Y10 Collecting, Representing and Interpreting

What do I need to be able to do?

By the end of this unit you should be able to:

- Construct and interpret frequency tables and polygon, two-way tables, line, bar, & pie charts
- Find and interpret averages from a list and a table
- Construct and interpret time series graphs, stem and leaf diagrams and scatter graphs

Keywords

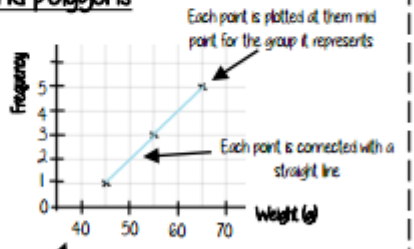
Population: the whole group that is being studied
Sample: a selection taken from the population that will let you find out information about the larger group
Representative: a sample group that accurately represents the population
Random sample: a group completely chosen by change. No predictability to who it will include.
Bias: a built-in error that makes all values wrong by a certain amount
Primary data: data collected from an original source for a purpose.
Secondary data: data taken from an external location. Not collected directly.
Outlier: a value that stands apart from the data set

Frequency tables and polygons

x Weight(g)	Frequency
$40 < x \leq 50$	1
$50 < x \leq 60$	3
$60 < x \leq 70$	5

We do not know from grouped data where each value is placed so have to use an estimate for calculations

MID POINTS
Mid-points are used as estimated values for grouped data. The middle of each group



Mid-point
Start point + End point
2

Two way tables

60 people visited the zoo one Saturday morning. 26 of them were adults. 13 of the adult's favourite animal was an elephant. 24 of the children's favourite animal was an elephant.

Extract information to input to the two-way table.

Subgroups each have their own heading

	Adult	Child	Total
Elephant	13	24	37
Other	13	10	23
Total	26	34	60

Needs subgroup totals — Overall total

Draw and interpret Pie Charts

Type of pet	Dog	Cat	Hamster
Frequency	32	25	3

There were 60 people asked in this survey (Total frequency)

$\frac{32}{60}$ "32 out of 60 people had a dog"

This fraction of the 360 degrees represents dogs

$\frac{32}{60} \times 360 = 192°$

Use a protractor to draw This is 192°

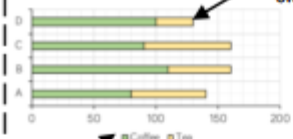
Multiple method
As 60 goes into 360 – 6 times
Each frequency can be multiplied by 6 to find the degrees (proportion of 360)

Comparing Pie Charts:
You NEED the overall frequency to make any comparisons

Bar and line charts

Compare the bars green compared to yellow. The size of each bar is the frequency. Overall total easily comparable

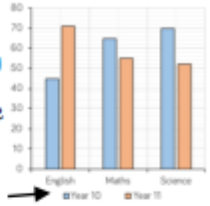
Composite bar charts



Categories clearly indicated

Dual bar charts

Bars are compared side by side
Easier to compare subgroups



Categories clearly indicated

Averages from a table

Non-grouped data

Number of Siblings	0	1	2
Frequency	6	8	6
Subtotal	0	8	12

Overall Frequency: 20

Total number of siblings: 20

The data in a list: 0,0,0,0,0,0,1,1,1,1,1,1,1,2,2,2,2,2,2,2

Mean: $\frac{\text{total number of siblings}}{\text{Total frequency}} = 1$

Grouped data

x Weight(g)	Frequency
$40 < x \leq 50$	1
$50 < x \leq 60$	3
$60 < x \leq 70$	5

Mid Point	MP x Freq
45	45
65	195
65	325

Overall Frequency: 9
Overall Total: 565
Mean: 62.8g

The data in a list: 45, 55, 55, 55, 65, 65, 65, 65, 65

Averages from lists

The Mean

A measure of average to find the central tendency... a typical value that represents the data

24, 8, 4, 11, 8

Find the sum of the data (add the values)
55
Divide the overall total by how many pieces of data you have
$55 \div 5$
Mean = 11

The Mode (The modal value)

This is the number OR the item that occurs the most (it does not have to be numerical)

24, 8, 4, 11, 8

This can still be easier if it the data is ordered first

Mode = 8

The Median

The value in the center (in the middle) of the data

24, 8, 4, 11, 8

Put the data in order 4, 8, 8, 11, 24
Find the value in the middle 4, 8, 8, 11, 24
Median = 8
NOTE: If there is no single middle value find the mean of the two numbers left

For Grouped Data

The modal group – which group has the highest frequency

Y10 Collecting, Representing and Interpreting

What do I need to be able to do?

