AP STATISTICS · UNIT 1

Revision Lesson

Past-paper questions with condensed solutions

22
questions

6
topic sections

62
slides in total

Built from the Unit 1 past-paper booklet

How to use this lesson

Key points first, then question, then solution

Key points	Each section opens with a summary slide of the ideas you need before you attempt anything.
Question	The question appears on its own slide, with the mark tariff in the corner. Attempt it before you turn the page.
Solution	The next slide gives the condensed solution — the reasoning that earns the mark, not just the answer.
Trap box	Every solution closes with the mistake that costs marks most often in that question.
Mark scheme	The final section lists the mark tariff for every part, plus a one-slide answer key.

01

Sampling methods

Stratified, cluster, simple random, systematic and convenience

- Questions 4, 8, 10 and 11
- Identify by what is chosen at the second stage: a person, or a whole group

Key points · sampling methods

Name the method, not the activity — the test is what gets chosen at the second stage

1	**The one test**	Ask what is selected the SECOND time: individuals → stratified; whole groups → cluster.
2	**Simple random (SRS)**	Every individual has an equal chance; selection is by a random list of individuals.
3	**Stratified**	Divide into groups that DIFFER, then take a random sample from EVERY group.
4	**Cluster**	Divide into groups that are SIMILAR, randomly select WHOLE groups, then survey everyone in them.
5	**Why stratify**	It reduces variability between samples. It does not reduce bias — an SRS is already unbiased.
6	**Why cluster**	It lowers fieldwork cost. It does not reduce bias or confounding, and usually increases variability.

Question 4

1 mark

Sampling methods

THE QUESTION

Which of these is a reason why a statistician may use stratified sampling instead of a simple random sample?
(A) To reduce the effects of confounding variables
(B) To introduce sample bias
(C) To reduce sample bias
(D) To reduce variability between possible samples
(E) To reduce question wording bias

Question 4 — solution

1 mark

Sampling methods

CONDENSED SOLUTION

- Answer: (D).
- By sampling every stratum, each stratum is represented in the right proportion, so estimates vary less from sample to sample.
- That is a reduction in the variability of the sampling distribution — not a change in bias.

TRAP (C) is close but wrong: stratified sampling is chosen for precision (less variability), and an SRS is already unbiased. Stratification does not remove bias from a study.

Question 8

1 mark

Sampling methods

THE QUESTION

The student council wants to survey the student body about a proposed spirit week theme. They randomly select 50 students from grade 9, 50 from grade 10, 50 from grade 11, and 50 from grade 12.

Which sampling method is this?
(A) simple random
(B) cluster random
(C) stratified random
(D) systematic with a random start
(E) convenience

Question 8 — solution

1 mark

Sampling methods

CONDENSED SOLUTION

- Answer: (C) stratified random sample.
- The population is first divided into strata that differ from one another — the four grade levels.
- Then a random sample is taken from EVERY stratum, 50 students in each.
- Both features must hold. An equal number drawn from each group is the giveaway.

TRAP Cluster sampling also uses groups, but there you select whole groups and sample everyone inside them — you do not sample from every group.

Question 10

1 mark

Sampling methods

THE QUESTION

The student council wants to survey the student body about a proposed spirit week theme. They randomly select 20 classrooms and survey everybody in those classrooms.

Which sampling method is this?
(A) simple random
(B) cluster random
(C) stratified random
(D) systematic with a random start
(E) convenience

Question 10 — solution

1 mark

Sampling methods

CONDENSED SOLUTION

- Answer: (B) cluster random sample.
- Groups are formed, 20 whole classrooms are chosen at random, and then every individual in the chosen classrooms is surveyed.
- The second stage takes an entire group, not a person — that is what makes it cluster.

TRAP Compare with Question 8: same setting, different structure. Ask whether students were chosen individually (stratified) or by the room they sit in (cluster).

Question 11

1 mark

Sampling methods

THE QUESTION

Which of the following is an advantage of using cluster random sampling over simple random sampling?
(A) It is less biased.
(B) It ensures every subgroup is included.
(C) It sometimes allows researchers to visit more subjects at lower cost.
(D) It reduces confounding.
(E) It always reduces variability.

