TJP TOP TIPS

FOR

IGCSE STATS & PROBABILITY

IGCSE STATS & PROBABILITY

First, some important words; know what they mean (get someone to test you):

Mean – the sum of the data values divided by the number of items.
The mean of 1, 2, 2, 3, 4, 6 is (1+2+2+3+4+6) ÷ 6 = 18 ÷ 6 = **3**

Median – the middle data value when all the numbers are listed in order.
The median of 1, 2, 2, 3, 4, 6 is (2+3) ÷ 2 = 5 ÷ 2 = **2.5**

Mode – the most common data value.
The mode of 1, 2, 2, 3, 4, 6 is **2**

Range – the largest data value minus the smallest data value.
The range of 1, 2, 2, 3, 4, 6 is 6 – 1 = **5**

Lower Quartile – the data value that is one quarter of the way up from the lowest value.
The lower quartile of 1, 2, 2, 3, 4, 6 is **2**

Upper Quartile – the data value that is three quarters of the way up from the lowest value.
The upper quartile of 1, 2, 2, 3, 4, 6 is **4**

Interquartile Range = Upper Quartile – Lower Quartile.
The interquartile range of 1, 2, 2, 3, 4, 6 is 4 – 2 = **2**

Frequency – the number of times that an event occurs.
If a roll a die 20 times and get the number three on four occasions, the frequency is 4.

Cumulative Frequency – the running total of the frequency values.
These running totals are often then plotted as an S-shaped curve.
We can then read off the median and the quartiles.

Grouped Frequency Table – a table where data values are grouped in 'bins'.
A survey might record the number of people aged 0-4, 5-9, 10-19, etc.

Class Width – the width of a 'bin' in a grouped frequency table.

Frequency Density – the frequency divided by the class width, used in histograms.

Histogram – a chart plotting frequency density on the y axis with classes along the x axis.
It is like a bar chart but corrected for the misleading effect of having differing bin widths.

Probability – the chance of an event happening.
Probability is always a number between 0 (impossible) and 1 (certain).

Outcome – the result of an event.
If a coin is tossed, the possible outcomes are Heads and Tails.

Expected Number – the number of times you would expect an outcome to occur.
If I roll a die 100 times and the chance of rolling a '3' is 0.2, I expect to get 20 '3's.

Mutually Exclusive Events – events that cannot both/all happen at the same time.
If you roll a die, getting a 1 or getting a 2 are mutually exclusive (can't happen together).
But having a beard or wearing glasses are **not** mutually exclusive (can happen together).

Independent Events – events that do not affect one another's outcomes.
If a coin is tossed twice, the outcomes are independent – there is no 'memory effect'.
But if sweets are picked from a bag and eaten, successive events are **not** independent.

Tree diagram – a way of showing all the possible outcomes when two or more events occur, along with their probabilities. The diagram branches repeatedly, like a tree (on its side).

IGCSE STATS & PROBABILITY

PROBABILITY

▷ **Probability** means **the chance of an event happening**.
It is always given as a **number between 0 and 1**, with 0 = impossible and 1 = certain.

You can give probabilities as **decimals** or **fractions**, but never as percentages.

▷ The **outcome** is the **result of an event**.

If you consider **all** the possible outcomes of an event, their **probabilities must add to 1**. This is because it is **certain** that one of the outcomes will happen (we just don't know which one).

▶ **SKILL: Complete a table of probabilities listing all outcomes.**

Q: If a spinner is numbered 1, 2, 3, 4 and 5 and it lands on these numbers with the following probabilities, complete the table.

Spinner	1	2	3	4	5
Probability	0.1	0.2	0.1	0.05	

A: Since the probabilities must add up to 1, the missing number is

1 – 0.1 – 0.2 – 0.1 – 0.05 = **0.55**.

If the probability of an event happening is p, the probability of it **not** happening is **1 – p**. This is because something either happens or it doesn't – there's no other option.

▷ Events are **mutually exclusive** if they can't both/all happen at once.
An example is getting Heads or getting Tails when you flip a coin; you can't get both.

We combine the probabilities of mutually exclusive events by **adding them**.

TJP TOP TIP: **Remember ADD-OR** (as in 'we add-or statistics').

▶ **SKILL: Combine probabilities using the OR rule.**

Q: For the spinner mentioned in the previous question, find the probability of getting a 2 or a 3.

A: Getting a 2 and getting a 3 are mutually exclusive, so we **add** the probabilities.

Prob(getting 2 or 3) = 0.2 + 0.1 = **0.3**.