By the end of this unit you should be able to:

- Construct and interpret frequency tables and polygon, two-way tables, line, bar, & pie charts
- Find and interpret averages from a list and a table
- Construct and interpret time series graphs, stem and leaf diagrams and scatter graphs

Keywords

Population: the whole group that is being studied
Sample: a selection taken from the population that will let you find out information about the larger group
Representative: a sample group that accurately represents the population
Random sample: a group completely chosen by change. No predictability to who it will include.
Bias: a built-in error that makes all values wrong by a certain amount
Primary data: data collected from an original source for a purpose.
Secondary data: data taken from an external location. Not collected directly.
Outlier: a value that stands apart from the data set

Stem and leaf

A way to represent data and use to find averages

This stem and leaf diagram shows the age of people in a line at the supermarket.

0	7 9
1	4 5 6 8 8
2	1 3
3	0

Key: 1|4 Means 14 years old

Stem and leaf diagrams:
Must include a key to explain what it represents
The information in the diagram should be ordered

Back to back stem and leaf diagrams

Girls		Boys
5	14	
7, 5, 5, 5, 4	15	3, 8, 9
8, 4, 2, 1, 0	16	2, 5, 7, 7, 7, 8, 8, 9
9, 8, 7, 6, 6, 4, 2, 1, 1, 0, 0	17	0, 2, 3, 6, 6, 7, 7
	18	0, 1, 4, 5

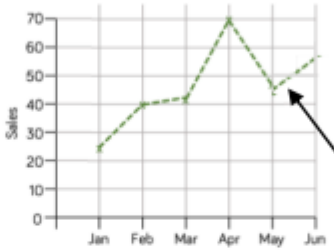
15 | 3
Means 153 cm tall

Back to back stem and leaf diagrams
Allow comparisons of similar groups
Allow representations of two sets of data

Time-Series

R

This time-series graph shows the total number of car sales in £1000 over time



Look for general trends in the data. Some data shows a clear increase or a clear decrease over time.

Readings in-between points are estimates (on the dotted lines). You can use them to make assumptions.

Comparing distributions

Comparisons should include a statement of average and central tendency, as well as a statement about spread and consistency

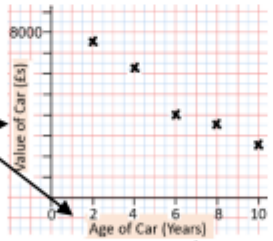
Mean, mode, median – allows for a comparison about more or less average
Range – allows for a comparison about reliability and consistency of data

Draw and interpret a scatter graph

R

Age of Car (Years)	2	4	6	8	10
Value of Car (£s)	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



Both axes should be labelled

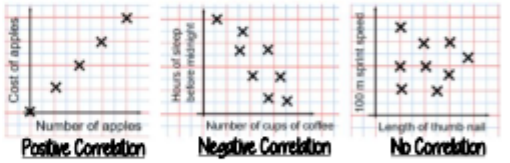
The axis should fit all the values on and be equally spread out

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

The link between the data can be explained verbally

Linear Correlation

R



Positive Correlation
As one variable increases so does the other variable

Negative Correlation
As one variable increases the other variable decreases

No Correlation
There is no relationship between the two variables

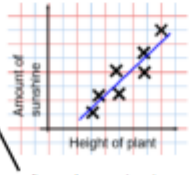
The line of best fit

R

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

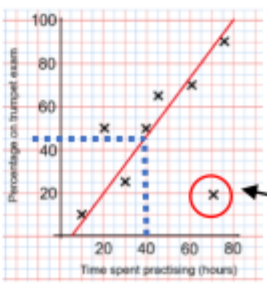
It is always a **straight line**

Using a line of best fit

R

Interpolation is using the line of best fit to estimate values inside our data point.

e.g. 40 hours revising predicts a percentage of 45.



Extrapolation is where we use our line of best fit to predict information outside of our data.
This is not always useful – in this example you cannot score more that 100%. So revising for longer can not be estimated

This point is an **"outlier"**
It is an outlier because it doesn't fit this model and stands apart from the data