Question 11 — solution

1 mark

Sampling methods

CONDENSED SOLUTION

- Answer: (C).
- Cluster sampling concentrates the work in a few locations, so travel and administration cost less and more subjects can be reached for the same budget.
- (B) describes stratified sampling, not cluster: cluster deliberately samples only some groups.

TRAP (A), (D) and (E) sound statistical but are false. Cluster sampling does not reduce bias or confounding, and it usually INCREASES variability compared with an SRS of the same size.

02

Studies and experiments

Observational study or experiment — and why blocking matters

- Questions 2, 3, 9 and 12
- The single test: did the researchers ASSIGN the treatment, or just watch what happened?

Key points · studies and experiments

One question decides the study type: did the researchers ASSIGN a treatment?

1	**The one test**	Treatments assigned by the researchers → experiment. Subjects already had it → observational study.
2	**Experiment**	Allows a cause-and-effect conclusion, because the researchers controlled who got what.
3	**Observational study**	Association only — never claim cause from a survey or an observational study.
4	**Selection vs assignment**	Random SELECTION supports generalising. Random ASSIGNMENT supports cause and effect. Never mix them.
5	**Blocking rule**	Block on the variable with the STRONGER association with the response. Blocking reduces variability, not bias.

Question 12

1 mark

Studies and experiments

THE QUESTION

Which of these characteristics of a study makes it an experiment rather than an observational study?
(A) Subjects are randomly selected from a population.
(B) Two or more groups are compared.
(C) Treatments are assigned to the subjects.
(D) Subjects already have the treatments built in.
(E) Results can be generalized to a larger population.

Question 12 — solution

1 mark

Studies and experiments

CONDENSED SOLUTION

- Answer: (C).
- A treatment is deliberately imposed on the subjects by the researchers.
- That single act is what separates an experiment from an observational study.

TRAP (A) and (E) are about random selection, which affects generalisation, not the type of study. (D) describes the opposite — that is an observational study.

Question 9

1 mark

Studies and experiments

THE QUESTION

Which of these is NOT true about a survey conducted on a random sample from a population?
(A) It can establish a causal relationship.
(B) The results can be generalized to a larger population.
(C) It is an observational study.
(D) Bias can be introduced in the way questions are asked.
(E) It can be used by researchers to make estimates about the population.

Question 9 — solution

1 mark

Studies and experiments

CONDENSED SOLUTION

- Answer: (A) is NOT true.
- A survey observes; it does not assign a treatment. Without random assignment you cannot conclude cause and effect.
- (B), (C), (D) and (E) are all true statements about a well-run random survey.

TRAP Confusing random selection with the power to prove cause. Good sampling buys generalisation, never causation.

Question 2

1 mark

Studies and experiments

THE QUESTION

A sports scientist is investigating whether a new athletic shoe improves running speed. Fifty runners volunteer. The scientist must decide whether to block by running experience (beginner, intermediate, advanced) or by resting heart rate before randomly assigning runners within each block to one of the two shoe types.

How should they decide?
(A) the variable most different among all participants
(B) the variable most similar among all participants
(C) the variable with the weaker association with running speed
(D) the variable with the stronger association with running speed
(E) randomly select the variable on which to block

Question 2 — solution

1 mark

Studies and experiments

CONDENSED SOLUTION

- Answer: (D). Block by the variable with the STRONGER association with the response.
- Running experience is strongly related to running speed, so blocking on it removes a large source of variation and makes the treatment comparison clearer.
- Blocking works by grouping similar subjects together, so that differences we see are more likely due to the treatment.

TRAP (A) and (B) use "most different" or "most similar" among participants, and (C) is backwards. The reason to block is a strong link to the RESPONSE variable.

Question 3

1 mark

Studies and experiments

THE QUESTION

A nutritionist wants to see which of two diet and exercise plans results in better improvement in cholesterol level. 80 participants volunteer. The nutritionist must decide whether to block by age group or by gender.

