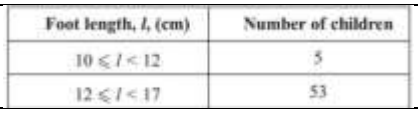
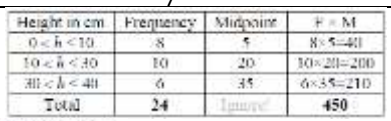
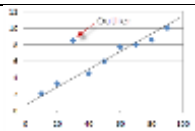
Core Knowledge

Topic: Summarising Data

Topic/Skill	Definition/Tips	Example
1. Types of Data	**Qualitative** Data – **non-numerical** data **Quantitative** Data – **numerical** data **Continuous** Data – data that can take **any numerical value** within a given range. **Discrete** Data – data that can take **only specific values** within a given range.	Qualitative Data – eye colour, gender etc. Continuous Data – weight, voltage etc. Discrete Data – number of children, shoe size etc.
2. Grouped Data	Data that has been **bundled in to categories**. Seen in grouped frequency tables, histograms, cumulative frequency etc.	 Foot length, l, (cm) | Number of children $10 \leqslant l < 12$ | 5 $12 \leqslant l < 17$ | 53
3. Primary /Secondary Data	**Primary** Data – **collected yourself** for a specific purpose. **Secondary** Data – **collected by someone else** for another purpose.	Primary Data – data collected by a student for their own research project. Secondary Data – Census data used to analyse link between education and earnings.
4. Mean	**Add** up the values and **divide** by how many values there are.	The mean of 3, 4, 7, 6, 0, 4, 6 is $\frac{3+4+7+6+0+4+6}{7} = 5$
5. Mean from a Table	1. Find the midpoints (if necessary) 2. Multiply Frequency by values or midpoints 3. Add up these values 4. Divide this total by the Total Frequency If **grouped** data is used, the answer will be an **estimate**.	 Height in cm | Frequency | Midpoint | F × M $0 < h < 10$ | 8 | 5 | 8×5=40 $10 < h < 30$ | 10 | 20 | 10×20=200 $30 < h < 40$ | 6 | 35 | 6×35=210 Total | 24 | | 450 **Estimated** Mean height: 450 ÷ 24 = 18.75cm
6. Median Value	The **middle** value. Put the data in order and find the middle one. If there are **two middle values**, find the number half way between them by **adding them together and dividing by 2**.	Find the median of: 4, 5, 2, 3, 6, 7, 6 Ordered: 2, 3, 4, **5**, 6, 6, 7 Median = 5
7. Median from a Table	Use the formula $\frac{(n+1)}{2}$ to find the position of the median. n is the total frequency.	If the total frequency is 15, the median will be the $\left(\frac{15+1}{2}\right) = 8th$ position
8. Mode /Modal Value	**Most** frequent/common. Can have more than one mode (called bi-modal or multi-modal) or no mode (if all values appear once)	Find the mode: 4, 5, 2, 3, 6, 4, 7, 8, 4 Mode = 4
9. Range	**Highest value subtract the Smallest value** Range is a 'measure of spread'. The smaller the range the more <u>consistent</u> the data.	Find the range: 3, 31, 26, 102, 37, 97. Range = 102-3 = 99
10. Outlier	A value that '**lies outside**' most of the other values in a set of data. An outlier is **much smaller or much larger** than the other values in a set of data.	
11. Lower Quartile	**Divides** the **bottom half** of the data into **two halves**. **LQ = $Q_1 = \frac{(n+1)}{4}$ th value**	Find the lower quartile of: 2, **<u>3</u>**, 4, 5, 6, 6, 7 $Q_1 = \frac{(7+1)}{4} = 2nd$ value → 3
12. Lower Quartile	**Divides** the **top half** of the data into **two halves**. **UQ = $Q_3 = \frac{3(n+1)}{4}$ th value**	Find the upper quartile of: 2, 3, 4, 5, 6, **<u>6</u>**, 7 $Q_3 = \frac{3(7+1)}{4} = 6th$ value → 6
13. Interquartile Range	The **difference** between the **upper quartile and lower quartile.** $IQR = Q_3 - Q_1$ The **smaller** the **interquartile range**, the **more consistent** the data.	Find the IQR of: 2, 3, 4, 5, 6, 6, 7 $IQR = Q_3 - Q_1 = 6 - 3 = 3$

Core Knowledge

<table>
<tr><td>Sample size</td><td>Data which is representative of the population</td><td>The greater the precision required to represent the population, the larger the sample size needs to be.</td></tr>
<tr><td>Sampling methods</td><td>Random each member of a population is equally likely to be selected

Stratified sample used to select a sample that is representative of different groups.</td><td>Possible methods include using a random number generator from a computer programme, rolling a number of dice or using the random number button on a scientific calculator.

Example Billy wants to survey 25 customers of a restaurant to find out which dessert they prefer. He decides to use a stratified sampling technique to work out how many people of each age group he should select.

The table below shows how many customers attended the restaurant in the last week. This is the total population. The sample size is the number of customers Billy wants to survey, 25 in this example.The strata size is the number of people in each group, 12, 34, 48, 21 and 3 in this example.

<table>
<tr><th>Age group</th><th>Number of customers</th></tr>
<tr><td>11-20</td><td>12</td></tr>
<tr><td>21-30</td><td>34</td></tr>
<tr><td>31-40</td><td>48</td></tr>
<tr><td>41-50</td><td>21</td></tr>
<tr><td>51+</td><td>3</td></tr>
</table>
The total number of customers = 12 + 34 + 48 + 21 + 3 = 118.

He then uses the equation:

$Number\ selected\ from\ each\ strata = (\frac{strata\ size}{total\ population}) \times sample\ size$

<table>
<tr><th>Age group</th><th>Number in sample</th></tr>
<tr><td>11-20</td><td>$(\frac{3}{118})$ x 25 = 2.54 (3 customers)</td></tr>
<tr><td>21-30</td><td>$(\frac{3}{118})$ x 25 = 7.20 (7 customers)</td></tr>
<tr><td>31-40</td><td>$(\frac{3}{118})$ x 25 = 10.17 (10 customers)</td></tr>
<tr><td>41-50</td><td>$(\frac{3}{118})$ x 25 = 4.45 (4 customers)</td></tr>
<tr><td>51+</td><td>$(\frac{3}{118})$ x 25 = 0.63 (1 customer)</td></tr>
</table>
</td></tr>
</table>

Links to Data diagrams such as Cumulative Frequency diagrams, Histograms, Biology – make sure you know the differences as to how to calculate the range.