Year 8 Knowledge Organiser – Statistics

Objectives

Use and interpret scatter graphs of bivariate data

Recognise correlation

Interpret, analyse and compare the distributions of data sets from univariate empirical distributions through appropriate graphical representation involving discrete, continuous and grouped data

Interpret, analyse and compare the distributions of data sets from univariate empirical distributions through appropriate measures of central tendency (median, mean, mode and modal class) and spread (range, including consideration of outliers)

Apply statistics to describe a population

Key Vocabulary

Scatter graph – a graph of plotted points that show the relationship between two sets of data

Univariate data - a type of data which consists of observations on only a single characteristic or attribute

Bivariate data – data for two variables (usually two types of related data). E.g: Ice cream sales versus the temperature on that day

Correlation – the measure of the strength of a linear relationship between two sets of data

Discrete data – numerical data that can only take certain values e.g. shoe sizes

Continuous data – numerical data that can take any value within a given range

Central tendency – a single value that attempts to describe a set of data by identifying the central position within that set of data

> Discrete data takes certain numerical values – e.g shoe size, goals scored, number of pupils.
> Continuous data can be any value within a range – e.g height, temperature, weight.

Comparing distributions

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

Here are the number of runs scored last month by Lucy and James in cricket matches
Lucy: 45, 32, 37, 41, 48, 35
James: 60, 90, 41, 23, 14, 23

Lucy
Mean: 39.6 (1dp), Median: 38 Mode: no mode, Range: 16

James
Mean: 41.8 (1dp), Median: 32, Mode: 23, Range: 76

James has two extreme values that have a big impact on the range

"James is less consistent that Lucy because his scores have a greater range. Lucy performed better on average because her scores have a similar mean and a higher median"

Find and interpret the range

The range is a measure of **spread**

Difference between the biggest and smallest values
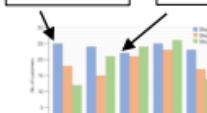
Shop 1 highest value
Shop 1 lowest value

A smaller range means there is less variation in the results – it is more consistent data
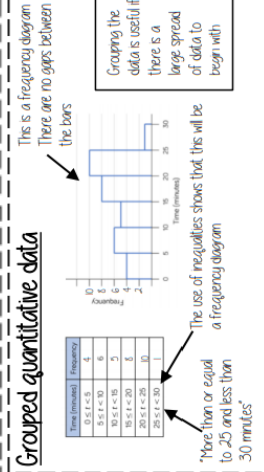
A range of 0 means all the data is the same value

Shop 1 has the smallest range – this indicates it has a more consistent flow of customers each week

Range of customers - 25 – 22 = 3 (Shop 1)

Grouped quantitative data

This is a frequency diagram. There are no gaps between the bars

Grouping the data is useful if there is a large spread of data to begin with

The use of inequalities shows that this will be a frequency diagram

Time (minutes)	Frequency
0 ≤ t < 5	4
5 ≤ t < 10	6
10 ≤ t < 15	15
15 ≤ t < 20	8
20 ≤ t < 25	6
25 ≤ t < 30	10

"More than or equal to 25 and less than 30 minutes"

Mean, Median, Mode

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 ÷ 5

Mean = 11

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

NOTE: If there is no single middle value find the mean of the two numbers left

Median = 8

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
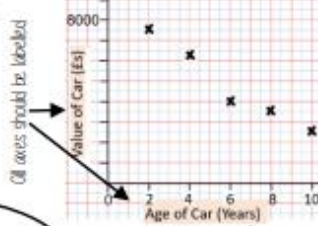
This can still be easier if the data is ordered first

4, 8, 8, 11, 24

Mode = 8

Draw and interpret a scatter graph

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
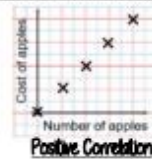- Not all data has a relationship

All axes should be labelled

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

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

The link between the data can be explained verbally

Linear Correlation

Positive Correlation
Number of apples
As one variable increases so does the other variable
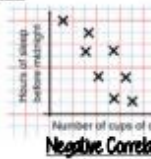
Negative Correlation
Number of cups of coffee
As one variable increases the other variable decreases
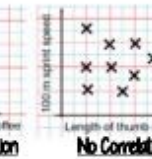
No Correlation
Length of thumb nail
There is no relationship between the two variables