How should they decide?
(A) the variable with the weaker association with cholesterol level
(B) the variable with the stronger association with cholesterol level
(C) the variable most similar among all participants
(D) the variable most different among all participants
(E) randomly select the variable on which to block

Question 3 — solution

1 mark

Studies and experiments

CONDENSED SOLUTION

- Answer: (B). Block by the variable with the stronger association with cholesterol level.
- Age is strongly associated with cholesterol, so blocking by age group removes more of the variation between the groups.
- This is exactly the same rule as Question 2 with different wording.

TRAP Wording swaps. Do not be thrown by "age group or gender" — ask which one the response variable depends on more.

03

Summarising data

Variables, measurements, medians and count tables

- Questions 1, 5, 6, 7 and 13
- Read the graph or table before you read the options — the data decides

Key points · summarising data

What the data can and cannot tell you

1. **Quantitative** — A number you can do arithmetic with: height, time, number of accounts. You can find a mean, median and range.
2. **Categorical** — A label or group: favourite genre, eye colour, shirt size label. Count and compare percentages only.
3. **Fixed vs variable** — An unknown constant (the flagpole's true height) does NOT vary. Repeated measurements vary — that is measurement variability.
4. **Median from a table** — Cumulate the percentages until you cross 50%. The median lies in that class, never at a single exact value.
5. **Reading counts** — Check totals and convert to percentages before judging words like "majority" (more than half) or "one third".

Question 6

1 mark

Summarising data

THE QUESTION

A teacher collected some personal data from their students to use in class. Which of the following are quantitative variables?
I. Height
II. Favourite music genre
III. Number of social media accounts they have

(A) I only (B) II only (C) III only (D) I and III (E) None of these

Question 6 — solution

1 mark

Summarising data

CONDENSED SOLUTION

- Answer: (D) I and III.
- Height is a measurement — quantitative. Number of accounts is a count — quantitative.
- Favourite music genre is a category, so it is categorical, not quantitative.

TRAP A number you have coded for a category (for example 1 = rock, 2 = pop) is still categorical. Ask whether arithmetic makes sense.

Question 5

1 mark

Summarising data

THE QUESTION

A geometry teacher assigns groups to measure the height of the flagpole. One group uses three different techniques and takes the mean of the three results.

Which is NOT a correct statistical justification for that plan?
(A) The calculated heights are a variable.
(B) The actual height of the flagpole is a variable, and three measurements show its variability.
(C) The mean of the three is more likely close to the actual height than a single measurement.
(D) The variability in the measurements shows how far off they might be.
(E) If one measurement differs greatly, students will notice a possible problem.

Question 5 — solution

1 mark

Summarising data

CONDENSED SOLUTION

- Answer: (B) is NOT correct.
- The actual height of the flagpole is a fixed constant, not a variable. It does not vary between the three measurements.
- What varies is the MEASUREMENT, so repeated measurements show the variability of the measurement process — which is what (A), (C), (D) and (E) correctly describe.

TRAP The most common wrong idea in Unit 1 — treating a fixed, unknown quantity as if it were a variable.

Question 7

1 mark

Summarising data

THE QUESTION

A clothing company classifies men's shirt sizes by neck circumference. The percentage of men in each size range:

XS 33–35: 1% S 35–38: 13% M 38–40: 37% L 40–43: 26%
XL 43–45: 14% XXL 45–48: 6% 3XL 48–52: 2% 4XL 52–55: 1%

Which statement about the median neck size must be true?
(A) $38 \le \text{median} < 40$ (B) $40 \le \text{median} < 43$ (C) $43 \le \text{median} < 45$
(D) it is 39 cm (E) it is 43 cm

Question 7 — solution

1 mark

Summarising data

CONDENSED SOLUTION

- Answer: (A). Medians need cumulative percentages.
- Cumulative: XS 1%, S 14%, M 51%. The 50th percentile is reached inside the M band.
- Because the median falls in the interval 38 < size < 40, it must be at least 38 cm and less than 40 cm.
- The table gives no individual values, so the exact median cannot be found — only its interval.

