Statistics & Probability · Unit S11 · Standard

Who Do You Ask?

You cannot ask everybody. So you ask some people and speak about everybody — and that only works if the some were chosen properly and enough of them answered. This unit is the machinery: the frame you draw from, the four ways of drawing, the arithmetic of fair shares, and the number that quietly wrecks good surveys — the response rate.

Standard (MYP 4–5) · about 50 minutes · 60 practice questions · Criteria A, B, C and D

BY THE END OF THIS UNIT YOU CAN

• **Explain** what a **sampling frame** is, and identify when a frame does not cover the population it claims to

• Draw a **simple random sample** from a numbered frame using a row of random numbers, discarding correctly

• **Calculate** the interval *k* for a **systematic** sample, choose a random start, and **explain** the edge cases

• **Calculate** the shares of a **stratified** sample, and resolve the rounding so that the shares add to the sample size

• **Explain** why **convenience** and **quota** samples are biased, and **discuss** what each is still good for

• **Calculate** a **response rate** and **discuss** what a low one does to a conclusion, even from a perfect sample

• **Discuss** the trade-off between sample size, cost and accuracy, and choose a technique to fit a purpose

THE ONE IDEA BEHIND ALL OF IT

A sample is a promise: *these people stand for all those people*. Two separate things can break the promise. The first is **how you chose** — if some members of the population could never have been picked, the sample was broken before anybody answered. The second is **who answered** — if only a third reply, and the third who reply are the ones who care most, the numbers you hold describe them and not the population. A perfect method and a poor response rate still give you a

poor survey.

frame → method → sample → replies

and a survey is only as good as the weakest of those four

WHERE THIS UNIT SITS

Where this sits. S03 *Don’t Be Fooled* gave you the vocabulary of bias in ordinary words — who was asked, where, and when. This unit makes the same ideas technical: the frame, the four named techniques, and the arithmetic that goes with each one.

What you already have. The stratified share is a **ratio split**: the sample is shared over the groups in the ratio of their sizes, exactly as Number Unit N11 *Share It Out* shares a quantity in a given ratio. Nothing about ratio is retaught here — only the one new problem it raises, which is that people do not come in fractions.

What comes next. S12 shows what a chart can do to honest numbers. Everything after that — averages, spread, curves, correlation — is computed *from a sample*, so every conclusion in the rest of the strand rests on the choices made in this unit.

CONCEPT 1 · THE FRAME YOU DRAW FROM

population, frame, and the gap between them

The **population** is everybody the question is about. The **sampling frame** is the list you actually draw from — the register, the membership roll, the route’s rider list. The frame is what makes sampling possible, and it is also the first place a survey goes wrong: if the frame leaves people out, no amount of clever drawing puts them back. A whole-school question answered from a class WhatsApp group has a frame of one class, whatever the report says. Always name the population, then name the frame, then say out loud who is in one and not the other.

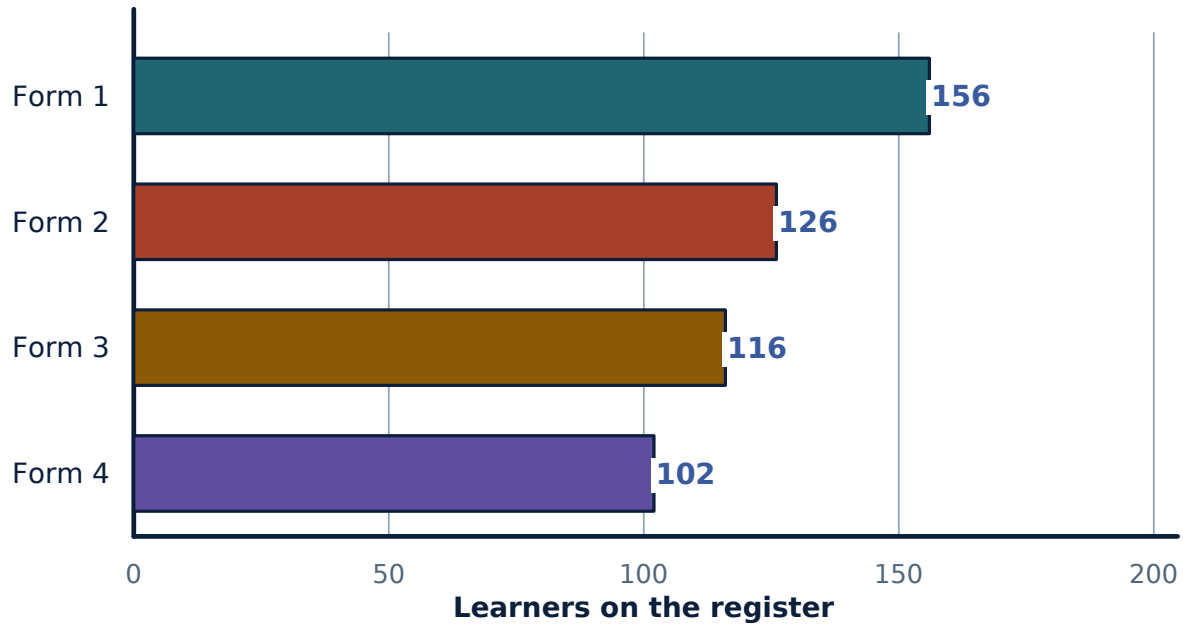
The sampling frame · one school register

Form group	Learners on the register
Form 1	156
Form 2	126
Form 3	116
Form 4	102
Total	500

The frame: every learner who could be picked, counted once.

The whole register, counted. Population and frame agree here, which is what lets a sample drawn from it speak for the school.

Strata block panel · the register split into four groups



Each block is one form group, drawn to the size of its roll.

The same four numbers as blocks. A stratified sample takes a slice of each block in proportion to its length.

EXAMPLE. The question is about all 500 learners in the school. The list you have is the register: all 500 names, one line each. Population and frame agree, so a sample drawn from that register can speak for the school. Draw instead from the 40 names of one Form 3 class and the frame is 40 — a fine frame for a question about that class, and the wrong frame for a question about the school.

