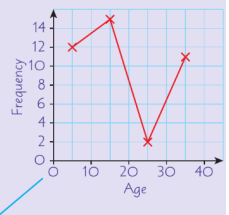
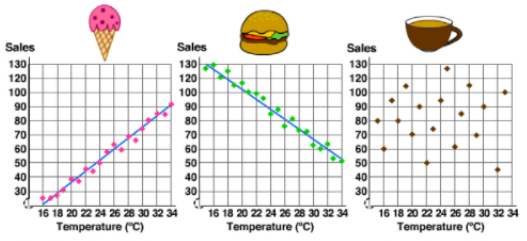
| | | | |
|---|---|---|---|
| **Prior Knowledge** | | | |
| 1 | Frequency | | The number of times something occurs. |
| 2 | Mode from a list | 3, 4, 5, 5, 5, 6, 6, 7, 8, 8, 9<br>mode = 5 | The value that occurs **most** often, or has the highest frequency. |
| 3 | Median from a list | 1, 3, 3, **6**, 7, 8, 9<br>Median = 6<br>1, 2, 3, **4**, **5**, 6, 8, 9<br>Median = (4 + 5) ÷ 2<br>= 4.5 | The middle value when the data is in **order**.<br>If there are an even number of values the median lies halfway between the two middle values. |
| 4 | Mean from a list | $\frac{\text{sum of amounts}}{\text{number of amounts}}$ | The total of the values divided by the number of values. |
| 5 | mean from a table | Jack asked students in his class how many pets they had. Here are his results. Work out the mean.<br>Number of pets / Frequency / Total number of pets<br>0 / 7 / 0 × 7 = 0<br>1 / 8 / 1 × 8 = 8<br>2 / 6 / 2 × 6 = 12<br>3 / 3 / 3 × 3 = 9<br>4 / 1 / 4 × 1 = 4<br>Total / 25 / 33<br>Add a column to the table.<br>6 people with 2 pets each.<br>Work out the total number of pets.<br>Work out the total frequency (number of people).<br>mean = $\frac{33}{25}$ = 1.32<br>mean = total number of pets ÷ total number of people<br>Use a calculator. | To calculate mean use<br>$\text{mean} = \frac{\sum f \times x}{\sum x}$<br>where f is frequency<br>x is the variable<br>$\Sigma$ means the sum of |
| 6 | Range | range = biggest value - smallest value | The spread of the data. Calculate by subtracting the smallest value from the biggest. |
| 7 | Modal class | | The class with the highest frequency in a grouped frequency table. |
| 8 | Comparing data sets | *"Paul's jumps are less consistent than Daniel's because his jumps have a greater range. Daniel jumps further on average than Paul because his jumps have a greater mean."* | Find an average and the range, then write 2 sentences comparing the data. |
| **Core Knowledge** | | | |
| 8 | Primary data | | Data that you collect yourself |
| 9 | Secondary data | | Data that is collected by someone else |
| 10 | Population | eg. If surveying favourite food of students at TBS the population would be all of the students at TBS | Total number of items that a survey relates to |
| 11 | Sample | A good sized sample is usually about 10% of the population | A group within a population. A sample is used when asking the whole population would be too difficult. |
| 12 | Hypothesis | eg: 'Most households have more than one tv' (it may not be true but is a statement to be tested) | A statement that you can test by collecting data in a questionnaire, survey or experiment |
| 13 | Bias | eg If you only asked your friends in a survey it might not represent the whole population. It would be a *biased* sample | **Bias** is the tendency of a **statistic** to overestimate or underestimate a parameter |
| 14 | Random sample | eg. names out of a hat | In a random sample every item is equally likely to be chosen |
| 15 | Data collection sheet | Car colour / Tally / Frequency<br>red<br>blue | A table or chart for collecting data |
| 16 | Grouped data | The class $4 \leqslant l < 6$ includes all values of length $l$ from $l = 4$ cm up to, but not including, 6 cm.<br>4 5 6<br>cm<br>2cm | Data can be grouped into classes. |

| | | | |
|---|---|---|---|
| 17 | Grouped frequency table | For discrete data<br>Mark / Frequency: 1–10: 4; 11–20: 23; 21–30: 26<br>For continuous data<br>Length, $l$ (mm) / Frequency: $3 < l \leqslant 4$: 7; $4 < l \leqslant 5$: 18; $5 < l \leqslant 6$: 27 | A frequency table that has several equal classes |
| 18 | Discrete data | *eg Number or siblings or shoe size* | Can only take certain values |
| 19 | Continuous data | *eg height, weight, time* | Is measured and can take any value |
| 20 | Class width | Height / Frequency / Class width: 140≤h<150: 8: 10; 150≤h<155: 9: 5; 155≤h<160: 4: 5 | The width of a class in a frequency table. |
| 21 | Median (position) | In a set of $n$ data values, the median is the $\frac{n+1}{2}$th one. | When data is grouped you can identify the position of the median value by adding 1 to the number of values (n) and dividing by 2. |
| 22 | median from a frequency table | The table shows the numbers of pets people own. Find the median number of pets.<br>Number of pets / Frequency: 0: 7; 1: 5; 2: 2; 3: 4; 4: 2; Total: 20<br>1st–7th data values; 8th–12th data values<br>Median is the $\frac{20+1}{2}$th $= \frac{21}{2} = 10.5$th value<br>Median is the $\frac{n+1}{2}$th value.<br>The median is 1 pet. | Find the $\frac{n+1}{2}$th number in the table |
| 23 | Estimate mean from grouped data | Score / Frequency, $f$ / Midpoint of class, $m$ / $m \times f$: 1–5: 5: 3: 15; 6–10: 6: 8: 48; 11–15: 9: 13: 117; 16–20: 10: 18: 180; Total: 30: Total: 360<br>Add a column, $m \times f$ to calculate an estimate of the total score for each class.<br>Estimate of mean $= \frac{360}{30} = 12$<br>Divide the total of the $m \times f$ column by the total frequency.<br>The mean is an estimate because we don't know the exact test scores. | When the data is grouped you can calculate an ESTIMATE of the mean by using the MIDPOINTS of the classes. |
| 24 | Outlier | | An extreme value, or anomaly, that doesn't fit the pattern of the other data values. |
| 25 | Line graph for grouped data | **Worked example**<br>Draw a line graph to represent this data.<br>Age, $a$ / Frequency / Midpoint: $0 \leqslant a < 10$: 12: 5; $10 \leqslant a < 20$: 15: 15; $20 \leqslant a < 30$: 2: 25; $30 \leqslant a < 40$: 11: 35<br>First work out the midpoint of each class.<br>Plot each frequency against the midpoint age.<br>  | To draw a line graph for grouped data you need to plot the frequencies at the midpoints of the classes |
| 26 | Scatter graph |  <br>Positive Correlation: A positive trend - as one set of values increases, the other set increases.<br>Negative Correlation: A negative trend - as one set of values increases, the other set decreases.<br>Zero Correlation: No trend - the points are scattered randomly with no visible pattern. | This shows two sets of data on the same graph. The shape of the graph shows if there is a correlation between the data sets. |
| 27 | Line of best fit | | A line of best fit represents the trend of the correlation. When drawing, aim for a balance of points above and below the line. |
| 28 | writing a report | | A report could include:<br>-the hypothesis you are investigating<br>-the data shown in a graph or chart<br>-averages and range<br>-a conclusion<br>-what else you could investigate. |