TRAP (D) and (E) offer exact values. A grouped table can never pin the median to a single number, only to a class.

Question 13

1 mark

Summarising data

THE QUESTION

This is the same table as Question 7. Which statement about the median neck size must be true?

XS 33–35: 1% S 35–38: 13% M 38–40: 37% L 40–43: 26%
XL 43–45: 14% XXL 45–48: 6% 3XL 48–52: 2% 4XL 52–55: 1%

Question 13 — solution

1 mark

Summarising data

CONDENSED SOLUTION

- Answer: (A) again — the item is repeated in the booklet.
- Same working: cumulate to 1%, 14%, 51%. The 50% point is crossed inside the M class.
- So 38 cm ≤ median < 40 cm.

TRAP If you see a repeated question, treat it as a free mark. Check your earlier answer rather than re-deriving from scratch.

Question 1

1 mark

Summarising data

THE QUESTION

An international school has students from all over the world. Number of students by passport continent:

North America 605 South and Central America 152 Europe 95
Asia 912 Africa 98 Oceania 38

Which of the following statements is NOT supported by the table?
(A) There are 1,900 students at the school.
(B) More students have passports from South and Central America than from Africa and Oceania combined.
(C) Fewer than one third hold passports from a North American country.
(D) The majority of the students hold passports from Asia.
(E) More than one third hold passports from North America or Europe.

Question 1 — solution

1 mark

Summarising data

CONDENSED SOLUTION

- Answer: (D) is NOT supported.
- Total = 605 + 152 + 95 + 912 + 98 + 38 = 1,900, so (A) is true.
- Asia = 912 ÷ 1,900 = 48%, which is NOT a majority (that needs more than 50%), so (D) fails.
- (B) 152 > 98 + 38 = 136 ✓ (C) 605 ÷ 1,900 = 31.8% < 1/3 ✓ (E) (605 + 95) ÷ 1,900 = 36.8% > 1/3 ✓

TRAP "Majority" means more than half. 48% is the largest share but not a majority — the single most common reason students lose this mark.

04

A full inference FRQ

Question 14 — two-proportion z-test, from state to conclude

- One question, four full marks
- Write the state, plan, do, conclude cycle every single time

Key points · the inference routine

Write state · plan · do · conclude, every single time

1	**State**	Name the parameters, write H_0 and H_a in terms of them, and state α. "A difference" = two-sided.
2	**Plan**	Name the test, then check and SAY you checked the conditions: random assignment, and large counts for proportions.
3	**Do**	For two proportions use the POOLED proportion in the standard error, then find z and the p-value.
4	**Conclude**	Compare p with α, give the decision, then answer in CONTEXT: the adverts, the people, the proportions.
5	**What earns marks**	The conclusion needs all three parts: the decision, the link to the p-value, and the answer in context.

Question 14

4 marks

Full inference FRQ

THE QUESTION

Neil tested two adverts. 500 volunteers were randomly assigned: 250 watched version A and 250 watched version B.

Version A: 83 of 250 texted the code.
Version B: 54 of 250 texted the code.

At the 0.05 level of significance, complete the appropriate inference procedure to determine if there is convincing statistical evidence of a difference in the proportion of adults similar to those in the study who will watch version A and text the code, compared with the proportion who will watch version B and text the code.

Question 14 — solution

4 marks

Full inference FRQ

CONDENSED SOLUTION

- STATE: let pA and pB be the true proportions who text for each version. H_0: pA − pB = 0; H_a: pA − pB ≠ 0; α = 0.05.
- PLAN: two-sample z-test for a difference in proportions. Random assignment ✓. Large counts — pooled $\hat{p}$ = 137/500 = 0.274, giving 250(0.274) = 68.5 and 250(0.726) = 181.5, all ≥ 10 ✓.
- DO: $\hat{p}_A$ = 83/250 = 0.332, $\hat{p}_B$ = 54/250 = 0.216. SE = √(0.274(0.726)(2/250)) = 0.0399. z = (0.332 − 0.216)/0.0399 = 2.91. p-value ≈ 0.0036.
- CONCLUDE: 0.0036 < 0.05, so reject H_0. There is convincing evidence of a difference in the proportion who text for the two versions.

