Averages and Measures of Spread

Individual Data Values

Averages

Suppose we need to compare the lengths of leaves from two different varieties of tree. We can use an AVERAGE such as the MEAN, MEDIAN or MODE to see which has longer leaves.

- MEAN: Add up the values and divide by the number of values

- MEDIAN Arrange the numbers in order. The median is the middle value, or halfway between the two middle values.

- MODE (or Modal Value) The most common value

The mode is very rarely used. The mean and median are commonly used.

Why do we need two different averages?

Often, either is suitable, but if the data contains any extreme values or OUTLIERS then the median is preferable as the mean can be greatly affected by one value. For example:

Suppose a class did a test and the marks were: 16 20 22 25 28 30 32 95

The mean mark is $\frac{268}{8} = 33.5$, but if we use this as the average, then 7 of the 8 students are below average.

In this situation it is better to use the median, which is $\frac{25+28}{2} = 26.5$.

Example:

Lengths of leaves in centimetres.

TREE A: 2 3 5 8 9 10 (11) 12 13 15 17 18 21
(13 values)

Mean = $\frac{144}{13} = 11.08$cm   Median = $\frac{13+1}{2} = 7^{\text{th}}$ value = 11cm

TREE B: 6 8 9 9 10 10 10 11 12 13 15 16 18 19
(14 values)   10.5

Mean = $\frac{166}{15} = 11.86$cm   Median = $\frac{14+1}{2} = $ "7.5$^{\text{th}}$" value = $\frac{10+11}{2} = 10.5$cm

On average the lengths of leaves from the two trees are very similar. The mean length of the leaves from Tree B is slightly greater, but the median length of the leaves from Tree B is slightly less.

Measuring Spread

It may also be important to measure the SPREAD or dispersion of the leaves.

There are 2 statistics we need to know for measuring spread:

- RANGE: The difference between the largest and the smallest values
  eg for tree A, range=$21 - 2 = 19$cm
      for tree B, range= $19 - 6 = 13$cm
  The range can be greatly affected by outliers, so it is not often used.

- INTERQUARTILE RANGE: The difference between the lower quartile (Q_1) and the upper quartile (Q_3).
  The quartiles are 3 values (Q_1, Q_2 and Q_3) which divide the distribution into quarters (so Q_2 is the same as the median.)
  Q_1 is the median of the numbers below Q_2
  Q_3 is the median of the numbers above Q_2

Example:

TREE A: 2 3 5 8 9 10 (11) 12 13 15 17 18 21
          $Q_1 = 6.5$ Q_2 $Q_3 = 16$

$Q_1 = \frac{5+8}{2} = 6.5$
$Q_3 = \frac{15+17}{2} = 16$ $IQR = 16 - 6.5 = 9.5$cm

TREE B: 6 8 9 (9) 10 10 10 11 12 13 (15) 16 18 19
        Q_1 $Q_2 = 10.5$ Q_3

$IQR = 15 - 9 = 6$cm.

From this we see that there is less variation in the size of leaves from Tree B, since the IQR is smaller.

Frequency Tables

Mean

Example: A class of 30 students collected data about the number of children in their families, as follows:

No of children (x)	Frequency (f)	fx
1	4	4
2	11	22
3	9	27
4	5	20
5	1	5
Totals:	30	78

Find the mean number of children per family.

When data is given in a frequency table, we can work out the mean as follows:

- Multiply each data value (x) by its frequency (f)

- Find the total of the f column and the fx column

- $\text{Mean} = \dfrac{\text{Total of fx}}{\text{Total of f}}$

So the mean number of children per family = $\dfrac{78}{30} = 2.6$

Median

To find the median from a frequency table we need to write out (or imagine) all the numbers in a long list, and find the median in the usual way.

Example: For the data above, find the median number of children in a family

1 1 1 1 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 5

15 below ←   → 15 above

Median $= \dfrac{2+3}{2} = 2.5$

Grouped Frequency Tables

Mean

Example: A sample of 150 leaves from a tree had lengths as shown:

Length (cm)	No of Leaves (f)	Midpoint (x)	fx
0–10	18	5	90
10–15	23	12.5	287.5
15–20	45	17.5	787.5
20-25	37	22.5	832.5
25–40	27	$\frac{25+40}{2}=32.5$	877.5
Totals:	150 ~~Totals:~~	NOT NEEDED	2875

Find the mean length of the leaves

The method for finding the mean of a grouped frequency table is the same as for an ordinary frequency table, except that we have to use the MIDPOINT of each class as the value of "x".