CONCEPT 2 · NUMBER AND DRAW

simple random sampling

In a **simple random sample** every member of the frame has an equal chance of being chosen, and so does every possible group of that size. You number the frame, then read random numbers — from a printed row, a calculator or a spreadsheet — and take the names those numbers point at. Three discards keep it honest: **00** (there is no name 0), any number **bigger than the frame**, and any **repeat** (a person cannot be in the sample twice). It is the fairest method there is; its cost is that you need a complete numbered frame before you start.

Simple random sample · six of the forty

Name drawn	Number in the frame
Masego	21
Lebogang	18
Thato	37
Kgosi	16
Karabo	10
Lesedi	19

Six numbers kept from the row, in the order they were read.

The six names the printed row of random numbers picked out, in the order they were read.

The frame, printed in full — forty learners of one Form 3 class, numbered in register order.

#	Name	#	Name	#	Name	#	Name
1	Amogelang	11	Katlego	21	Masego	31	Palesa
2	Bakang	12	Keabetswe	22	Mmoloki	32	Refilwe

#	Name	#	Name	#	Name	#	Name
3	Boitumelo	13	Kefilwe	23	Mpho	33	Reneilwe
4	Bonolo	14	Kelebogile	24	Naledi	34	Sefako
5	Gaone	15	Keneilwe	25	Neo	35	Segomotso
6	Goitseone	16	Kgosi	26	Nthabiseng	36	Thabo
7	Gorata	17	Khumo	27	Oarabile	37	Thato
8	Kabelo	18	Lebogang	28	Obakeng	38	Tshepo
9	Kagiso	19	Lesedi	29	Onalenna	39	Tumelo
10	Karabo	20	Lorato	30	Pako	40	Wame

EXAMPLE · the row of random numbers, in full. Read across, two digits at a time: 69 45 78 21 81 78 50 18 37 78 56 16 10 95 37 19 22 82 60 67 48 25 70 58 92 40 66 85 28 96 46 16 12 72 19 83.

Worked example 1 · Number and draw

1	Number the frame 1 to 40, in register order, so every learner has exactly one number.	**40 numbers**
2	Read 21: it is in 1–40 and is new, so keep it.	**21. Masego**
3	Read 18: it is in 1–40 and is new, so keep it.	**18. Lebogang**
4	Read 37: it is in 1–40 and is new, so keep it.	**37. Thato**
5	Read 16: it is in 1–40 and is new, so keep it.	**16. Kgosi**
6	Read 10: it is in 1–40 and is new, so keep it.	**10. Karabo**
7	Read 37 again: Thato is already held, so discard it.	**discard**
8	Read 19: it is in 1–40 and is new, so keep it.	**19. Lesedi**

9	Stop at six names and read the sample back.	**Masego, Lebogang, Thato, Kgosi, Karabo, Lesedi**

Three discard rules and nothing else: 00, anything above the frame size, and any repeat. Write the discards down as you go — a marker cannot see that a number was rejected for the right reason unless the working says so.

EXAMPLE. Reading the printed row 69 45 78 21 81 78 − the first number, 69, is bigger than 40, so it goes. So do 45 and 78. The fourth, 21, is kept: it is Masego. You carry on until six names are held.

CONCEPT 3 · EVERY K-TH

systematic sampling

A **systematic sample** takes every *k*-th name off the list. Work out **k = N ÷ n** — the frame size divided by the sample size — take the whole-number part, then pick a **random start** between 1 and *k* and step on by *k* from there. It is quick, it spreads the sample evenly through the list, and it needs no random numbers after the first one. Two edge cases matter. When *N* is not a multiple of *n* the last step may run past the end of the list, so you say in advance what you will do. And if the list itself repeats with period *k* — girl, boy, girl, boy — every pick can land on the same kind, and the sample is worthless.

EXAMPLE · the systematic sample, in full. Register of 500, sample of 25, k = 20, random start 9: 9, 29, 49, 69, 89, 109, 129, 149, 169, 189, 209, 229, 249, 269, 289, 309, 329, 349, 369, 389, 409, 429, 449, 469, 489.

Worked example 2 · Every k-th

1	Work out the interval: k = N ÷ n.	**k = 500 ÷ 25 = 20**
2	Choose a random start between 1 and k.	**start = 9**
3	Step on by k from the start.	**9, 29, 49, 69, 89, …**
4	Check the last one is still on the register.	**9 + 24 × 20 = 489**
5	Edge case: what if the roll were 510, not 500?	**510 ÷ 25 = 20.4**

6	Take k as the whole-number part and say in advance what happens at the end of the list.	**k = 20, and stop at 25 names**
7	Edge case: what if the register repeated girl, boy, girl, boy?	**an even k picks one gender only**

k comes from the two sizes, and the start comes from chance. Skip the random start and you have not sampled at all: you have taken a fixed slice of the register, and the register's own order decides your answer.

EXAMPLE. A register of 500 and a sample of 25 give k = 500 ÷ 25 = 20. A random start of 9 gives 9, 29, 49, 69, and so on, ending at 489 — twenty-five names, evenly spread down the whole register.

CONCEPT 4 · FAIR SHARES

stratified sampling and the rounding

A **stratified sample** first splits the population into **strata** — groups that differ in a way that matters, such as form, suburb or shift — and then takes from each group in proportion to its size: share = n × (group roll) ÷ (whole roll) Each group is then sampled at random inside itself. It guarantees the shape of the sample matches the shape of the population, which no other method here does. The one problem is arithmetic: the shares almost never come out whole, and rounding each one to the nearest whole number can leave you one over or one under. That has to be fixed, and the fix has to be stated.

Stratified sample · fifty shared over four groups

Form group	Learners in the sample
Form 1	16
Form 2	13
Form 3	11
Form 4	10
Total	50

The allocation that adds to 50, rounding tension resolved.

The allocation after the rounding was repaired: 16 + 13 + 11 + 10 = 50, which is the sample size asked for.