Q: A pupil is picked at random from a class. The probability of picking someone wearing glasses is 0.3 and the probability of picking a girl is 0.5. Explain why the probability of picking a girl or someone wearing glasses is **not** 0.8.

A: Wearing glasses and being a girl are not mutually exclusive; there could be one or more girls who wear glasses. So we can't just add the probabilities; we'd be counting any girls with glasses twice.

▷ Two events A and B are **independent** if they have no effect on one another.
In this case, **Prob(A and B) = Prob(A) × Prob(B)**

This is the **multiplication law for independent events**.

▶ **SKILL: Combine probabilities using the AND rule.**

Q: The probability of spinning a 2 is 0.2 and the probability of picking a red ball out of a bag is 0.3. Find the probability of spinning a 2 **and** picking a red ball.

A: These are independent events; they don't affect each other. So we multiply:

Prob(2 on spinner and red ball) = 0.2 × 0.3 = **0.06**.

▶ **SKILL: List all possible outcomes to solve probability questions.**

To **list all the possible outcomes**, you need to be **systematic**.
Here are two common examples:

Q: Three fair coins are tossed. List all the possible outcomes.
Hence find the probability of getting three tails.

A: HHH HHT HTH HTT THH THT TTH TTT

There are 8 possible outcomes (spot the pattern...)
So the probability of getting TTT is **1/8**.

Q: Two spinners are marked with numbers from 1 to 4.
Draw a table to show all the possible outcomes.
If each number is equally likely, find the probability of getting a total of 6.

A:

	1	2	3	4
1	.	.	.	.
2	.	.	.	x
3	.	.	x	.
4	.	x	.	.

There are 16 possible outcomes (4 × 4), of which the 3 marked ones add up to 6.
So the probability of getting a total of 6 is **3/16**.

▷ The **expected number** or **expected frequency** is the number of times you would expect an event to happen.

Simply **multiply the probability by the number of trials**.

▶ **SKILL: Find the expected frequency.**

Q: The probability of getting a '3' when rolling a die is 0.15. How many times would you expect to get a 3 is you roll it 200 times?

A: 0.15 × 200 = **30 times**.

SIMPLE TREE DIAGRAMS

▷ Probability problems are often tackled using a **tree diagram**. This looks like a tree on its side, where the branches show the different outcomes along with their probabilities.

The first set of branches from the left correspond to the first event to happen.
The next set of branches coming off these correspond to the second event, etc.

TJP TOP TIP: Remember **MAAD** for tree diagrams: **Multiply Across, Add Down**. This refers to how to combine the probabilities marked on the branches.

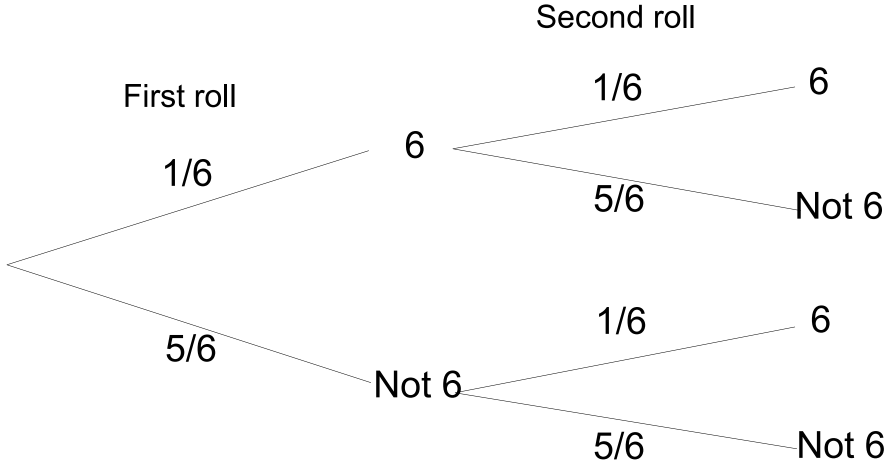
▶ **SKILL: Construct a tree diagram and use it to answer a probability question.**

Q: The probability of Esmee rolling a six on a fair die is 1/6.

a) Draw a tree diagram to show the possible outcomes when she rolls this die twice.

b) Use this tree diagram to find the probability of Esmee rolling:
i) two sixes
ii) exactly one six

A: a) Draw the tree diagram.



b) i) P(two sixes) = $\frac{1}{6}\times\frac{1}{6} = \frac{1}{36}$

ii) P(exactly one six) = $\frac{1}{6}\times\frac{5}{6} + \frac{5}{6}\times\frac{1}{6} = \frac{5}{36}+\frac{5}{36} = \frac{10}{36} = \frac{5}{18}$

[This could be '6' followed by 'not 6', or 'not 6' followed by '6'.]

