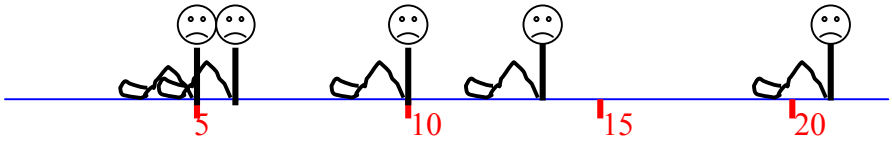
# **STATISTICS** (The mean)

There are at least three types of average - The mean, the median and the mode are all averages.

If we were to work out the average of a set of numbers we tend to go for the mean.

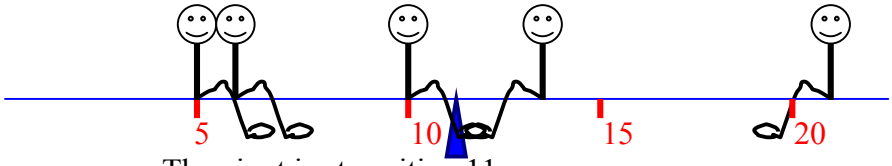
This simply involves adding up all the numbers and dividing by the number of **observations.**

Supposing I observed the following observations: 5, 6, 10, 13, 21 like the positions of 5 people of the same weight sitting along a plank.




To balance the plank we would have to locate a position for a pivot

This is the mean: (5 + 6 + 10 + 13 + 21) divided by 5 which equals 55 ÷ 5 = 11




The pivot is at position 11

You could look at the problem as "5 people have £5, £6, £10, £13, £21".
This is not fair!
So share the money out equally
Put it all in a pot: 5 + 6 + 10 + 13 + 21 = £55
Share it out between the five people
55 ÷ 5 = 11 Each person gets £11

**Remember** The mean is always found by adding all the observations and dividing by the number of observations.

This method is used if we have lots of observations with repetition 5,6,2,3,3,5,2,6,4,3,6,1,5,2,4,..... We still have to add 'em all up and divide by however many there are.
It is usual to look at the observations in a more tidy form - A frequency table:-

| Observation | Frequency | | | Observation x Frequency |
|---|---|---|---|---|
| 1 | 3 | We need to add all the observations 1+1+1+1+2+2 +2+3+.... four ones three twos etc. So.... | Introduce a new column of all observations multiplied by frequency and add the contents of this column. | 3 |
| 2 | 3 | | | 6 |
| 3 | 7 | | | 21 |
| 4 | 5 | | | 20 |
| 5 | 6 | | | 30 |
| 6 | 5 | | | 30 |
| 10 | 1 | | | 10 |
| | | | | 120 |

The number of observations can be found by adding all the frequencies giving 30

The mean is 120 ÷ 30 = 4

If there were lots of observations over a wide range we would be faced with a Grouped Frequency Table.

The raw data may look like this: 25 12 2 9 16 21 4 2 27 .....

| Group or class interval | Tallies | Frequency | Mid point | Frequency × mid point |
|---|---|---|---|---|
| 1 - 5 | 𝍸 \| | 8 | 3 | 24 |
| 6 - 10 | 𝍸 𝍸 \|\|\| | 13 | 8 | 104 |
| 11 - 15 | | 27 | 13 | 351 |
| 16 - 20 | | 31 | 18 | 558 |
| 21 - 25 | | 15 | 23 | 345 |
| 26 - 30 | 𝍸 | 6 | 28 | 168 |
| TOTALS | | 100 | | 1550 |

The data is put into groups

You may have to do a tally chart

This shows how many observations in each group

Pretend each value in a class is the same Use the mid point

This gives the grand total before dividing by the total number of observations

1550 ÷ 100 = 15.5

The mean is 15.5

## Problems

The mean is found by dividing the total by the number of observations

$$\text{Mean} = \frac{TOTAL}{NUMBEROFOBSERVATIONS}$$

So, we can get back to the total by multiplying the mean by the number of observations

If the mean for 25 boys was 60, the total must have been $25 \times 60 = 1500$

If the mean for 30 girls was 65, the total must have been $30 \times 65 = 1950$

The grand total for boys and girls was 3450 and so the combined mean is $3450 \div 55$

(We get the 55 by adding the 25 boys to the 30 girls). The mean is 62.7 (1 dp)

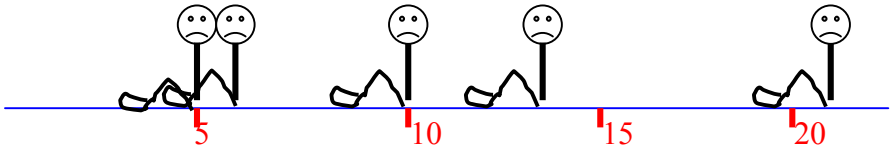
# **STATISTICS** (The median)

The median is the middle observation after the observations are put in order.

