Advanced Statistics

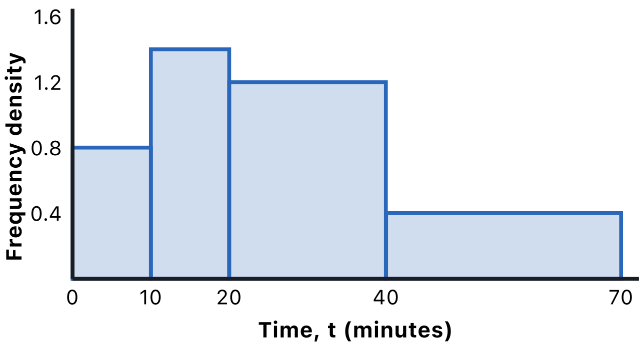
YEAR 11 · UNIT 21 KNOWLEDGE ORGANISER

HISTOGRAMS — FREQUENCY DENSITY

$$\textbf{frequency density} = \frac{\textbf{frequency}}{\textbf{class width}}$$

- For **grouped continuous** data with **unequal** class widths.
- The **area** of each bar is proportional to the **frequency** — not the height.
- Bars **touch**; label the axis where they join.

DRAWING A HISTOGRAM

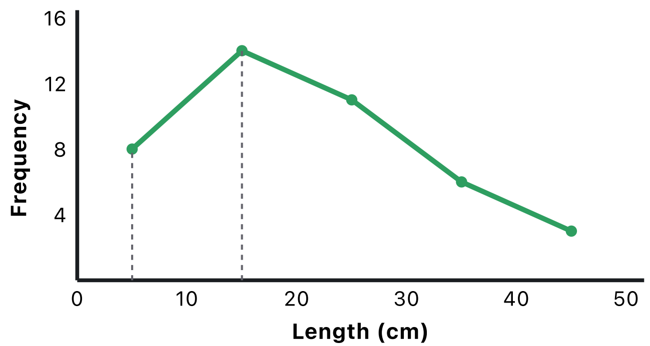


WORKING TABLE — ALWAYS ADD A COLUMN

Time, t (min)	Freq	Width	FD
$0 < t \le 10$	8	10	0.8
$10 < t \le 20$	14	10	1.4
$20 < t \le 40$	24	20	1.2
$40 < t \le 70$	12	30	0.4

Reading **back**: frequency = FD × width, so 1.2 × 20 = **24**.

FREQUENCY POLYGONS



Plot at the **midpoint** of each class; join with **straight lines**.

ESTIMATED MEAN FROM A TABLE

Estimate the mean length.

Midpoints: 5, 15, 25, 35, 45

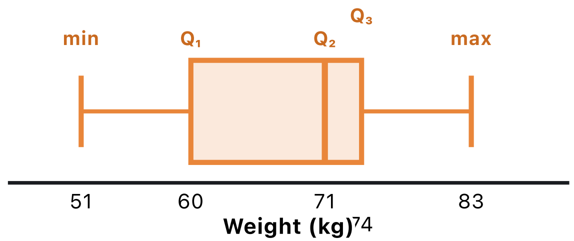
$\sum fx = 5(8)+15(14)+25(11)+35(6)+45(3) = 870$

$\sum f = 8 + 14 + 11 + 6 + 3 = 42$

Mean = 870 ÷ 42 = **20.7 cm (3 s.f.)**

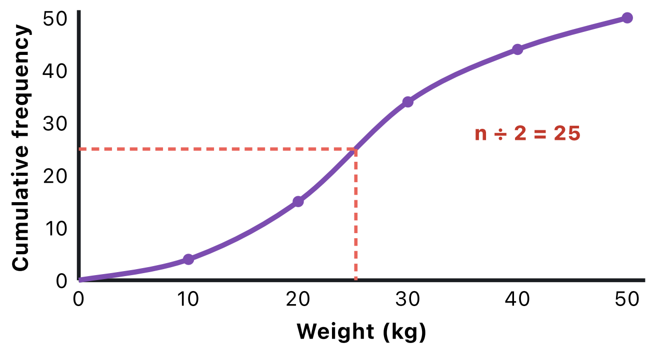
It is an **estimate** — the original values are lost in the grouping.

BOX PLOTS — FIVE VALUES



IQR = 74 − 60 = **14 kg** · Range = 83 − 51 = **32 kg**

CUMULATIVE FREQUENCY GRAPHS



Plot at the **upper class boundary**; join with a **smooth curve**.

READING OFF A CF GRAPH (N = 50)

- **Median (Q_2)** — read across from n ÷ 2 = 25 → **25 kg**.
- **Lower quartile (Q_1)** — from n ÷ 4 = 12.5 → **18 kg**.
- **Upper quartile (Q_3)** — from 3n ÷ 4 = 37.5 → **34 kg**.
- **IQR** = $Q_3 - Q_1$ = 34 − 18 = **16 kg**.
- An **outlier** is more than **1.5 × IQR** beyond a quartile.

