

Cumulative Frequency

We have already learned that when data has been grouped and we no longer have the raw data, it is only possible to estimate some measures of average and spread.

The table below shows my scores at throwing darts over 60 sets of throws:

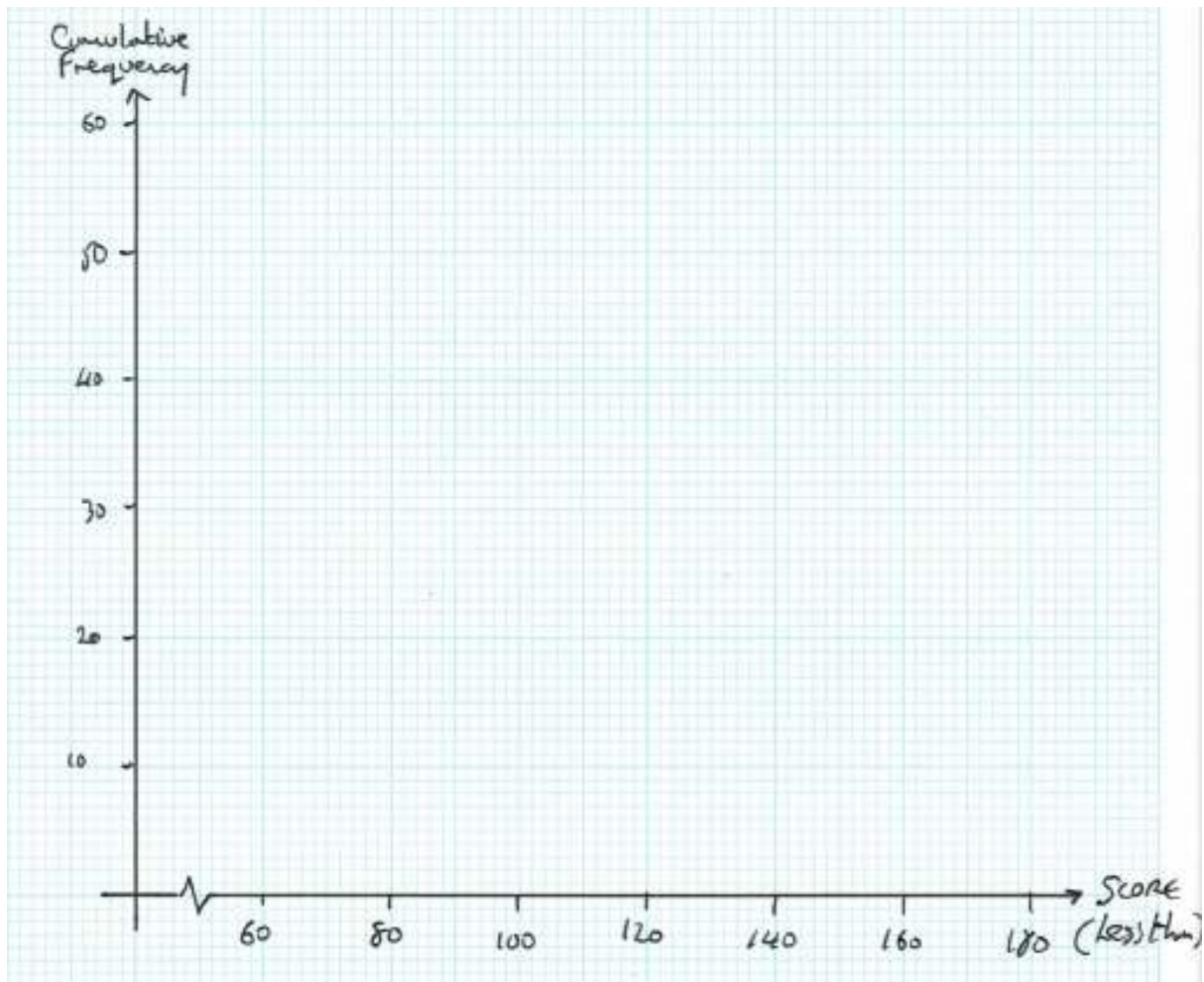
Score, s	Frequency		
$30 < s \leq 60$	10		
$60 < s \leq 90$	4		
$90 < s \leq 120$	13		
$120 < s \leq 150$	23		
$150 < s \leq 180$	10		

Estimate my mean score, the modal and median groups.

Another way of working with grouped data is to produce a **cumulative frequency curve (or graph)** which can be used to **estimate a median value** rather than a group and a different measure of range, the **inter-quartile range**, which focuses on how spread out the middle half of the data is. This is useful because it isn't affected by extreme data values (think of the example I used by including my age in the ages of the people in the classroom)

Before we can draw the graph, we need to produce a running total (accumulating, cumulative) for our data:

Score, s	Frequency	$s \leq$	cumulative frequency
$30 < s \leq 60$	10		
$60 < s \leq 90$	4		
$90 < s \leq 120$	13		
$120 < s \leq 150$	23		
$150 < s \leq 180$	10		



We can now use the graph to estimate the median

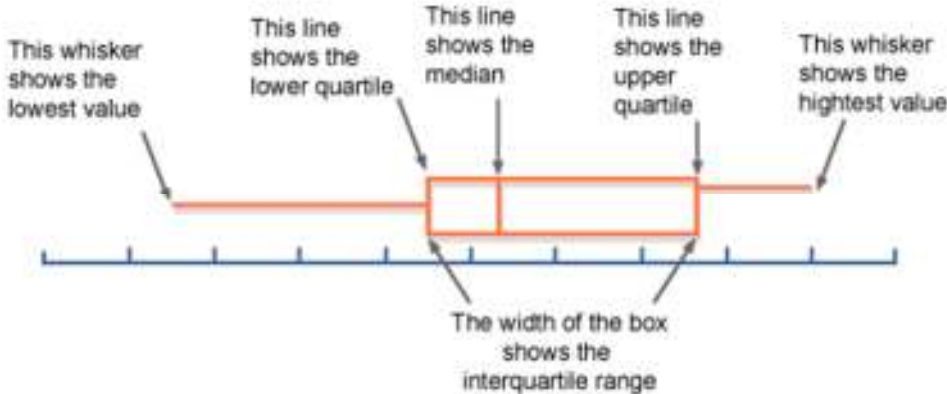
To find the inter-quartile range, we find the data values located at the quarter (lower quartile) and three quarter (upper quartile) points and then find their difference:

We can also use the curve to estimate things like:

How many of my throws scored less than 110?

How many of my throws were over 160?

We can also display this information as a Box & Whisker plot:



So for my darts scores

Median	124
Upper Quartile	144
Lower Quartile	92
Lowest possible score	31
Highest possible score	180