Worked example 3 · Fair shares

1	Write the raw share of each form: 50 × roll ÷ 500.	**Form 1 15.6; Form 2 12.6; Form 3 11.6; Form 4 10.2**
2	Round each one to the nearest whole number and add them up.	**16 + 13 + 12 + 10 = 51**
3	That is 51 learners, not 50. The obvious method has overshot.	**one too many**
4	Start again: give every form the whole-number part of its share.	**15 + 12 + 11 + 10 = 48**
5	Hand out the seats still missing to the largest fractional parts, one each; a tie goes to the larger form.	**Form 1 0.6; Form 2 0.6; Form 3 0.6; Form 4 0.2**
6	Read the allocation back and check the total.	**16 + 13 + 11 + 10 = 50**

The rounding is not a detail to be tidied away in silence. State the rule you used, show that the seats add to the sample size, and the marker can follow the whole allocation — which is exactly what the mark is for.

EXAMPLE. Form 1 has 156 of the 500 learners. Its share of a sample of 50 is 50 × 156 ÷ 500 = 15.6 learners — and nobody can send 0.6 of a learner, which is where the work in this concept starts.

CONCEPT 5 · THE CHEAP SHORTCUTS

convenience and quota

A **convenience sample** is whoever is easy to reach — the queue at the gate, your own class, the people who answer an online poll. A **quota sample** fixes how many of each kind you want — 20 girls, 20 boys — but then lets the interviewer choose *which ones*. Both are fast and cheap, and both are **biased**, because who is easy to reach is never a random feature of a population: the gate at 07:00 is full of early arrivals, and an interviewer avoids people who look unfriendly. Bias does not shrink when the sample grows — a biased sample of 400 is a bigger picture of the wrong people. They are still useful for a quick pilot, for testing whether a question is understood, or when no frame exists at all; what they cannot do is carry a number back to the population.

Worked example 5 · Pick the method

1	Does the tuck shop queue move faster after the new till? You have the till’s own list of 500 sales, numbered by receipt.	**simple random sample**
2	How long do learners wait for a combi after school? A prefect stands at the stop with a clipboard and no list of riders exists.	**systematic sample**
3	Do learners want the sports day moved to April? Junior and senior forms are likely to disagree, and the four forms differ in size.	**stratified sample**
4	In every case, say what the choice costs as well as what it buys.	**the trade-off is part of the answer**

There is no best technique, only a best fit. A frame that already exists points to random; a stream with no frame points to systematic; groups that differ point to stratified. Convenience is on none of these lists.

> **EXAMPLE.** “We asked 400 people online and 78% agreed.” Four hundred is a big number, and it is a big number of people who happened to see the post and felt like clicking. A properly drawn random 50 from the register would tell you far more about the school than that 400 ever can.

CONCEPT 6 · WHO ACTUALLY REPLIED

response rate and sample size

The **response rate** is the fraction of the people you asked who actually answered: response rate = replies ÷ people asked A perfectly drawn sample with a 30% response rate is no longer a random sample of the population: it is a sample of *the people who reply*, and they usually differ from the rest in exactly the way the survey is asking about. That is **non-response bias**, and it is not cured by asking more people — only by getting more of the people you asked to answer, or by checking who the missing ones were. Sample size still matters: a bigger sample narrows ordinary sampling variation, but it costs time and money, and it never repairs a bad frame or a low response rate.

Response rate · the survey that went out to 200

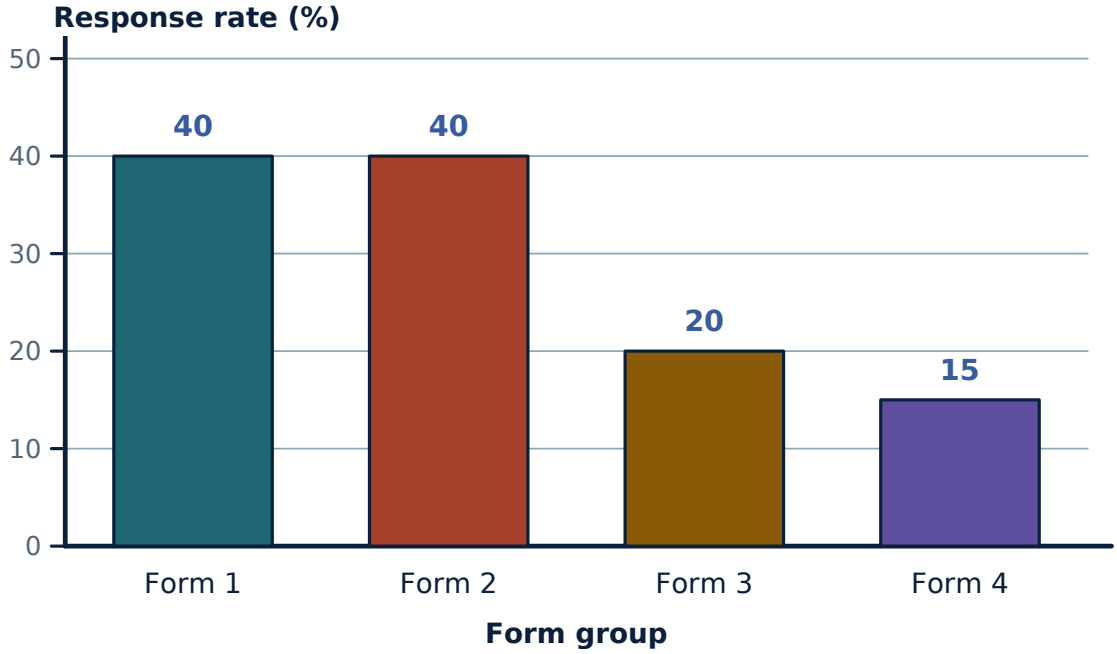
The homework survey, in three numbers

200	60	30%
learners asked	replied	response rate

Seven in every ten of them never answered at all.

The same survey in three numbers. Two of them are counts; the third is what those counts do to each other.

Who actually replied · response rate by form group



Same survey, same day. The seniors simply did not reply.

Response rate by form group. The survey was sent to a balanced sample; what came back was not balanced at all.

The survey, printed in full.

Form group	Asked	Replied	Response rate
Form 1	60	24	40%

Form group	Asked	Replied	Response rate
Form 2	50	20	40%
Form 3	50	10	20%
Form 4	40	6	15%
Whole survey	200	60	30%

Worked example 4 · The 30% reply

1	Work out the response rate of the whole survey.	**60 ÷ 200 = 30%**
2	Work out the rate for each form group.	**Form 1 40%; Form 2 40%; Form 3 20%; Form 4 15%**
3	Find the share of the 200 asked that was Form 1.	**30%**
4	Find the share of the 60 replies that was Form 1.	**40%**
5	Do the same for Form 4.	**asked 20%, replied 10%**
6	Say what the 60 replies now describe.	**the learners who reply, not the school**

The sample that was drawn and the sample that answered are two different samples. Only the second one is in your hands, and the report has to say so.