WHICH CHART FOR WHICH DATA?

Chart	Use it for
Bar chart	Categorical — bars do **not** touch
Histogram	Grouped continuous, unequal widths
Cum. freq.	Medians & quartiles, n > 30

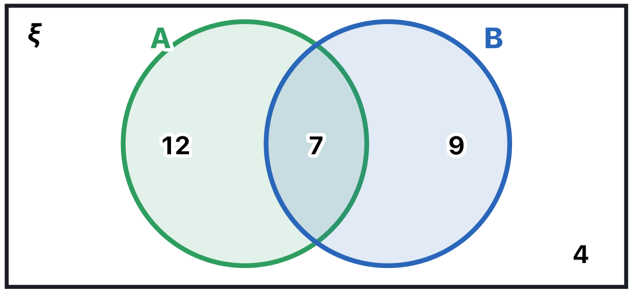
⚠ COMMON EXAM MISTAKES (recurring errors flagged in examiners' reports)

- ✕ Plotting **frequency** up the axis of a histogram instead of **frequency density**.
- ✕ Leaving **gaps** between histogram bars, or drawing all bars the same width.
- ✕ Forgetting to **multiply back** (frequency = FD × width) when reading a histogram.
- ✕ Plotting a cumulative frequency point at the **midpoint** instead of the **upper boundary**.
- ✕ Using **n + 1** instead of **n** when reading quartiles off a cumulative frequency graph.

2 Advanced Probability

YEAR 11 · UNIT 21 KNOWLEDGE ORGANISER

VENN DIAGRAMS & SET NOTATION



$n(\xi) = 12 + 7 + 9 + 4 = \mathbf{32}$

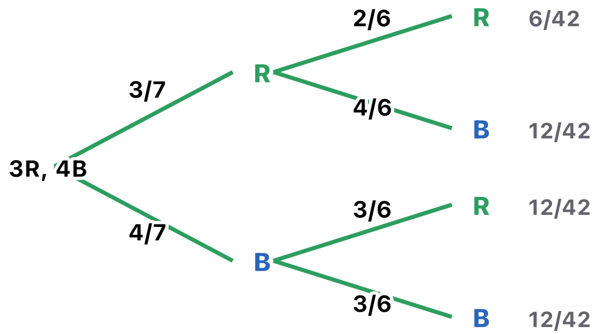
READING THE SYMBOLS

Symbol	Means	Here
A ∩ B	in A **and** B (overlap)	7
A ∪ B	in A **or** B or both	28
A′	**not** in A	13
A	how many are in A	19
∈	is a member of	—
∅	the empty set	—

CONDITIONAL PROBABILITY

- Read **P(A | B)** as "the probability of A **given that** B has happened".
- B has already happened, so you only count the outcomes **inside B** — B becomes the new total.
- Above: $P(A \mid B) = \frac{7}{16}$, **not** $\frac{7}{32}$.

TREE DIAGRAMS — WITHOUT REPLACEMENT



Second-stage fractions have denominator **6** — one ball has gone.

ALONG × · DOWN +

- **Multiply** along the branches for one outcome.
- **Add** the branch ends for "or" outcomes.
- All the end probabilities must total **1**.
- If the question does **not** say, presume **without replacement**.
- Without replacement: the total drops by 1 **and** that colour drops by 1.

"AT LEAST ONE" — USE THE COMPLEMENT

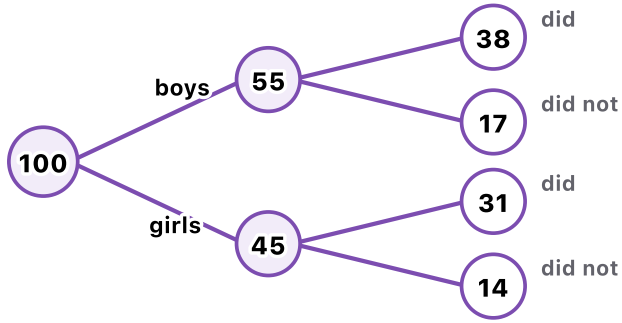
3 red, 4 blue. Two taken. P(at least one red)?

$P(\text{no red}) = \frac{4}{7} \times \frac{3}{6} = \frac{12}{42} = \frac{2}{7}$

$P(\text{at least one red}) = 1 - \frac{2}{7} = \frac{5}{7}$

Far quicker than adding the three separate branches.

FREQUENCY TREES



Whole numbers, not fractions. $P(\text{boy did not}) = \frac{17}{55}$

ALGEBRAIC PROBABILITY

n sweets, 6 orange. Two eaten. P(both orange) = $\frac{1}{3}$. Find n.