So the mean length of the leaves $= \dfrac{2875}{150} = 19.2\text{cm}$

Median and Interquartile Range – Cumulative Frequency Graphs

When data is given in the form of a grouped frequency distribution, we can find the median IQR using a CUMULATIVE FREQUENCY GRAPH.
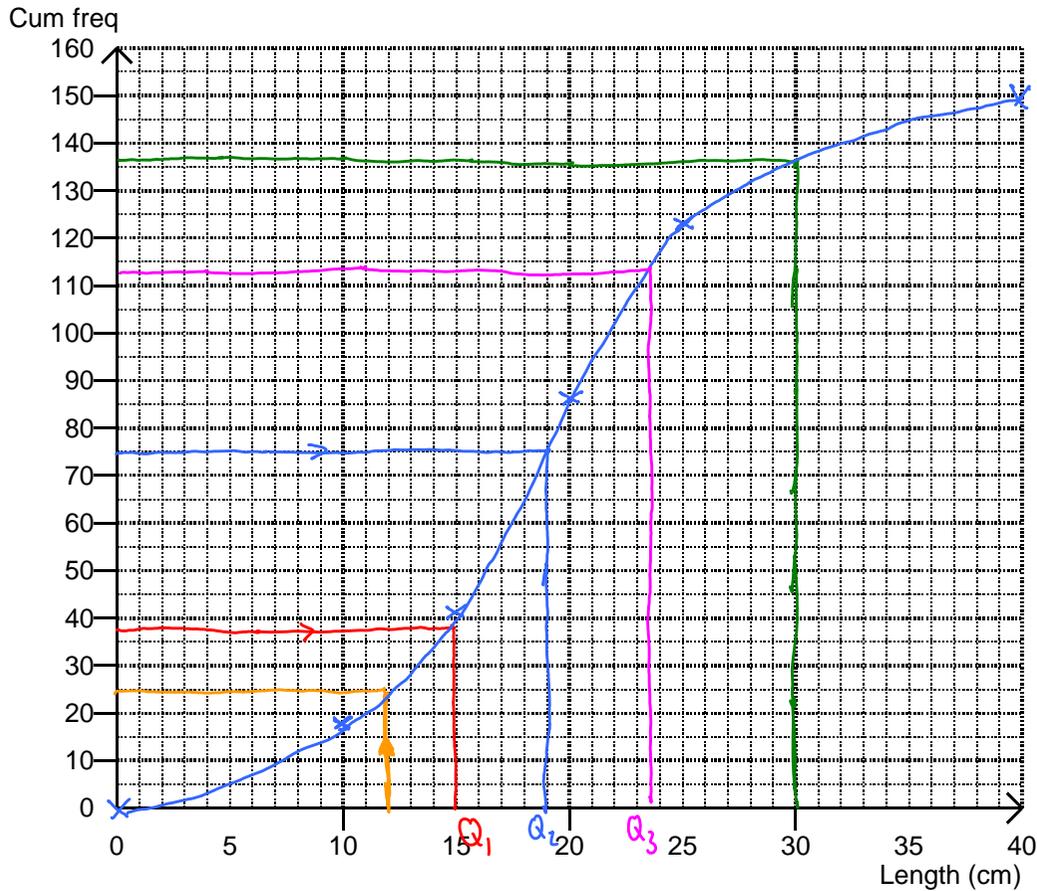
Example: Using the data about leaves given above,

(a) Draw up a cumulative frequency table:

Length (cm)	Cumulative Frequency
0–10	18
0–15	41
0–20	86
0-25	123
0–40	150

(b) Draw a cumulative frequency graph

Always plot the points at the __end__ of the given interval

Cum freq

Length (cm)

When reading off values, draw lines on the graph – the lines are the "working" for these questions

(b) Find the median length of the leaves

- Work out ½ of the total frequency   ½ of 150 = 75
- Draw a line ▮▮▮ at this frequency, and read off down

  Median = 19 cm

(c) Find the IQR of the lengths of the leaves

To find Q₁:

- Work out ¼ of the total frequency   ¼ of 150 = 37.5
- Draw a line across at this frequency, and read off down

  Q_1 = 15 cm

To find Q₃:

- Work out ¾ of the total frequency   ¾ of 150 = 112.5
- Draw a line across at this frequency, and read off down

  Q_3 = 23.5 cm

$IQR = Q_3 - Q_1 = 23.5 - 15$
$= 8.5 cm$

(d) Find how many leaves were under 12cm long

- Draw a line UP at 12cm (and read off across)

  25 leaves

(e) Find how many leaves were over 30 cm long

Draw a line UP at 30cm

136 leaves are LESS THAN 30cm.

150 - 136 = 14 leaves are over 30cm long

A cumulative frequency graph always shows the number of items LESS THAN a certain value.