Sampling and capture–recapture

S1 · Statistics · revision notes

Populations and samples: why samples are used, how to take a random sample, why a sample may be biased, how to use a sample to estimate for a whole population, and capture–recapture to estimate the size of an animal population. This comes up on both tiers, usually as short 'explain' questions or a capture–recapture calculation.

Key facts

- Population: the whole group being studied
- Random sample: every member has an equal chance of being chosen
- Estimate for the population = proportion in the sample × population size
- M = number marked the first time, C = number caught the second time, R = number of these that are marked
- Grade 6+ only Assumptions: population unchanged, marked animals mix evenly, marks not lost, equal chance of being caught
- Census: data from every member; sample: data from part of the population
- Biased sample: some groups are more likely to be chosen, so it does not represent the population
- Capture–recapture: $N = \frac{M \times C}{R}$
- A larger random sample gives a more reliable estimate

Notes

Populations and samples

- The **population** is the whole group you want to find out about, e.g. all 900 students at a school.
- A **census** collects data from every member of the population. A **sample** collects data from only part of it.
- A sample is quicker and cheaper, and is the only option when testing destroys the item (e.g. testing how long batteries last).
- A sample may not represent the population, so any conclusion is only an estimate. A larger sample usually gives a more reliable estimate.

Random samples and bias

- In a **random sample**, every member of the population has an equal chance of being chosen.
- To take one, list the population, give each member a number, then use a random number generator (or a calculator) to choose numbers, ignoring repeats.
- A **biased** sample does not represent the population fairly, because some groups are more likely to be chosen than others, e.g. asking only people at a gym how often people exercise.
- A good sample includes the different groups (ages, places, times of day) in roughly the same proportions as the population.

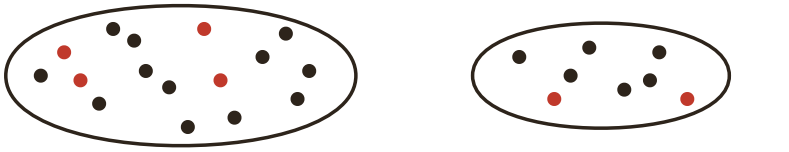
Estimating from a sample

- Estimate for the population = (number in the sample with the feature ÷ sample size) × population size.
- e.g. 12 out of a random sample of 50 students walk to school. For a school of 900 students: 12 ÷ 50 × 900 = 216 students.

Capture–recapture

- Catch M animals, mark them and release them.
- Later, catch a second sample of C animals and count how many of them, R, are marked.
- Assume the fraction marked is the same in the second sample as in the whole population, N: $\frac{R}{C} = \frac{M}{N}$, so $N = \frac{M \times C}{R}$

R ÷ C = M ÷ N



Population: N animals, M of them marked

Second sample: C animals, R of them marked

Marked animals are red. The fraction marked in the sample should match the fraction marked in the population.

- Assumptions: the population does not change between the samples (no births, deaths or animals moving in or out), the marked animals mix evenly with the others, the marks do not come off, and marked and unmarked animals are equally likely to be caught. Grade 6+ only
- If too few marked animals are recaptured (e.g. marks wash off, or marked animals hide), R is too small, so the estimate of N is too big. If too many are recaptured (e.g. the second sample is taken where they were released, before they have mixed), the estimate is too small. Grade 7+ only

How to answer each type of question

M = method mark, A = answer mark, B = a mark on its own, C = a communication mark.

Explain why a sample may be biased and how to improve it 1 to 2 marks **Grade 3**

1. Ask: could every member of the population have been chosen with the same chance?
2. Name the group that is over- or under-represented, in context.
3. To improve it, suggest a random sample from the whole population, or a larger one.

Example. Amir wants to find out how often the adults in his town use the leisure centre. He asks 40 adults as they leave the leisure centre on a Saturday morning.
(a) Give one reason why his sample is likely to be biased.
(b) Suggest one way he could improve his sample.

Model answer.

(a) Everyone he asks has just used the leisure centre, so they are likely to use it more often than other adults in the town (C1)

(b) Choose a random sample of adults from the whole town, e.g. from a numbered list using random numbers (C1)

Describe how to take a random sample 2 marks **Grade 4**

1. Say how you get a list of the whole population and number it.
2. Say how you choose: random numbers from a calculator or computer, ignoring repeats.
3. Say how many you choose.

Example. A school has 850 students.
Describe how the head teacher could choose a random sample of 50 of these students.

Model answer.

