Data Analysis (3)

Dealing with grouped continuous data

When continuous data has been grouped, the original values have been lost to us and so we can only *estimate* the value of the mean and identify the group that is the mode and median.

Example (6)

The table below shows the heights in cm of 110 sunflower plants measured two months after planting:

Height of plant, h (cm)	Frequency
$100 \le h < 110$	6
$110 \le h < 120$	13
$120 \le h < 130$	35
$130 \le h < 140$	29
$140 \le h < 150$	16
$150 \le h < 160$	11

As with any grouped table, make sure you understand what the data is telling you by selecting a row of data and putting it in context:

Height of plant, h (cm)	Frequency
$100 \le h < 110$	6
$110 \le h < 120$	13
$120 \le h < 130$	35
$130 \le h < 140$	29
$140 \le h < 150$	16
$150 \le h < 160$	11

As we aren't given the raw data, we don't know the exact measurements of these 35 sunflowers. So in order to do the maths, we assume they will all average out around the middle, ie that we have 35 sunflowers each measuring 125cm in height. The total height of the flowers in that group will then be

We then repeat this for each of the groups in the table:

Height of plant, h (cm)	Frequency	Mid-Value	Total
100 ≤ h < 110	6		
110 ≤ h < 120	13		
120 ≤ h < 130	35		
130 ≤ h < 140	29		
140 ≤ h < 150	16		
150 ≤ h < 160	11		

Remember now that **mean = total ÷ count**, which in this case is the total height of all the sunflowers divided by the number of sunflowers measured.

This will be an estimate, because we have had to use the mid-values in each group.

The modal group will be:

The median group will be:

Example (7)

The table shows the length, in seconds, of 100 calls made from a mobile phone:

Length of call, L (sec)	Frequency
0 ≤ L < 30	51
30 ≤ L < 50	25
50 ≤ L < 60	13
60 ≤ L < 120	7
120 ≤ L < 150	

Calculate an estimate for the mean and identify the modal and median groups.