TRAP The conclusion needs all three parts: the decision, the link to the p-value, and the answer in context. Random assignment is what lets you speak about cause here.

05

Data collection plans

Observational study or experiment, generalisation, and advantages

- Questions 15, 16 and 17
- Same study, three different marks: classify it, judge it, then write the question

Key points · judging a data collection plan

Classify it, judge it, then write the question

1	**The three-part question**	Population + variable with units + comparison of interest. All three or lose the mark.
2	**Classify**	Look for “randomly assign”: assignment by researchers → experiment; groups they already belonged to → observational study.
3	**Generalise?**	Only if the sample was selected at random from that population. One gym or one region does not represent a whole country.
4	**Which is better?**	The experiment — random assignment supports a cause-and-effect conclusion, which observation never can.
5	**What loses marks**	Naming only the sample, asking about individuals instead of the population, or asserting the answer instead of asking.

Question 15

2 marks

Data collection plans

THE QUESTION

A study asks whether people who exercise with a training partner burn more calories on average than those who exercise alone.

Plan One: recruit volunteers, randomly assign each to partner or alone, record calories burned for six weeks.
Plan Two: take a random sample at one fitness centre, ask whether they usually exercise with a partner, and form the groups from their answers.

Part A
i. Identify whether Plan One is an observational study or an experiment.
ii. Identify whether Plan Two is an observational study or an experiment.

Question 15 — solution

2 marks

Data collection plans

CONDENSED SOLUTION

- i. Plan One: an EXPERIMENT (1 mark).
- The researchers randomly assign each participant to a treatment — partner or alone. Assignment is done by the researchers, so it is an experiment.
- ii. Plan Two: an OBSERVATIONAL STUDY (1 mark).
- Participants are grouped by a characteristic they already have. The researchers assign nothing.

TRAP The whole question turns on the words "randomly assign". If the subjects chose, or already belonged, it is observational.

Question 16

2 marks

Data collection plans

THE QUESTION

Part B

i. Explain whether using the sampling method in Plan Two would allow the researchers to generalise to all people in the country.

ii. Explain why Plan One may be more advantageous than Plan Two.

Question 16 — solution

2 marks

Data collection plans

CONDENSED SOLUTION

- i. No (1 mark). Plan Two samples only people at ONE fitness centre in the western region.
- That group is not representative of all people in the country — people who do not use gyms had no chance of being selected, and other regions are excluded.
- ii. Plan One uses RANDOM ASSIGNMENT (1 mark), so it can support a cause-and-effect conclusion about exercising with a partner and calories burned. Plan Two cannot, because group membership was not assigned.

TRAP Two different ideas are being asked side by side: generalisation (who was selected) in part i, and causation (who was assigned) in part ii. Do not swap them.

Question 17

2 marks

Data collection plans

THE QUESTION

Part C

The researchers want to replicate their study in the northern region of the country. Determine a valid investigative question for their study. Ensure all three components of a valid investigative question are included in your response.

Question 17 — solution

2 marks

Data collection plans

CONDENSED SOLUTION

- A valid investigative question has three parts: the population, the variable with its units, and the comparison of interest.
- Model answer: “For all people in the northern region who exercise, is there a difference between the average number of calories burned per workout by those who exercise with a training partner and the average number of calories burned per workout by those who exercise alone?”
- Marking: 1 mark for two components, 2 marks for all three.
- Check: population = people who exercise in the northern region; variable = average calories burned per workout; comparison = partner vs alone.

TRAP Do not name only the sample, and do not predict the answer ("do partner exercisers burn more?"). An investigative question asks; it does not assert.