Note: Tree diagrams can have three or more branches at each stage, and three or more stages. But if you end up drawing hundreds of branches, there's probably an easier way...

TREE DIAGRAMS FOR CONDITIONAL PROBABILITY

▷ Read the question carefully to see if it says objects are picked **without replacement**. If so, the probabilities will **change** after each selection. This is an example of **conditional probability**, where the probabilities depend on what has already happened.

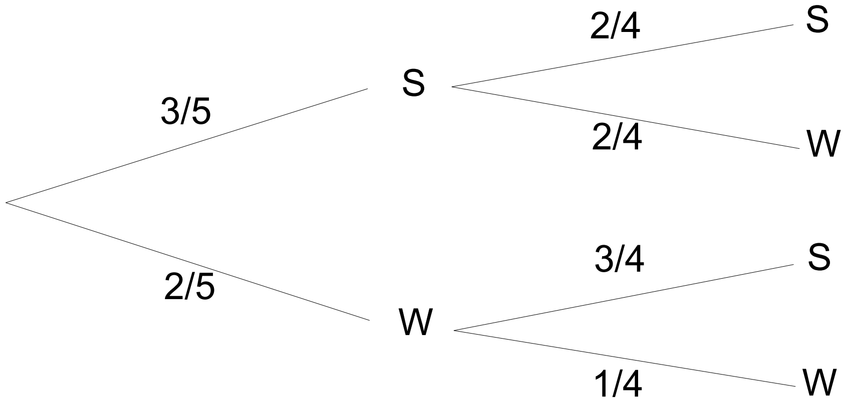
▶ **SKILL: Use a tree diagram to answer a conditional probability question.**

Q: A bag contains 3 stoats and 2 weasels; two animals are then picked at random without replacement.

Use a tree diagram to calculate:

a) Prob(2 stoats)
b) Prob(1 stoat and 1 weasel, in either order)
c) Prob(at least one weasel)

A: First, draw the tree diagram.



[The top right probability is 2/4 because if we remove a stoat, there are now only 2 stoats left out of 4 animals still in the bag.]

Now use the tree diagram to answer the questions.

a) P(2 stoats) = $\frac{3}{5}\times\frac{2}{4} = \frac{6}{20} = \frac{3}{10}$

b) P(1 stoat and 1 weasel) = $\frac{3}{5}\times\frac{2}{4} + \frac{2}{5}\times\frac{3}{4} = \frac{6}{20}+\frac{6}{20} = \frac{12}{20} = \frac{3}{5}$

[Note: this could be Stoat then Weasel, or Weasel then Stoat.]

c) P(at least one weasel) = 1 – P(no weasels) = $1 - \frac{3}{5}\times\frac{2}{4} = 1 - \frac{6}{20} = \frac{14}{20} = \frac{7}{10}$

TJP TOP TIP: If it's a '**one of each**' question, remember that there is **more than one way** to get this on your tree diagram. For example, 'A then B' or 'B then A'.

MEAN, MEDIAN, QUARTILES, MODE AND RANGE FROM A LIST

▷ **Mean** = **total of all data values ÷ total number of items**.
It's sensitive to any 'freak results' that are unusually high or low.

▷ **Median** = **middle data value** when **sorted into increasing order**.
If there are two middle data values, take their mean.
The median is not sensitive to 'freak results'.

▷ **Lower Quartile** = the **median** of the **bottom half** of the list.
It's the value ¼ of the way up the list.

▷ **Upper Quartile** = the **median** of the **top half** of the list.
It's the value ¾ of the way up the list.

▷ **Interquartile Range** = **upper quartile – lower quartile**.
It indicates how spread out the data values are.

▷ **Mode** = **most common** data value.
If there are two most common values, the distribution is **bimodal**.

▷ **Range** = **highest value – lowest value**.

▶ **SKILL: Find the mean, median, quartiles, mode and range from a list of data.**

Q: Find the quartiles and median of 4, 5, 6, 8, 10, 13, 15, 16, 19.

A: The median is the middle number, **10**.
The lower quartile is the median of the bottom half 4, 5, 6, 8 which is **5.5**.
The upper quartile is the median of the top half 13, 15, 16, 19 which is **15.5**.
The interquartile range = 15.5 – 5.5 = **10**.
[Note: don't include the middle number 10 in the bottom or top half of the list.]