EXAMPLE. Two hundred learners were asked about homework and 60 replied: a response rate of 30%. Split by form the rates were 40%, 40%, 20% and 15% — so the 60 replies hold far more juniors than the 200 who were asked did. The sample that was *drawn* was balanced; the sample that *answered* is not.

THE SAMPLING ROUTINE

1. Name the population, then the frame. Who is the question about, and what list will you actually draw from? Say who is in one and not the other before anything else.

2. Choose the technique for the purpose. Random for fairness, systematic for speed on a long list, stratified when groups differ in a way the question cares about.

3. Do the arithmetic and make it add up. k = N ÷ n for systematic; share = n × group ÷ total for stratified — then check the shares add to n, and say how you fixed the rounding.

4. Ask what came back. Response rate = replies ÷ asked. Then ask who the missing ones were, and what that does to the answer you are about to give.

THE MISTAKES THAT COST THE MOST MARKS

Trap 1 · Stratified shares that do not add to the sample size

Wrong. “50 × 156 ÷ 500 = 15.6, so 16; and 12.6, so 13; and 11.6, so 12; and 10.2, so 10. Sample of 50 — done.”

Right. 16 + 13 + 12 + 10 = **51**, which is one learner more than the sample asked for. Take the whole-number parts (15, 12, 11, 10 = 48), then give the two missing seats to the largest fractional parts, larger form first: **16, 13, 11, 10 = 50**.

Rounding each share on its own is rounding four numbers separately and hoping. The total is the check, and it has to be done before the sample is drawn, not after. Whichever rule you use, **say which rule you used** — an allocation that adds up but is not explained loses the explanation mark, not the arithmetic mark.

Trap 2 · Sampling from the wrong frame

Wrong. “I posted the question in the Form 4 WhatsApp group and 63 learners replied, so 63 learners of the school think the sports day should move.”

Right. The population is the **school**; the frame was **one form’s WhatsApp group**. Learners with no data bundle, no phone, or no place in that group could never have been picked at all. The finding belongs to that group and to nobody else.

This is the failure that no arithmetic can repair, because it happens before the arithmetic starts. Name the population and name the frame as two separate sentences, then say who is in the first and not the second. If that sentence is uncomfortable to write, the frame is wrong.

Trap 3 · Treating a big convenience sample as better than a small random one

Wrong. “400 people answered the online poll and only 50 were on the random list, so the poll is eight times as reliable.”

Right. The 400 are whoever saw the post and chose to click — a **self-selected** group. The 50 were drawn so that every learner had an equal chance. The random 50 speaks for the school; the 400 speaks for the clickers, however many of them there are.

Sample size fixes only ordinary sampling variation. It does nothing at all to bias: a biased sample of 400 is the same wrong answer measured more precisely. Ask “could every member of the population have been picked?” before you ever ask “how many?”

PRACTICE · 60 QUESTIONS

Answers are at the back. Work each one out before you look.

Concept 1 · The frame you draw from

P1 (Fluency)	**State** the name for the complete list of everybody who could be chosen for a sample.
P2 (Fluency)	A survey asks what the whole school thinks of the sports day. **State** the population.
P3 (Fluency)	The register of the whole school is used as the list to draw from. **State** the size of the sampling frame.
P4 (Fluency)	One Form 3 class list of 40 names is used instead. **State** the size of the sampling frame.
P5 (Applied)	A whole-school question is posted in one form's WhatsApp group. **Explain** what is wrong with that sampling frame.
P6 (Applied)	A survey is planned about riders on one combi route. **Explain** the difference between the population and the frame here.
P7 (Applied)	The school register lists all 500 learners, one line each. **Explain** why it is a good frame for a whole-school question.
P8 (Applied)	The frame used is one Form 3 class of 40 names. **Calculate** how many learners of the school are outside that frame.
P9 (Challenge)	A rider survey draws its names from a list of *households* along the route. **Discuss** whether that frame fits a question about riders.
P10 (Challenge)	"We will fix the WhatsApp-group survey by collecting 200 replies instead of 60." **Explain** why that does not fix it.

Concept 2 · Number and draw

P11 (Fluency)	**State** the three kinds of number that are discarded when reading random numbers against a frame of 40.
P12 (Fluency)	The printed row begins 69 45 78 21. **State** the first number that is kept.
P13 (Fluency)	**State** the name that number 16 draws from the frame.
P14 (Fluency)	**State** the number that draws Thato from the frame.
P15 (Applied)	Reading the row, the number 78 comes up. **Explain** what is done with it and why.
P16 (Applied)	Later in the same row the number 37 appears for a second time. **Explain** what is done with it.
P17 (Applied)	**State** the six names the printed row draws, in order.
P18 (Applied)	A simple random sample of 6 is drawn from the frame of 40. **Calculate** the probability that one particular learner is in it.
P19 (Challenge)	All six drawn numbers happen to be under 40. A learner says the draw was not random. **Discuss** the claim.
P20 (Challenge)	The printed row runs out before six usable numbers are found. **Explain** what to do next without spoiling the sample.

Concept 3 · Every k-th

P21 (Fluency)	A register of 500 names is to give a systematic sample of 25. **Calculate** k.
P22 (Fluency)	**State** the numbers the random start may take when k is 20.

P23 (Fluency)	The start is 9 and k is 20. **State** the second name’s number.
P24 (Fluency)	The same sample runs to 25 names. **State** the last number.
P25 (Applied)	A register of 510 names is to give a systematic sample of 25. **Calculate** k and **explain** your choice.
P26 (Applied)	Stepping by k passes the end of the register before the sample is full. **Explain** two ways of dealing with it.
P27 (Applied)	**Explain** why the start of a systematic sample must be chosen at random.
P28 (Applied)	Every 20th name is taken from the register of 500, starting at 1. **Calculate** how many names the sample holds.
P29 (Challenge)	A register lists learners in pairs — one girl, one boy, one girl, one boy — and k comes out even. **Discuss** the effect on the sample.
P30 (Challenge)	**Discuss** when a systematic sample is a better choice than a simple random one, and when it is worse.

