CALCULATING STATISTICS

Averages

There are 3 types of average :—

Mean : Add up all numbers and divide by how many numbers there are

Median : Put numbers in order. Half of the numbers are below median and half above.

Mode : The most frequent number.

Example The number of children in the families of 10 girls are :—

1 2 2 2 2 3 3 3 4 7

Mean number of children $= \frac{29}{10} = \underline{2.9 \text{ children}}$

Median : There are 10 numbers, so 5 below median and 5 above.

So median is halfway between 5^{th} and 6^{th} numbers

So median $= \frac{2+3}{2} = \frac{5}{2} = \underline{\underline{2.5}}$

Mode : Most frequent number = $\underline{\underline{2}}$

[Notes :

① If there are an odd number of numbers, the median is the actual middle number

e.g 1 2 2 3 4 5 7

Median = $\underline{\underline{3}}$

② Data may be BIMODAL (have 2 modes)

e g 1 2 2 2 3 4 4 4 7

Modes are $\underline{\underline{2}}$ and $\underline{\underline{4}}$]

p 271 Ex 16A Q 1(abc), 2(abc), 3(abc)

Averages from a Frequency Table

Mean The number of children in 30 families is as shown below

No of children (x)	Frequency (f)	fx
1	4	4
2	13	26
3	8	24
4	3	12
5	2	10
	30	76

$$\text{Mean} = \frac{76}{30}$$

$$= \underline{\underline{2.53}}$$ children

Method: Multiply each value by its frequency to get the "fx" column. Divide the total of the "fx" by the total of the "f".

Median Find the median of the family sizes above.

There are 30 families, so 15 families should be below the median, ie the median is between the 15^{th} and 16^{th} number.

If we list the numbers in order,

1, 1, 1, 1, 2, 2, 2, ...

both the 15^{th} and 16^{th} numbers are '2'.

So the median is **2**

Mode of the above family sizes is **2**

(since 2 has the highest frequency)

Mean of a Grouped Frequency Table

This is the same as finding the mean of an ordinary frequency table, except that we use the MIDPOINT of each class to represent all the values in that class.

This means that the answer we get is only an ESTIMATE of the mean - it is impossible to find the actual mean as we do not know every individual value.

Example The distance travelled to school by each student in a class is shown below. Find an estimate of the mean distance travelled.

Distance (miles)	Frequency (f)	Midpoint (x)	fx
0 – 10	15	5	75
10 – 20	10	15	150
20 – 30	3	25	75
30 – 50	2	40	80
Sum	30		380

Mean distance = $\frac{380}{30} = \underline{\underline{12.\dot{6} \text{ miles}}}$

Measuring Spread (or Dispersion)

Two sets of data may have the same average but one may be more spread out

e.g. Two sets of test marks out of 20:

A :	12	13	14	14	15	16	Mean = 14
B :	8	9	14	19	20		Mean = 14

Both sets of marks have a mean of 14 but the second set is much more spread out.

We can measure spread using the range.

Group A : Range = 16 − 12 = 4

Group B : Range = 20 − 8 = 12

Interquartile Range

Sometimes the range can be greatly affected by one unusual result

e.g Two more sets of test marks :

Group C	8	9	9	10	12	14	14	16	18	19	19	20
Group D	1	11	12	13	13	14	16	16	17	18	18	19

Q_1 Q_2 (median) Q_3

(Again the mean of each group is 14)

The RANGE of Group C is $20 - 8 = 12$
and of Group D is $19 - 1 = 18$

This makes it seem that Group D is more spread out. But this is only caused by the single mark of 1; if we ignore this, Group C is actually more spread out.

The Interquartile Range gives a better comparison of the spread :

The 3 quartiles, Q_1 Q_2 and Q_3, divide the values into 4 groups of equal size. (Q_2 is the same as the median.)

So for Group C, $Q_1 = 9.5$
$Q_2 = 14$
$Q_3 = 18.5$

and for Group D, $Q_1 = 12.5$
$Q_2 = 15$
$Q_3 = 17.5$

$\boxed{\text{The Interquartile Range} = Q_3 - Q_1}$

So for Group C, $IQR = 18.5 - 9.5 = \underline{\underline{9}}$
for Group D, $IQR = 17.5 - 12.5 = \underline{\underline{5}}$

This suggests that group C is more spread out which is true.
(Note that if the mark of 1 was changed into 10, this would make no difference to the IQR.)