Q: Find the mean, median, quartiles, mode and range of 1, 3, 3, 3, 4, 5, 6, 7, 10, 11.

A: Mean = (1+3+3+3+4+5+6+7+10+11) ÷ 10 = **5.3**
Median = (4+5) ÷ 2 = **4.5**
Lower Quartile = **3**
Upper Quartile = **7**
Interquartile Range = 7 – 3 = **4**
Mode = **3**
Range = 11 – 1 = **10**

Q: Three numbers are 6, x and $2x$ (with $x>6$).
Show that the mean is 2 greater than the median.

A: The median is x.
The mean is $(6 + x + 2x) \div 3 = (6 + 3x) \div 3 = 2 + x$.

So the mean $(x + 2)$ is 2 greater than the median (x).

MEAN, MEDIAN, MODE AND RANGE FROM A TABLE

▷ To work out these quantities if the data values are listed in a **table**, do the following.

▶ **SKILL: Find the mean, median, mode and range from a table.**

Q: Find the mean, median, mode and range of the following data.

length	frequency
1	3
2	2
3	6
4	5
5	4

A: Add a column to the table for length × frequency.

length	frequency	length × frequency
1	3	3
2	2	4
3	6	18
4	5	20
5	4	20
	20	65

Mean = (1×3 + 2×2 + 3×6 + 4×5 + 5×4) ÷ (3 + 2 + 5 + 6 + 4) = 65 ÷ 20 = **3.25**.
[Use MAAD – Multiply Across, Add Down – on the table.]

Median = the length category containing the middle (two) items when listed in order.
There are 20 items altogether, so we need the 10^{th} and 11^{th} items.
Count down from the top:
Items 1-3 have length 1;
Items 4-5 have length 2;
Items 6-11 have length 3.
So the median is **3**.

Mode is the length category with the most (the biggest frequency) = **3**.

Range = biggest length – smallest length = 5 – 1 = **4**.

TJP TOP TIP: To find the **position of the median**, do the **mean of the first and last positions**. So in a list of 123 items, the median is at position (1 + 123) ÷ 2 = 62.

If this position is 'X and a half', the two middle numbers are at X and X+1.

MEAN, MEDIAN CLASS, MODAL CLASS AND RANGE FROM A GROUPED TABLE

▷ If you need to work out these quantities from a **grouped table** (where data values are grouped into 'bins' so we don't know their exact values any more):

- Find the **mean** of **grouped** data using the **middle value** of each class.
- Find the **class containing the median** (see previous page).
- The **modal class** is the group or class with the most (the highest frequency).
- The **range** is the upper limit of the highest group (class) minus the lower limit of the lowest group (class).

▶ **SKILL: Find the mean, median class, modal class and range from a grouped table.**

Q: Find the mean, median class, modal class and range of the following grouped data.

height (cm)	frequency
101-120	1
121-130	3
131-140	5
141-150	7
151-160	4
161-170	2
171-190	1

A: Add two columns to the table, for the midpoint of the class and for freq × midpoint.

height (cm)	frequency	midpoint	freq × midpoint
101-120	1	110.5	110.5
121-130	3	125.5	376.5
131-140	5	135.5	677.5
141-150	7	145.5	1018.5
151-160	4	155.5	622
161-170	2	165.5	331
171-190	1	180.5	180.5
	23		3316.5

Mean = 3316.5 ÷ 23 = **144** (midpoint MAADness...)
This is just an **estimate** because we don't know the exact data values.

Median class = the class containing the middle item, no. 12 in the list.
The 12^{th} item occurs in the **141-150 class**.

Modal class = **141-150** because it has the highest frequency.

Range = 190 – 101 = **89**.

CUMULATIVE FREQUENCY

▷ To find the **median** and **quartiles** accurately from **grouped data**, it is helpful to draw a **cumulative frequency graph** and read off the values from it.

In a **cumulative frequency** graph, we find the **running total** of the frequencies.
We then plot this against the **upper end** of each class interval to show how many data values there are **up to** a particular limit.