Concept 4 · Fair shares

P31 (Fluency)	Form 1 has 156 of the 500 learners, and the sample is 50. **Calculate** its raw share.
P32 (Fluency)	Form 2 has 126 learners. **Calculate** its raw share of the same sample of 50.
P33 (Fluency)	The allocation rule has been applied to those shares. **State** how many learners are drawn from Form 1.
P34 (Fluency)	**Calculate** the total of the four allocated shares 16, 13, 11, 10.

P35 (Applied)	The four raw shares are 15.6, 12.6, 11.6, 10.2. **Calculate** what rounding each to the nearest whole number gives altogether, and **explain** the problem.
P36 (Applied)	**Explain** the rule that always makes the stratified shares add to the sample size.
P37 (Applied)	School B has rolls 240, 180, 150, 130 and wants a sample of 40. **Calculate** the share drawn from Form 1.
P38 (Applied)	School C has rolls 338, 229, 188, 145 and wants a sample of 45. **Calculate** what rounding its raw shares to the nearest whole number gives altogether.
P39 (Challenge)	**Explain** why the fractional parts of the raw shares always add up to a whole number.
P40 (Challenge)	Three groups tie with a fractional part of 0.6 and only two seats are left. **Discuss** a fair way to break the tie.

Concept 5 · The cheap shortcuts

P41 (Fluency)	**State** the name for a sample made of whoever happens to be easy to reach.
P42 (Fluency)	**State** the name for a sample where the numbers of each kind are fixed but the interviewer chooses the individuals.
P43 (Fluency)	Learners are asked at the school gate at 07:00. **Identify** the kind of sample.
P44 (Fluency)	**State** whether making a convenience sample bigger removes its bias.
P45 (Applied)	An online poll collects 400 replies. **Explain** why it is still a biased sample.

P46 (Applied)	A quota sample fills exactly 20 girls and 20 boys. **Explain** why it can still be biased.
P47 (Applied)	**Explain** one job a convenience sample can still do honestly.
P48 (Applied)	No list of riders on a route exists at all. **Discuss** how a fair sample could still be taken.
P49 (Challenge)	**Discuss** the difference between a quota sample and a stratified sample, given that both fix how many come from each group.
P50 (Challenge)	“A sample of 400 must be more accurate than a sample of 50.” **Discuss** the claim.

Concept 6 · Who actually replied

P51 (Fluency)	200 learners were asked and 60 replied. **Calculate** the response rate.
P52 (Fluency)	In Form 3, 50 learners were asked and 10 replied. **Calculate** the response rate.
P53 (Fluency)	**State** how many of the 60 replies came from Form 1.
P54 (Fluency)	In Form 4, 40 were asked and 6 replied. **Calculate** the response rate.
P55 (Applied)	**Calculate** the share of the 60 replies that came from Form 1, and compare it with that form’s share of the 200 who were asked.
P56 (Applied)	The survey is about homework. **Explain** how a 30% response rate can bend the result.
P57 (Applied)	A survey expects the same 30% response rate and needs 90 replies. **Calculate** how many learners must be asked.

P58 (Applied)	“Ask 400 learners instead of 200 and the low response rate stops mattering.” **Discuss** the claim.
P59 (Challenge)	A survey team may spend its remaining time either asking 100 more learners or chasing replies from the ones already asked. **Discuss** which is the better use of the time.
P60 (Challenge)	A sample is drawn perfectly at random and 30% of it replies. **Explain** why the replies are no longer a random sample.

CRITERION B · INVESTIGATING PATTERNS

The rounding seats

A stratified sample shares n places over the groups of a population in proportion to their sizes: **share = n × (group roll) ÷ (whole roll)**. The shares almost never come out whole, and people do not come in fractions. Rounding each share on its own is the obvious repair — and it does not always give a sample of n. Three schools are printed below, with every roll. Find out when the obvious method fails, and what to do instead.

The three schools — every roll printed

School	Group	Roll	Whole roll	Sample wanted
School A	Form 1	156	500	50
	Form 2	126		
	Form 3	116		
	Form 4	102		
School B	Form 1	240	700	40
	Form 2	180		
	Form 3	150		
	Form 4	130		
School C	Form 1	338	900	45
	Form 2	229		
	Form 3	188		
	Form 4	145		

Three invented schools, written for this task.

Part 1 · the structured investigation

1. **Predict**, before working anything out, whether rounding each raw share to the nearest whole number will give exactly the sample size asked for in each of the three schools. Write the three predictions down.

2. **Tabulate**, for each school and each group: the raw share, its whole-number part, its fractional part, and the nearest whole number. Add each column of nearest whole numbers and compare the total with the sample size.

3. **Identify** the pattern. Look at the fractional parts of one school and say what they always add up to, then say how the number of parts that are 0.5 or more decides whether the nearest-rounding total is too big, too small or exactly right.

4. **Determine** a general allocation rule that gives a total of n every time, for any number of groups and any sample size. Write it as instructions somebody else could follow without seeing your table.

5. **Test and verify** your rule on all three schools, and then on a fourth school of your own invention whose rolls do not divide neatly. Show the totals adding to n each time.

6. **Justify** the rule’s fairness: explain which group each seat goes to and why that is a defensible choice, and say what your rule does to a group whose raw share is below 0.5 in the first place.

School A is done for you. The blank rows are for School B, and you will need another four for School C.

Group	Roll	Raw share	Whole part	Fraction	Nearest	Seats
Form 1	156	15.6	15	0.6	16	16
Form 2	126	12.6	12	0.6	13	13
Form 3	116	11.6	11	0.6	12	11
Form 4	102	10.2	10	0.2	10	10
Total	500	50	48	2	51	50

Part 2 · your own method choice

Part 2 · your own investigation. Somewhere in your table two fractional parts are equal, and only one seat is left. The mathematics cannot break that tie — a person has to. **Invent** your own tie-break rule and **defend** it. Say exactly what it does, test it on a case where three parts tie at once, and argue why it is fair. Then find a situation in which

your own rule gives an uncomfortable result, and say whether you would still keep it. You choose the rule, the tests and the way you present the argument — a rule that is stated in advance and defended honestly is worth more than one that happens to be popular.