If there are an odd number of observations, say 11, the median is the 6th ( it is obvious, but we could have found it by adding 1 to 11 and then dividing by 2).
If there were 67 observations the median is the 34th.

If there are an even number of observations, say 12, the median is half-way between the 6th and 7th. (Adding 1 and dividing by 2 gives us 6.5).

Supposing I observed the following observations: 5, 6, 10, 13, 21 like the positions of 5 people of the same weight sitting along a plank.



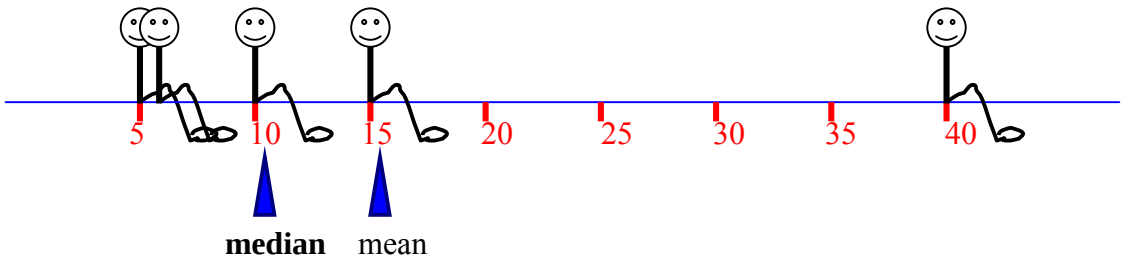

The median is 10.

The median is used as a better measure of location instead of the mean, when there are one or two extreme values.

For example: 5, 6, 10, 15 and 40. The mean would be 15.2

This is not representative of all the observations (the 40 has inflated the mean somewhat).

Using the median gives 10.




**Remember** The median is always found by locating the middle observation or half-way between the middle two.

This method is used if we have lots of observations with repetition 5,6,2,3,3,5,2,6,4,3,6,1,5,2,4,..... We still have to add 'em all up and divide by however many there are.

It is usual to look at the observations in a more tidy form - A frequency table

| Observation | Frequency | Cumulative Frequency |
|---|---|---|
| 1 | 3 | 3 |
| 2 | 3 | 6 |
| 3 | 7 | 13 |
| 4 | 5 | 18 |
| 6 | 5 | 23 |
| 5 | 6 | 29 |
| 10 | 1 | 30 |

There are 30 observations so the median is half way between the 15th and 16th. So....

**Accumulate** the frequencies. Build them up to see how they grow up to 30.

The first **13 observations** are 1s, 2s and 3s

The next five observations are all 4s and take us up to the **18th observation**

The 15th and 16th observations must both be 4s, **so the median is 4.**

If there were lots of observations over a wide range we would be faced with a Grouped Frequency Table. They could be test scores in a test marked out of 30 for 100 pupils.
The raw data may look like this: 25 12 2 9 16 21 4 2 27 .....

| Group or class interval | Tallies | Frequency | "LESS THAN" |
|---|---|---|---|
| 1 - 5 | 𝍸 \| | 8 | 8 less than 5.5 |
| 6 - 10 | 𝍸 𝍸 \| | 13 | 21 less than 10.5 |
| 11 - 15 | | 27 | 48 less than 15.5 |
| 16 - 20 | | 31 | 79 less than 20.5 |
| 21 - 25 | | 15 | 94 less than 25.5 |
| 26 - 30 | 𝍸 | 6 | 100 less than 30.5 |
| TOTALS | | 100 | |

The data is put into groups giving different class intervals.