▶ **SKILL: Plot a cumulative frequency curve and find the median and quartiles.**

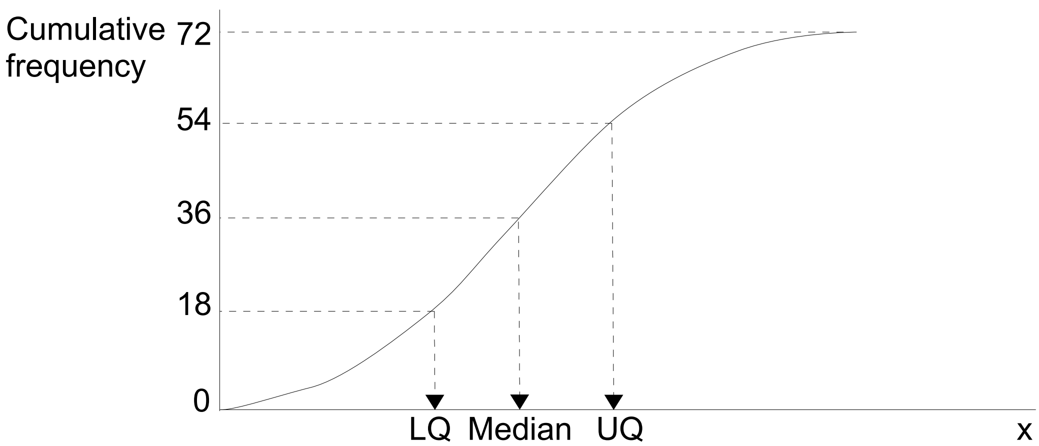
Q: Plot a cumulative frequency curve from this table and find the median and quartiles.

Value x	Frequency f
0-20	12
20-30	20
30-60	15
60-100	25

A: First work out the cumulative frequency (running total).

Value x	Frequency f	Cumulative Freq
0-20	12	12
20-30	20	32
30-60	15	47
60-100	25	72

We now plot points at (0, 0) (20, 12) (30, 32) (60, 47) (100, 72)
and draw a smooth curve **through** these points to give that classic S-shaped curve.



Then we can read off the **Median** and the **Upper and Lower Quartiles** off the x-axis.
(No actual numbers here this time, but there will be in the exam.)

The **Interquartile Range** = Upper Quartile – Lower Quartile.

HISTOGRAMS

▷ **Histograms** are a bit like bar charts except that we must plot **frequency density = frequency ÷ class width**, not the frequency.

This means that the **area** of each bar is equal to the **frequency**.
Also remember that the **bars should not have gaps between them**.

TJP TOP TIP: **Two utterly essentially vitally important facts for histograms**:
- Always plot **frequency density (= frequency ÷ width)**.
  Hint: think alphabetically... frequency comes before width.
- The **frequency** is given by the **area** of each bar, not the height.

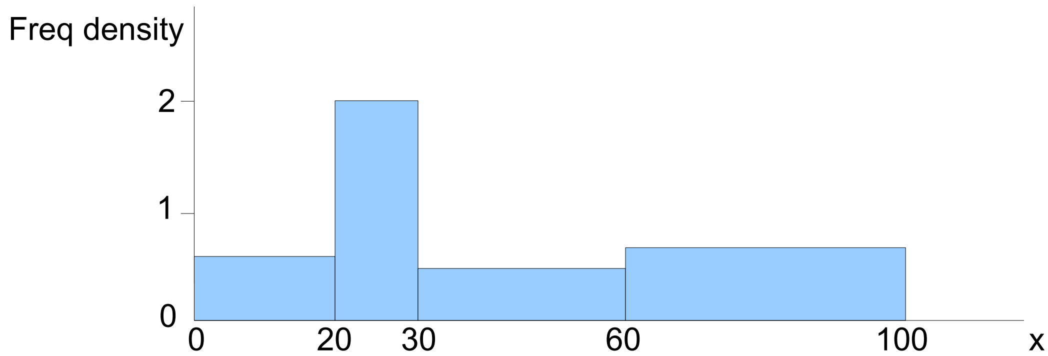
▶ **SKILL: Plot a histogram from a grouped frequency table.**

Q: Display the following data on a histogram.

Value x	Frequency f
0-20	12
20-30	20
30-60	15
60-100	25

A: We **must** begin by working out the **frequency density** = frequency ÷ width.
Draw an extra column (or row) on the table if necessary.

Value x	Frequency f	Freq density
0-20	12	12÷20 = 0.6
20-30	20	20÷10 = 2.0
30-60	15	15÷30 = 0.5
60-100	25	25÷40 = 0.625



TJP TOP TIP: Sometimes they give us a **bar** that is **already drawn** on the histogram as well as featuring in the table. Use this known bar to **label the y-axis** correctly.

To **read values off** the histogram, remember that the **area** of a bar gives the **frequency**.

IGCSE STATS & PROBABILITY

CONTENTS