

Topic: Summarising Data

Topic/Skill	Definition/Tips	Example																				
1. Types of Data	<p>Qualitative Data – non-numerical data</p> <p>Quantitative Data – numerical data</p> <p>Continuous Data – data that can take any numerical value within a given range.</p> <p>Discrete Data – data that can take only specific values within a given range.</p>	<p>Qualitative Data – eye colour, gender etc.</p> <p>Continuous Data – weight, voltage etc.</p> <p>Discrete Data – number of children, shoe size etc.</p>																				
2. Grouped Data	<p>Data that has been bundled in to categories.</p> <p>Seen in grouped frequency tables, histograms, cumulative frequency etc.</p>	<table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Foot length, l, (cm)</th> <th>Number of children</th> </tr> </thead> <tbody> <tr> <td>$10 \leq l < 12$</td> <td>5</td> </tr> <tr> <td>$12 \leq l < 17$</td> <td>53</td> </tr> </tbody> </table>	Foot length, l , (cm)	Number of children	$10 \leq l < 12$	5	$12 \leq l < 17$	53														
Foot length, l , (cm)	Number of children																					
$10 \leq l < 12$	5																					
$12 \leq l < 17$	53																					
3. Primary /Secondary Data	<p>Primary Data – collected yourself for a specific purpose.</p> <p>Secondary Data – collected by someone else for another purpose.</p>	<p>Primary Data – data collected by a student for their own research project.</p> <p>Secondary Data – Census data used to analyse link between education and earnings.</p>																				
4. Mean	<p>Add up the values and divide by how many values there are.</p>	<p>The mean of 3, 4, 7, 6, 0, 4, 6 is</p> $\frac{3 + 4 + 7 + 6 + 0 + 4 + 6}{7} = 5$																				
5. Mean from a Table	<ol style="list-style-type: none"> 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 <p>If grouped data is used, the answer will be an estimate.</p>	<table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Height in cm</th> <th>Frequency</th> <th>Midpoint</th> <th>F × M</th> </tr> </thead> <tbody> <tr> <td>$0 < h \leq 10$</td> <td>8</td> <td>5</td> <td>$8 \times 5 = 40$</td> </tr> <tr> <td>$10 < h \leq 30$</td> <td>10</td> <td>20</td> <td>$10 \times 20 = 200$</td> </tr> <tr> <td>$30 < h \leq 40$</td> <td>6</td> <td>35</td> <td>$6 \times 35 = 210$</td> </tr> <tr> <td>Total</td> <td>24</td> <td>Ignore!</td> <td>450</td> </tr> </tbody> </table> <p>Estimated Mean height: $450 \div 24 = 18.75\text{cm}$</p>	Height in cm	Frequency	Midpoint	F × M	$0 < h \leq 10$	8	5	$8 \times 5 = 40$	$10 < h \leq 30$	10	20	$10 \times 20 = 200$	$30 < h \leq 40$	6	35	$6 \times 35 = 210$	Total	24	Ignore!	450
Height in cm	Frequency	Midpoint	F × M																			
$0 < h \leq 10$	8	5	$8 \times 5 = 40$																			
$10 < h \leq 30$	10	20	$10 \times 20 = 200$																			
$30 < h \leq 40$	6	35	$6 \times 35 = 210$																			
Total	24	Ignore!	450																			
6. Median Value	<p>The middle value.</p> <p>Put the data in order and find the middle one.</p> <p>If there are two middle values, find the number half way between them by adding them together and dividing by 2.</p>	<p>Find the median of: 4, 5, 2, 3, 6, 7, 6</p> <p>Ordered: 2, 3, 4, 5, 6, 6, 7</p> <p>Median = 5</p>																				
7. Median from a Table	<p>Use the formula $\frac{(n+1)}{2}$ to find the position of the median.</p> <p>n is the total frequency.</p>	<p>If the total frequency is 15, the median will be the $\left(\frac{15+1}{2}\right) = 8\text{th}$ position</p>																				
8. Mode /Modal Value	<p>Most frequent/common.</p> <p>Can have more than one mode (called bi-modal or multi-modal) or no mode (if all values appear once)</p>	<p>Find the mode: 4, 5, 2, 3, 6, 4, 7, 8, 4</p> <p>Mode = 4</p>																				
9. Range	<p>Highest value subtract the Smallest value</p>	<p>Find the range: 3, 31, 26, 102, 37, 97.</p> <p>Range = $102 - 3 = 99$</p>																				

	Range is a 'measure of spread'. The smaller the range the more <u>consistent</u> the data.	
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 . $LQ = Q_1 = \frac{(n+1)}{4} \text{th value}$	Find the lower quartile of: 2, <u>3</u> , 4, 5, 6, 6, 7 $Q_1 = \frac{(7+1)}{4} = 2nd \text{ value} \rightarrow 3$
12. Lower Quartile	Divides the top half of the data into two halves . $UQ = Q_3 = \frac{3(n+1)}{4} \text{th value}$	Find the upper quartile of: 2, 3, 4, 5, 6, <u>6</u> , 7 $Q_3 = \frac{3(7+1)}{4} = 6th \text{ value} \rightarrow 6$
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 $IQR = Q_3 - Q_1 = 6 - 3 = 3$