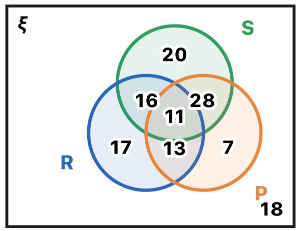
$\frac{6}{n} \times \frac{5}{n-1} = \frac{1}{3} \rightarrow \frac{30}{n(n-1)} = \frac{1}{3}$

$90 = n(n-1) = n^2 - n$

$n^2 - n - 90 = 0 \rightarrow (n - 10)(n + 9) = 0$

n = 10 — reject n = −9, you cannot have −9 sweets.

THREE SETS — START IN THE MIDDLE



130 pupils · S, R, P

All three → **11** in the centre

S and R = 27, so 27 − 11 = **16**

S = 75, so 75 − 16 − 11 − 28 = **20**

In a circle = 112, so outside = **18**

Always fill the centre first.

⚠ COMMON EXAM MISTAKES (recurring errors flagged in examiners' reports)

- ✕ Keeping the **same denominator** on the second stage of a without-replacement tree.
- ✕ Reducing the total but forgetting to reduce that **colour's** count as well.
- ✕ Writing the **overlap** number as the whole of a set when filling a Venn diagram.
- ✕ Starting a three-set Venn diagram at the outside — always start in the **middle**.
- ✕ Dividing by the **total** instead of by B in a conditional probability.
- ✕ Adding probabilities along a branch, or multiplying down the ends.
- ✕ Answering "at least one" the long way and missing a case.
- ✕ Keeping the **negative** root when solving an algebraic probability equation.

Revise it: draw the tree for 3 red and 4 blue, then find P(at least one blue).

3 Combinations & Permutations

YEAR 11 · UNIT 21 KNOWLEDGE ORGANISER

COUNTING BY MULTIPLYING

- Make a **separate choice** at each stage → **multiply** the number of options.
- One from each group: multiply **all** the group sizes.
- "**Or**" between whole scenarios → work each out, then **add**.

total = $\text{choices}_1 \times \text{choices}_2 \times \text{choices}_3$ …

WORKED EXAMPLE — ONE OF EACH

25 teachers, 250 girls, 200 boys. Choose one of each.

25 × 250 × 200

= 6250 × 200

= **1 250 000 ways**

WORKED EXAMPLE — FINDING A MISSING GROUP

10 flavours, 7 toppings, 3 flakes, s sauces → 1260 ways. Find s.

10 × 7 × 3 × s = 1260

210s = 1260

s = 1260 ÷ 210 = **6 sauces**

WORKED EXAMPLE — "OR" SCENARIOS

7 starters, 9 mains, 3 desserts. Choose starter+main, main+dessert, or all three.

Starter + main: 7 × 9 = 63

Main + dessert: 9 × 3 = 27

All three: 7 × 9 × 3 = 189

Total = 63 + 27 + 189 = **279 choices**

CAN ITEMS REPEAT?

- **Repeats allowed** — the count stays the same every time.
- **No repeats** — the count goes **down by one** each time.
- Passwords and lock codes usually **allow** repeats.
- Picking different people or objects usually does **not**.

WORKED EXAMPLE — REPEATS ALLOWED

A password is 3 lowercase letters, repeats allowed.

26 × 26 × 26

= 26^3

= **17 576 passwords**

WORKED EXAMPLE — NO REPEATS, ORDER MATTERS

73 girls: pick a Head Girl and a different Deputy.

73 choices, then 72 left

73 × 72 = **5256 ways**

Head Girl then Deputy is **different** from Deputy then Head Girl, so **do not divide**.

ORDERED OR NOT?

Wording	Order?	Do
1st, 2nd, 3rd	Yes	Multiply
Captain & vice	Yes	Multiply
A code / a number	Yes	Multiply
A team of 2	No	Then divide
"combinations of"	No	Then divide

WHEN ORDER DOES NOT MATTER

- Count as if order **did** matter, then **divide** by the number of ways your picks could be arranged.
- Choosing **2**: divide by 2 × 1 = **2**.
- Choosing **3**: divide by 3 × 2 × 1 = **6**.
- Choosing **4**: divide by 4 × 3 × 2 × 1 = **24**.

WORKED EXAMPLE — A TEAM OF TWO

20 players. Pick 2 for a training camp.

Ordered: 20 × 19 = 380

Each pair counted **twice** (AB and BA)

380 ÷ 2 = **190 combinations**

WORKED EXAMPLE — CHOOSING THREE

10 books. Pick 3 to take on holiday.