You may have to do a tally chart

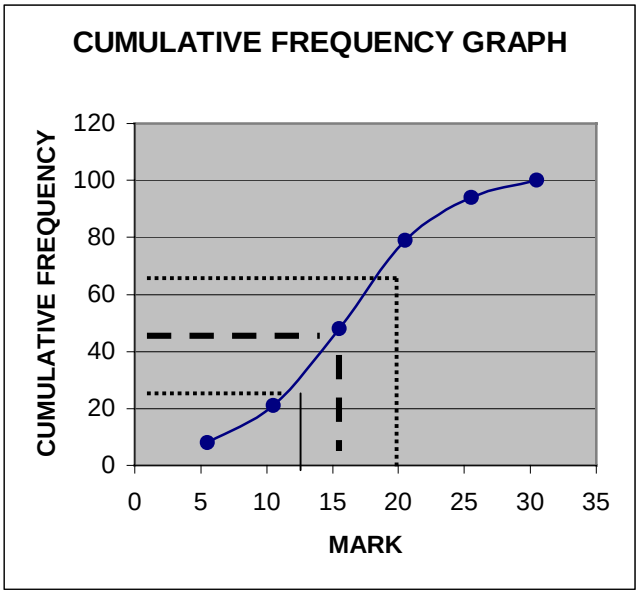
This shows how many observations in each group

Accumulate, or build up the numbers from the previous column to get a **cumulative frequency table** in this column.
You have obtained it by making statements like "8 pupils got less than 5.5".
Always use the right hand value of each class interval. **Bold type.**
If there is a gap between the right hand value in one group and the left hand value in the next group, use the half-way value as we have here.

Data may be put in intervals like
$0 \leq x < \mathbf{10}$ 8
$10 \leq x < \mathbf{20}$ 13
Use
8 less than 10
13 less than 20 etc.

| 5.5 | 10.5 | 15.5 | 20.5 | 25.5 | 30.5 |
|---|---|---|---|---|---|
| 8 | 21 | 48 | 79 | 94 | 100 |

From a cumulative frequency table comes a cumulative frequency graph.
After this has been constructed, we can READ OFF the median.
We can also read off the **upper and lower quartiles.**




The Median is read off by going across from 50.
We can say that 50 pupils (half the pupils) got less than 16. **The median is 16.**
75 pupils got less than 19 (from the graph). **So the upper quartile is 19.**
25 pupils (a quarter of all pupils) got less than 12. **So the lower quartile is 12.**

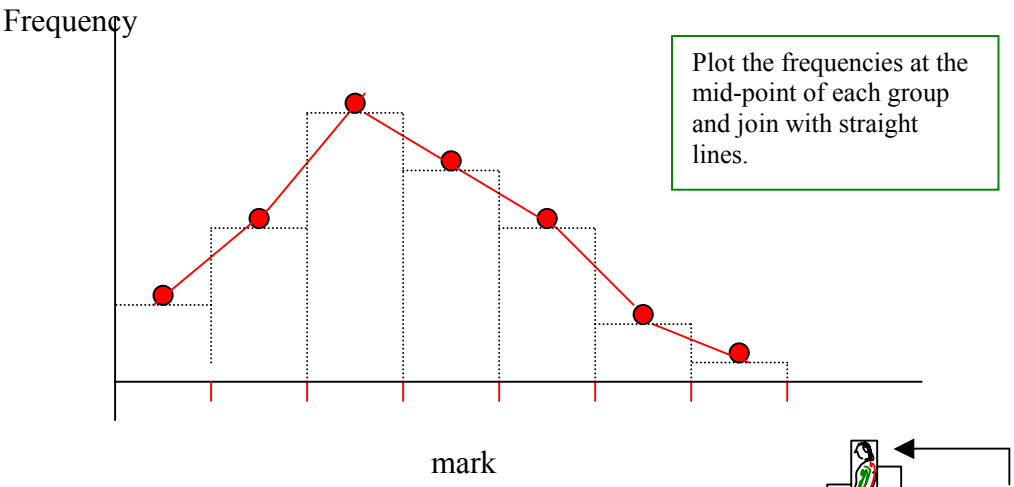
A measure of spread is the **inter-quartile range**, which is the
Upper Quartile minus Lower Quartile. This is 19 - 12 = **7.**

The simplest measure of spread is the Range, which is simply the highest observation minus the lowest.

Go back to the top of the sheet and check the range for the observations 5, 6, 10, 13, 21.

The range is 21 - 5 = **16**

An exam question may ask you to construct a frequency polygon:




But what about the mode? Click here to find the factsheet on the mode.....