Criterion B · Investigating patterns. Levels 1–2: computes some raw shares correctly and records them. **Levels 3–4:** completes the table for all three schools and describes in words when the nearest-rounding total misses. **Levels 5–6:** states a correct general allocation rule that always totals n, and verifies it on all three schools and on a fourth of the student's own. **Levels 7–8:** in addition, justifies the rule from the structure of the fractional parts — they sum to a whole number, so the seats still to hand out are always countable and the rule can always finish — and carries Part 2 through to a tie-break rule that is stated in advance, tested on a three-way tie, and defended against its own worst case.

Remember, you should communicate in an organized and coherent manner.

CRITERION D · APPLYING MATHEMATICS IN REAL-LIFE CONTEXTS

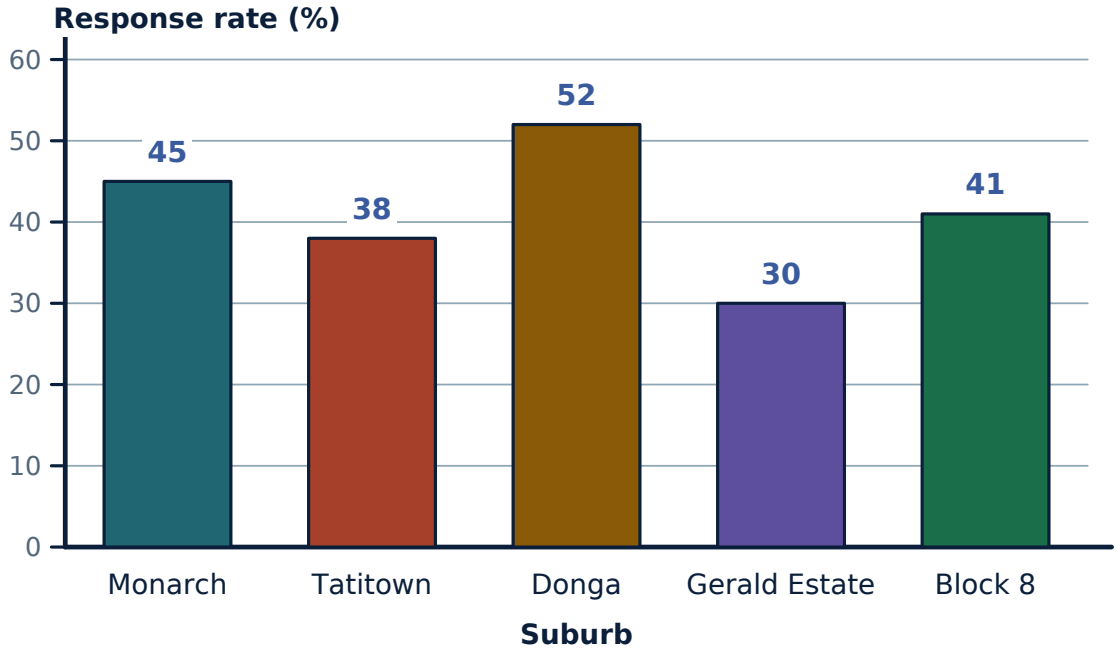
The rider survey on a new combi route

A proposed combi route in Francistown would serve five suburbs. Before it is approved, riders are to be surveyed about fares, waiting times and stops. The planners want **240 completed replies**, shared over the suburbs in proportion to the number of adult riders in each. The sheet below carries last year's survey figures as well, and more columns than this question needs.

Suburb	Adult riders on the route	Asked last year	Replied last year	Walk to the stop (min)	Fare (P)
Monarch	1450	200	90	12	5.50
Tatitown	1100	250	95	18	5.50
Donga	900	150	78	9	4.50
Gerald Estate	750	180	54	22	6.00
Block 8	600	200	82	15	6.00
Total	4800	980	399		

This data sheet is an invented illustrative example written for the task. It is not taken from any business, price list or publication, and no organisation, person or source is named or implied.

Rider survey · last year's response rate by suburb



Last year's rates, from the asked and replied columns.

Last year’s response rate in each suburb, computed from the asked and replied columns of the data sheet.

1. **Interpreting the situation. Explain** which columns of the sheet are needed to plan the sample and which are surplus to it. Name the **population** and the **sampling frame** as two separate sentences, and say who is in the first and not the second.

2. **Selecting the mathematics. Explain** which sampling technique fits this enquiry and why, and say what your choice costs compared with the two you rejected. State the formula you will use for the shares.

3. **Applying the mathematics. Calculate** the raw share of the 240 replies for each suburb, resolve the rounding so that the shares add to 240 exactly, and state the rule you used. Then **calculate**, from last year’s response rates, how many replies each suburb’s list would actually be expected to bring back.

4. **Accuracy. Check** that your shares add to 240 and that each response rate is the replied column divided by the asked column. **Explain** which of your figures are exact and which are estimates, and say which way you rounded the expected replies and why that direction is the safe one.

5. **Sense in context. Discuss** what last year’s response rates do to this year’s conclusion: which suburb ends up under-represented among the replies, and what the planners should do about it. **Discuss** whether

approaching more riders is the right answer, or whether the money is better spent somewhere else.

Criterion D · Applying mathematics in real-life contexts. Levels 1–2: identifies some of the needed columns and computes some raw shares. **Levels 3–4:** selects stratified sampling with a reason, computes all five raw shares and makes a start on the rounding. **Levels 5–6:** shares that add to 240 with the rule stated, all five response rates correct, expected replies computed and clearly labelled as estimates, and an explained conclusion about the suburb that is under-represented. **Levels 7–8:** in addition, discusses the trade-off between approaching more riders and chasing the non-responders — noting that a bigger approach list leaves the tilt exactly where it was — and recognises what the sheet cannot show at all, such as riders who board between the listed suburbs and are on no row of it.

ANSWERS

Every answer here is recomputed from the datasets printed in the questions before this file is written.

