2	5	$2 \times 5 = 10$
3	1	$3 \times 1 = 3$
Total	10	15

Step 1: calculate the total frequency
Step 2: calculate $f \times x$
Step 4: calculate the mean

$$\text{Mean} = \frac{\text{Total } fx}{\text{Total } f}$$

$$\frac{\text{Total } fx}{\text{Total } f} = \frac{15}{10} = 1.5 \text{ goals}$$

b) Find the mode

The mode is the one with the highest frequency

Highest frequency = 5

Mode = 2 goals

c) Find the median

$$\text{Median value} = \frac{\text{Total frequency} + 1}{2}$$

$$\frac{11}{2} = 5.5 \text{th value}$$

add the frequency column until you reach the value in-between the 5th and 6th value

Median = 2 goals

d) Find the range

Highest number of goals = 3
Smallest number of goals = 0

Range = $3 - 0 = 3$

a) Estimate the mean of this data

Length (L cm)	Frequency (f)	Midpoint (x)	f × x
$0 < L \leq 10$	10	5	$10 \times 5 = 50$
$10 < L \leq 20$	15	15	$15 \times 15 = 225$
$20 < L \leq 30$	23	25	$23 \times 25 = 575$
$30 < L \leq 40$	7	35	$7 \times 35 = 245$
Total	55		1095

Step 1: calculate the total frequency
Step 2: find the midpoint of each group
Step 3: calculate $f \times x$
Step 4: calculate the mean

$$\text{Mean} = \frac{\text{Total } fx}{\text{Total } f}$$

$$\frac{\text{Total } fx}{\text{Total } f} = \frac{1095}{55} = 19.9 \text{ cm}$$

STATISTICS

Summer Term

Statistics 1

Career Links

Being able to confidently work with data is a great skill to have and has lots of links with a number of careers such as:

- Statistician
- Business Analyst
- Biostatistician

Topics

- Scatter graphs
- Sampling

What do I need to be able to do?

- Be able to plot and interpret a scatter graph
- Understand the limitations of sampling

Key Terms:

Origin – Where two axes meet on a graph.

Outlier – A point that lies outside the trend of the graph.

Relationship – The link between two variables e.g. between sunny days and ice cream sales.

Correlation – The mathematical definition for the type of relationship.

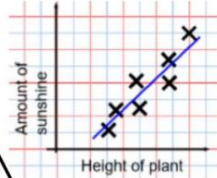
Line of Best Fit – A straight line on a graph that represents the data on a scatter graph.

The line of best fit

The Line of best fit is used to make estimates about the information in your scatter graph

Things to know:

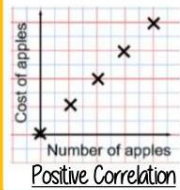
- The line of best fit **DOES NOT** need to go through the origin (The point the axes cross)
- There should be approximately the same number of points above and below the line (It may not go through any points)
- The line extends across the whole graph



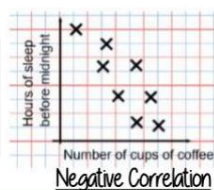
It is only an estimate because the line is designed to be an average representation of the data

It is always a straight line.

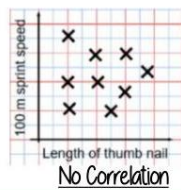
Linear Correlation



As one variable increases so does the other variable



As one variable increases the other variable decreases



There is no relationship between the two variables

Scatter Graphs – Explaining Patterns

A weatherman says "Temperatures are higher in towns that have more sunshine". Is this supported by the scatter graph?

Yes, the majority of points for high temperature appear when there are more hours of sunshine.

Interpolation and Extrapolation

Interpolation – making a prediction of a value that falls within the range of your data. This is more accurate.

Extrapolation – making a prediction of a value that falls outside the range of your data. This is less accurate.

Another town had a maximum temperature of 11°C that day. Use a line of best fit to estimate the hours of sunshine at this town.

- Step 1 – Draw a line of best fit = 9.5 hours
 Step 2 – Draw a line along from 11°C and down from the line of best fit

Comment on the reliability of your prediction.

= This is not a reliable estimate because it is extrapolation

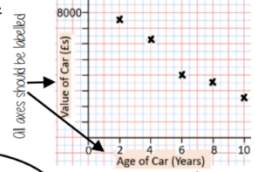
Draw and interpret a scatter graph

Age of Car (Years)	2	4	6	8	10
Value of Car (Es)	7500	6250	4000	3500	2500

- This data may not be given in size order
- The data forms information pairs for the scatter graph
- Not all data has a relationship

The link between the data can be explained verbally

"This scatter graph show as the age of a car increases the value decreases"



All axes should be labelled

The axis should fit all the values on and be equally spread out

Disadvantages of sampling

- Chances of bias.
- Difficulties in selecting truly a representative sample.
- Need for subject specific knowledge.
- changeability of sampling units.
- impossibility of sampling.

What is Sampling

Sampling might be characterized as the technique where a Sample is chosen from an individual or a gathering of a specific kind for research purposes. In Sampling, the population is isolated into various parts called examining units. Sampling guarantees accommodation, an assortment of concentrated and thorough data, suitability in restricted assets and better compatibility.

