

YEAR 8 — REPRESENTATIONS...

Representing Data

What do I need to be able to do?

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

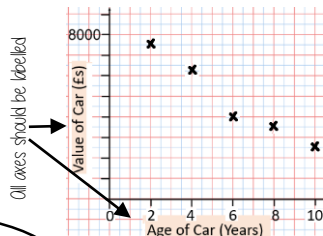
- Draw and interpret scatter graphs
- Describe correlation and relationships
- Identify different types of non-linear relationships
- Design and complete an ungrouped frequency table
- Read and interpret grouped tables (discrete and continuous data)
- Represent data in two way tables

Keywords

- Variable:** a quantity that may change within the context of the problem
- Relationship:** the link between two variables (items) Eg Between sunny days and ice cream sales
- Correlation:** the mathematical definition for the type of relationship.
- Origin:** where two axes meet on a graph
- Line of best fit:** a straight line on a graph that represents the data on a scatter graph
- Outlier:** a point that lies outside the trend of graph
- Quantitative:** numerical data
- Qualitative:** descriptive information, colours, genders, names, emotions etc.
- Continuous:** quantitative data that has an infinite number of possible values within its range.
- Discrete:** quantitative or qualitative data that only takes certain values
- Frequency:** the number of times a particular data value occurs

Draw and interpret a scatter graph

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



All axes should be labelled

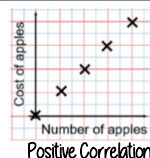
The axis should fit all the values on and be equally spread out

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

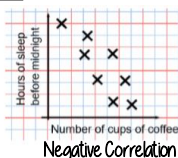
The link between the data can be explained verbally

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

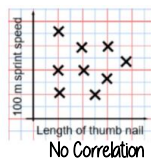
Linear Correlation



As one variable increases so does the other variable



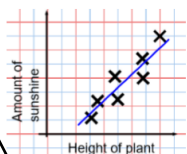
As one variable increases the other variable decreases



There is no relationship between the two variables

The line of best fit

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

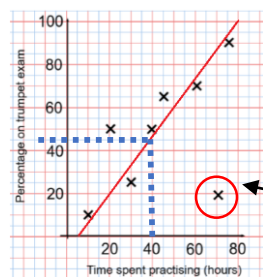
Things to know:

- 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

Using a line of best fit

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 than 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

Ungrouped Data

The table shows the number of siblings students have. The answers were 3, 1, 2, 2, 0, 3, 4, 1, 1, 2, 0, 2

The number of times an event happened

| Number of siblings | Frequency |
|--------------------|-----------|
| 0 | 2 |
| 1 | 3 |
| 2 | 4 |
| 3 | 2 |
| 4 | 1 |

Best represented by discrete data (Not always a number)

2 people had 0 siblings. This means there are 0 siblings to be counted here

0 → 0
3 → 3
2 + 2 + 2 + 2 OR 2 x 4 = 8

3 + 3 OR 3 x 2 = 6

2 people have 3 siblings so there are 6 siblings in total

OVERALL there are 0 + 3 + 8 + 6 + 4 Siblings = 21 siblings

Grouped Data

If we have a large spread of data it is better to group it. This is so it is easier to look for a trend. Form groups of equal size to make comparison more valid and spread the groups out from the smallest to the largest value.

| Cost of TV (£) | Tally | Frequency |
|----------------|---------|-----------|
| 101 - 150 | THH | 7 |
| 151 - 200 | THH THH | 11 |
| 201 - 250 | THH | 5 |
| 251 - 300 | | 3 |

We do not know the exact value of each item in a group — so an estimate would be used to calculate the overall total (Mid Point)

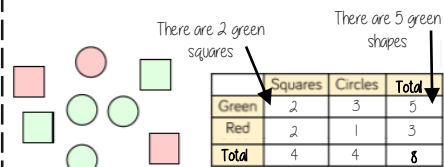
| x | Frequency |
|-------------|-----------|
| Weight(g) | |
| 40 < x ≤ 50 | 1 |
| 50 < x ≤ 60 | 3 |
| 60 < x ≤ 70 | 5 |

To make sure all values are included inequalities represent the subgroups

e.g this group includes every weight bigger than 60kg, up to and including 70kg

Representing data in two-way tables

Two way tables represent discrete information in a visual way that allows you to make conclusions, find probability or find totals of sub groups



Using your two way table

To find a fraction e.g What fraction of the items are red? 3 red items but 8 items in total = $\frac{3}{8}$

Interleaving Use your fraction, decimal percentage, equivalence knowledge