Ordered: 10 × 9 × 8 = 720

Each set of 3 counted 3 × 2 × 1 = 6 times

720 ÷ 6 = **120 combinations**

WORKED EXAMPLE — ARRANGING DIGITS

Cards 3, 5, 6, 7 make 4-digit numbers, each card used once. How many are odd?

Last digit must be 3, 5 or 7 → **3** choices

The other 3 cards fill 3 places: 3 × 2 × 1 = 6

3 × 6 = **18 odd numbers**

⚠ COMMON EXAM MISTAKES (recurring errors flagged in examiners' reports)

- ✕ **Adding** the group sizes instead of multiplying them.
- ✕ Keeping the count the same when repeats are **not** allowed.
- ✕ Forgetting to **divide** when order does not matter, so every team is counted twice.
- ✕ Dividing when order **does** matter, e.g. for Head Girl and Deputy.
- ✕ Fixing the restricted digit **last** — always deal with the restriction **first**.
- ✕ Missing the "**or**" in a question and giving only one of the scenarios.
- ✕ Writing 26 × 3 instead of 26^3 for a three-letter password.
- ✕ Rounding a counting answer — it must be a **whole number**.

4 Capture–Recapture

YEAR 11 · UNIT 21 KNOWLEDGE ORGANISER

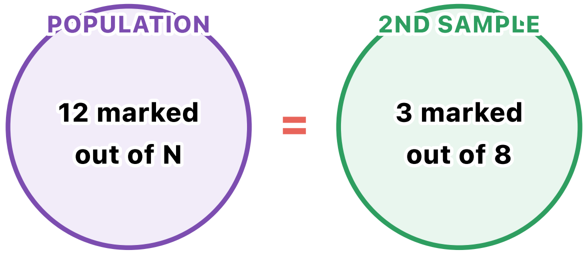
THE IDEA

- Used to **estimate** a population that is too large to count — fish in a lake, deer in a park.
- Devised by the Danish marine biologist **Carl Petersen** in 1896.
- The **proportion marked** in the second sample should match the **proportion marked** in the whole population.

THE METHOD

1. Take a **first sample** and count it.
2. **Mark** every item in that sample.
3. **Return** them and let them **mix thoroughly**.
4. Take a **second sample** and count it.
5. Count how many in it are **marked**.
6. Set the two **proportions equal** and solve.

THE TWO PROPORTIONS



THE EQUATION

$$\frac{\text{marked in 2nd sample}}{\text{size of 2nd sample}} = \frac{\text{total marked}}{\text{population}}$$

- Both fractions are the **proportion that is marked**.
- The **total marked** is the size of the **first** sample.
- Put the **unknown N** on the bottom right, then cross-multiply.

WORKED EXAMPLE

12 fish are caught and marked. The next day 8 fish are caught, of which 3 are marked. Estimate the population.

$\frac{3}{8} = \frac{12}{N}$

3N = 8 × 12

3N = 96

N = 96 ÷ 3 = **32 fish**

Check: $\frac{12}{32} = \frac{3}{8}$ ✓

YOUR TURN

18 fish are marked. The next day 7 are caught, of which 2 are marked.

$\frac{2}{7} = \frac{18}{N}$

2N = 7 × 18 = 126

N = **63 fish**

ASSUMPTIONS — WORTH MARKS

- The population has **not changed** between the samples — no births, deaths or migration.
- The marks do **not come off** and are not missed.
- Being marked does **not change** how likely an item is to be caught again.
- The marked items **mixed back in** evenly.
- Both samples were taken **at random**.

JUDGING THE INVESTIGATION

- A **bigger** sample gives a more reliable estimate.
- Very **few marked** in the second sample makes N unreliable — one extra fish would swing it a long way.
- Leaving a **long gap** between samples breaks the "no change" assumption.
- The answer is an **estimate**, so say so.

WHY THE ANSWER IS SENSITIVE

Marked in 2nd	Estimate N
2 of 8	48
3 of 8	32
4 of 8	24

One fish either way moves the estimate by **8 to 16** — that is why sample size matters.

⚠ COMMON EXAM MISTAKES (recurring errors flagged in examiners' reports)

- ✕ Putting the two fractions together the **wrong way round**, giving an N smaller than the sample.
- ✕ Using the **unmarked** count in the second sample instead of the marked count.
- ✕ Adding the two sample sizes to "get" the population.
- ✕ Giving an answer that is **smaller** than the number already marked — always check it is sensible.
- ✕ Leaving the answer as a **decimal** — round to a whole number of fish.
- ✕ Writing assumptions about the **method** rather than about the **population**.
- ✕ Saying "the population is N" instead of "**an estimate** of the population is N".
- ✕ Forgetting that the total marked is the size of the **first** sample.

Revise it: 20 marked, then 9 caught with 4 marked — estimate the population.