Use the school register to give each student a different number from 1 to 850 (C1)

Use a random number generator to choose 50 different numbers, ignoring repeats, and select those students (C1)

Use a sample to estimate for the population

2 marks **Grade 4**

1. Write the result of the sample as a fraction of the sample size.
2. Multiply by the size of the population.
3. Round to a whole number if you are counting items.

Example. A factory makes 6000 light bulbs each day. A random sample of 150 of these bulbs contains 4 faulty bulbs.
Work out an estimate for the number of faulty bulbs the factory makes each day.

Model answer.

$\frac{4}{150} \times 6000$ (M1)

= 160 (A1)

Estimate a population using capture–recapture

3 marks **Grade 5**

1. Identify M (marked the first time), C (caught the second time) and R (marked in the second sample).
2. Write the proportion equation $\frac{R}{C} = \frac{M}{N}$
3. Solve it: $N = \frac{M \times C}{R}$, and give a whole number.

Example. Priya wants to estimate the number of fish in a pond.
She catches 40 fish, marks each one and puts them back in the pond.
A week later she catches 50 fish. 8 of these fish are marked.
Work out an estimate for the number of fish in the pond.

Model answer.

$\frac{8}{50} = \frac{40}{N}$, or 50 ÷ 8 = 6.25 (M1)

$N = \frac{40 \times 50}{8}$ (M1)

= 250 fish (A1)

State an assumption, or decide if an estimate is too high or too low 1 to 3 marks **Grade 7**

1. For an assumption, give one condition the method relies on, in context.
2. For too high or too low, decide whether R (the number marked in the second sample) would be too small or too big.
3. A smaller R makes $\frac{M \times C}{R}$ bigger; a bigger R makes it smaller.

Example. A scientist uses capture–recapture to estimate the number of frogs in a pond.
(a) Write down one assumption the scientist makes.
(b) Some of the marks wash off the frogs before the second sample is taken. Explain the effect this has on the scientist's estimate.

Model answer.

(a) For example: no frogs join or leave the pond between the two samples (C1)

(b) The estimate will be too high (B1), because fewer of the frogs in the second sample are seen to be marked, so R is too small and $\frac{M \times C}{R}$ is too big (C1)

Don’t lose marks

- Dividing by the wrong number in capture–recapture. The bottom of $\frac{M \times C}{R}$ is the number of marked animals in the second sample.
- Writing 'pick people randomly' without saying how. Describe numbering the population and choosing with random numbers.
- Saying only 'the sample is too small' when asked why it is biased. A small sample is less reliable, but bias means some groups are more likely to be chosen, so name the group.
- Giving the proportion in the sample (e.g. $\frac{4}{150}$) instead of scaling it up to the population.
- Leaving a population estimate as a decimal, such as 312.5 animals. Round it to a whole number.

More tips

Memory tricks

- Capture–recapture as two equal fractions: marked in the second sample ÷ size of the second sample = number marked ÷ population.
- 'M times C over R': multiply the two catches, then divide by the number of marked animals recaptured.
- Sense check: a population estimate can never be smaller than either sample.
- Bias test: could everyone in the population have been picked with the same chance? If not, the sample may be biased.

Exam technique

- Answer sampling questions in context: name the people, places or times in the question.
- In capture–recapture, write the fractions or the formula before you substitute. This earns the first method mark.
- When asked if an estimate is too high or too low, state the direction first, then give the reason linked to the number of marked animals recaptured.
- For 'describe how to take a random sample', give two steps: numbering the whole population, and choosing with random numbers.

What each grade needs

What you need to be able to do, from the first marks up to the top grade.

Grade 2 — **Know the words population, sample and census**
The population is the whole group; a sample is part of it; a census collects data from every member.

Grade 3 — **Explain why a sample may be biased**
Say which group is more or less likely to be chosen, so the sample does not represent the population.

Grade 4 — **Describe how to take a random sample**
Number every member of the population and use random numbers to choose, so each has an equal chance.

Grade 4 — **Use a sample to estimate for a population**
Scale up the proportion in the sample: proportion × population size.

Grade 4 — **Suggest how to make a sample more reliable**
Use a larger random sample, taken from the whole population rather than one place or time.

Grade 5 — **Estimate a population using capture–recapture**
Assume the proportion marked in the second sample equals the proportion marked in the whole population.

Grade 6 — **State the assumptions of capture–recapture**
The population does not change between samples, marked animals mix evenly, and marks are not lost.

Grade 7 — **Decide if an estimate is too high or too low**
Work out how a broken assumption changes the number of marked animals recaptured, and so the estimate.

Free notes, Full packs, practice papers and mock exams at chhetriacademy.com. Written for this site; not taken from past papers. Stuck? 1-to-1 tutoring with a free 30-minute trial lesson.