P1	sampling frame	**P31**	15.6
P2	all 500 learners of the school	**P32**	12.6
P3	500	**P33**	16
P4	40	**P34**	50
P11	00, anything above 40, and any repeat	**P41**	convenience sample
P12	21	**P42**	quota sample
P13	Kgosi	**P43**	convenience sample
P14	37	**P44**	no
P21	20	**P51**	30%
P22	1 to 20	**P52**	20%
P23	29	**P53**	24
P24	489	**P54**	15%

Applied and Challenge · fully worked

P5 the frame is one form’s group, not the school, so most learners could never be chosen

The population is the school; the frame is the group. Learners in other forms, and learners in that form with no phone or no data, have a chance of zero of being picked. A frame that cannot reach part of the population is broken before a single answer arrives, and no amount of careful drawing inside it repairs the gap.

P6 the population is every adult who rides the route; the frame is the list the survey can actually draw from

The population is defined by the question — everybody who rides that route. The frame is whatever list exists: riders counted at the stops, or a suburb-by-suburb count of adult riders. They are the same only if the list covers every rider, and it usually does not: riders who board between stops, or who ride at night, may be on no list at all.

P7 it covers the whole population exactly once, so every learner can be chosen

Population and frame agree: 500 learners in the school and 500 names on the register, each appearing once. Every learner therefore has a chance of being picked, and no learner can be picked twice. Those two properties are what make a sample drawn from it able to speak for the school.

P8 460

500 − 40 = 460 learners are on the register but not on the class list. They have no chance at all of being sampled, so any finding belongs to the class and not to the school.

P9 it does not fit: households and riders are not the same people, so some riders are missed and some non-riders are included

A household list over-covers and under-covers at the same time. It includes people who never ride, and it counts a household with four riders once and a household with one rider once, so heavy-riding households are under-represented. It also misses riders who live outside the listed suburbs but board along the route. It could be used as a starting frame only if every adult in each household were then listed and screened for riding — which is a different, and much larger, piece of work.

P10 the frame is still one group, so the extra replies come from the same excluded-everybody-else list

More replies make the picture of that group sharper. They do not give a single learner outside the group a chance of being picked, so the gap between the frame and the population is exactly as wide as it was. Size is the cure for sampling variation; it is never the cure for a frame that leaves people out.

P15 it is discarded, because the frame only goes up to 40

There is no learner numbered 78 on a frame of 40, so the number points at nobody. It is crossed out and the reading moves on. Discarding is not a failure of the method — it is the method: every number from 01 to 40 is equally likely to survive, which is exactly what makes the sample random.

P16 it is discarded, because Thato is already in the sample

A person cannot be in a sample twice: the sample is six different learners, not six draws. So the repeat is crossed out and reading continues to the next usable number. (Drawing without replacement like this is why the six names are six distinct people.)

P17 Masego, Lebogang, Thato, Kgosi, Karabo, Lesedi

Reading across and keeping 01–40, new numbers only, gives 21, 18, 37, 16, 10, 19 — and those numbers point at Masego, Lebogang, Thato, Kgosi, Karabo, Lesedi. The order they were drawn in does not matter to the sample, but writing it down is what lets somebody else check the draw.

P18 3/20

Every learner has the same chance, and 6 of the 40 places are filled, so the probability is 6/40 = 3/20. Equal chances for every member of the frame is the definition of a simple random sample, and this fraction is that definition written as a number.

P19 the draw is still random: every number had an equal chance, and a clustered result is a possible outcome of a fair draw

Randomness is a property of the *method*, not of the pattern that comes out. A fair draw can produce a clumped-looking sample, exactly as a fair coin can give four heads in a row. The only way to judge the draw is to check the procedure: was the frame complete, was every number equally likely, were the discards applied consistently? Redrawing because a result “looks wrong” is itself a way of introducing bias.

P20 carry on into the next row of random numbers, using the same discard rules

The row is a supply of random digits, not the sample itself. Continue into the next row, or the next line of the calculator’s output, applying the same three discard rules. What must *not* happen is starting again from a different place chosen by eye, or filling the last name in by hand — either of those turns the last member of the sample into a choice, and a choice is not a random draw.

P25 20

510 ÷ 25 = 20.4, which is not a whole number, so k is taken as the whole-number part, 20. Stepping by 20 from a start of at most 20 reaches at most 500, which is still inside the register, and the sample stops once 25 names are held.

P26 either wrap round to the start of the list, or reduce k by one so the steps fit — and say which you did before you draw

Wrapping treats the register as a circle, so the step after the last name is the first name; every learner still has one chance. Reducing k keeps the sample inside the list but crowds it slightly towards the front. Both are defensible. What is not defensible is deciding afterwards, because by then you can see which learners each choice would give you.

P27 otherwise the register’s own order decides the sample, and it is not a sample at all

With k fixed, the start decides the whole sample: start at 1 and you get one fixed set of 25 names, start at 2 and you get a different fixed set. Choosing the start by chance is the only randomness in the method, so it is what gives every learner a chance of appearing. Always starting at 1 means the first name on the register is in every survey the school ever runs.

P28 25

The names taken are 1, 21, 41, …, 481, and there are 500 ÷ 20 = 25 of them. The count of a systematic sample is the frame size divided by the step, which is exactly how k was chosen in the first place.

P29 every pick lands on the same gender, so the sample is all girls or all boys and represents nobody

Stepping by an even number through a list whose pattern repeats every two names always lands on the same position in the pattern. The sample looks properly spread down the register and is in fact a perfect selection of half the school. This is the one failure mode systematic sampling has that simple random sampling does not, and the check for it is to look at the order of the list before choosing k, not after.

P30 better when the list is long or has no numbers and speed matters; worse when the list has a repeating pattern

Systematic sampling needs one random number instead of n of them, and it can be run on a moving queue where no frame exists at all — every k-th person through the gate. It also spreads the sample evenly through the list, which a random draw need not do. Against that, it is wrecked by periodicity, and it does not give every possible group of n an equal chance — only every k-th-name group. Speed and practicality against a known, checkable weakness.

P35 51

Rounded separately they are 16, 13, 12, 10, and 16 + 13 + 12 + 10 = 51 — one more learner than the sample of 50 asked for. Each rounding is correct on its own; it is their total that is wrong, because four separate roundings do not know about each other.

P36 give each group the whole-number part, then hand the missing seats to the largest fractional parts, one each

The whole-number parts can never overshoot, so after step one you are at or below the target; the number still missing is then handed out one seat at a time to the groups with the largest fractional parts, and the handing out stops the moment the target is reached. It cannot overshoot and it cannot undershoot. A tie between fractional parts needs its own stated rule — this unit gives the seat to the larger group.

P37 14

The raw share is 40 × 240 ÷ 700 = 13.7. Take the whole-number parts of all four raw shares first: 13, 10, 8, 7, which add to 38 — 2 seats short of the 40 wanted, not one. The fractional parts are 0.714, 0.286, 0.571, 0.429, and the 2 largest belong to Form 1 and Form 3, so those groups take a seat each. Form 1 therefore gets 13 + 1 = 14, and the four shares 14, 10, 9, 7 add to 40.

P38 44

The raw shares are 16.9, 11.45, 9.4, 7.25. Rounded to the nearest whole number they are 17, 11, 9, 7, which add to 44 — one learner *short* of the 45 wanted. The obvious method can miss in either direction, which is why the total has to be checked every time.

P39 because the raw shares themselves add to n exactly, so what is left after removing the whole-number parts is a whole number too

Each raw share is n × group ÷ total, and adding them gives n × (sum of groups) ÷ total = n exactly. Take away the whole-number parts, which are whole numbers, and a whole number is left. That is why the number of seats still to hand out is always a whole number, and why the rule can always finish.

P40 any rule works only if it is fixed in advance; this unit gives the seat to the larger group

The three claims are exactly equal, so the mathematics cannot choose — a person must, and the choice must be made before the numbers are seen. Giving it to the larger group loses the smallest fraction of each group’s representation and is easy to defend. Drawing lots is defensible too, and arguably fairer, but it makes the sample impossible to reproduce. What is indefensible is choosing after looking, because then the seat goes wherever the answer is wanted.

P45 the 400 are self-selected: only people who saw the post and chose to answer are in it

Everybody who never saw the post, has no data bundle, or simply scrolled past had no chance of being included, so the frame is “people who saw it and felt strongly enough to click”. People who feel strongly answer online polls at far higher rates than people who do not, so the result is tilted towards strong opinions however many replies arrive.

P46 the interviewer chooses which girls and which boys, and that choice is not random

The proportions are right and the selection inside them is not. An interviewer standing in one place, at one time, approaching people who look approachable, fills the quota from a narrow slice of each group. Quota sampling controls the *shape* of the sample; stratified sampling controls the shape **and** draws at random inside each group, which is the whole difference between them.

P47 a pilot: testing whether the questions are understood before the real sample is drawn

Handing a draft questionnaire to the nearest ten learners will show you a question nobody understands, an option that is missing, and a form that takes twice as long as planned. None of that needs a representative sample, because none of it is a claim about the population. The moment a number from those ten is reported as a finding, the convenience sample has been used dishonestly.

P48 sample the flow instead of the list: take every k-th rider boarding, with a random start, across different times and days

With no frame there is nothing to number, but there is a stream of people, and a stream can be sampled systematically. Spreading the counting across mornings, afternoons and different days stops the sample becoming a sample of one hour’s riders — the commuters at 07:00 are not the same population as the shoppers at 11:00. It is not a simple random sample and should not be called one, but it is defensible and it can be described exactly.

P49 stratified draws at random inside each group; quota lets the interviewer choose, so only stratified can be called random

Both start the same way, by deciding how many from each group. They part company at the last step. Stratified sampling numbers each group and draws from it at random, so every member of every group has a known chance of selection and the arithmetic of sampling error still applies. Quota sampling leaves that last choice to a person, so the chance of any individual being picked is unknown and unknowable. The proportions look identical on the page; only one of them supports a claim about the population.

P50 only if both are drawn the same way: size cuts sampling variation but does nothing to bias

Between two properly drawn random samples, 400 is indeed better than 50 — the ordinary wobble of a sample shrinks as it grows. Between a self-selected 400 and a random 50 the comparison collapses, because the 400 is measuring the wrong population very precisely. Accuracy has two parts: how far the sample scatters, which size fixes, and where it is centred, which only the method fixes.

P55 40%

24 ÷ 60 = 40% of the replies, against 60 ÷ 200 = 30% of those asked. The form is over-represented among the answers by ten percentage points, purely because more of its learners chose to answer.

P56 the learners who reply are likely to be the ones with strong feelings about homework, so the answers over-state those feelings

Non-response is almost never random. A learner who finds the homework load unbearable has a reason to reply; a learner who is untroubled by it has none. The 60 who replied are therefore a sample of the interested, not of the school, and the survey measures the strength of feeling among people with feelings. Nothing in the arithmetic reveals this — only knowing who did not answer does.

P57 300

At 30%, replies are 3/10 of those asked, so the number asked is 90 ÷ 3/10 = 300. Planning from the response rate is the only way to end with the sample you wanted; planning from the sample size alone leaves you a third of the way there.

P58 it gives more replies but the same tilt: the extra replies come from the same kind of learner

Doubling the number asked at the same 30% rate gives about 120 replies instead of 60, so ordinary sampling variation shrinks. The people who stay silent are still silent, and they are still the ones the survey cannot see, so the direction of the error does not move at all. Chasing the non-responders — a second reminder, a paper form for learners without data — buys more than doubling the mailing does.

P59 chasing the replies, because it attacks bias while asking more only attacks variation

The 100 extra learners bring in about 30 more replies at the current rate, and they bring in the same kind of learner, so the tilt survives. Chasing non-responders converts silent members of the drawn sample into answers, and those are exactly the people the survey is missing: every one recovered pulls the result back towards the population. Only when the response rate is already high does adding people become the better buy.

P60 the people who replied selected themselves, so the randomness of the draw does not carry through to the answers

The draw gave every learner an equal chance of being *asked*. It gave nobody an equal chance of *answering* — that was decided afterwards, by the learners themselves, for reasons connected to the subject of the survey. What is in your hands is a random sample filtered by a non-random rule, and a random sample that has been filtered is simply a sample of the filter. This is why the response rate is reported next to the sample size, always.