06

The commuting survey

Sampling methods, generalisation, displays, shape and questions

- Questions 18 to 22 — one scenario, five parts
- Regions X, Y and Z survey adults about their commute to work

Key points · the commuting survey

Keep the three regions straight before answering

1	**Region X**	Every adult surveyed, chased with reminders → a CENSUS (complete enumeration).
2	**Region Y**	Six similar areas, two chosen at random, everyone in them surveyed → a CLUSTER sample.
3	**Region Z**	Only people who answered a previous survey → undercoverage plus self-selection, so biased. Higher response rate does not fix it.
4	**Why Y generalises**	The areas are similar AND two were chosen at random — a random, representative subset of the region.
5	**Five-number summary**	With only min, Q1, median, Q3, max available → side-by-side BOXPLOTS. Histograms and dotplots need the raw data.
6	**Judging shape**	Skewed right when mean > median and the upper half (Q3 − median) is much wider than the lower half (median − Q1).

Question 18

2 marks

The commuting survey

THE QUESTION

Part A

i. Identify the sampling method for Region X.

ii. Identify the sampling method for Region Y.

Region X: each adult who lives in the region receives a survey and must complete it by mail; reminders are sent until they do.

Region Y: the region is divided into six similar areas; two are randomly selected and each adult in those two areas is surveyed by mail.

Question 18 — solution

2 marks

The commuting survey

CONDENSED SOLUTION

- i. Region X — a CENSUS (a complete enumeration) (1 mark).
- Every adult in the region is asked to take part and reminders are sent until they respond, so an attempt is made to survey the whole population.
- ii. Region Y — a CLUSTER sample (two-stage cluster sample) (1 mark).
- The region is split into six areas, two areas are selected at random, and then every adult within those areas is surveyed.

TRAP No credit for "survey" or "mail survey" on their own — the mark is for naming the METHOD. Region Y is not stratified, because not all six areas were sampled.

Question 19

4 marks

The commuting survey

THE QUESTION

Part B

i. Explain why the results of the sampling method for Region Y can be generalised to the population of the region.

ii. Tom claims the sampling method for Region Z may lead to a higher response rate than the other regions and could better represent the entire population in Region Z. Explain why Tom's claim is not correct.

Region Z: each adult who lives in the region and has responded to a previous official survey is asked to complete the survey by mail.

Question 19 — solution

4 marks

The commuting survey

CONDENSED SOLUTION

- i. The six areas are similar to one another, so any one area is broadly representative of the region (1 mark), and two areas were chosen AT RANDOM (1 mark).
- Together these mean the selected adults form a random, representative subset, so the results generalise to all adults in the region.
- ii. Region Z samples only people who responded to an earlier survey (1 mark).
- People who did not respond have no chance of selection (undercoverage), and responders are self-selected, so they may differ systematically in commuting behaviour (1 mark). A higher response rate does not remove that bias.

TRAP Both parts need a named bias plus the consequence. "It is biased" alone earns nothing — say who is missing and why that matters.

Question 20

2 marks

The commuting survey

THE QUESTION

Part C
Manuel, an insurance agent in Region Y, wants to compare the average commute times to work for the populations in each of the two areas of Region Y. His first step is to construct a graphical representation for each area using the summary statistics in the table.

i. Identify a graphical representation that is appropriate to use.
ii. Explain why the representation from part C(i) is appropriate to use.

Question 20 — solution

2 marks

The commuting survey

CONDENSED SOLUTION

- i. Side-by-side BOXPLOTS, one for each area (1 mark).
- ii. A boxplot is built from exactly the five-number summary the table provides — minimum, Q1, median, Q3, maximum (1 mark).
- Commute time is quantitative, and two boxplots on a common scale make centre and spread easy to compare.

TRAP No credit for a histogram, stemplot or dotplot: those need the individual data values, which the table does not give.

Question 21

2 marks

The commuting survey

THE QUESTION

Part D
Use comparisons of the summary statistics in the table to describe the most likely shape for the distribution of commute time in Area Five.

Area Five: $\bar{x}$ 42, n 4,502, min 5, Q1 24, median 32, Q3 68, max 83

Question 21 — solution

2 marks

The commuting survey

CONDENSED SOLUTION

- Answer: skewed to the RIGHT (positively skewed).
- Mean (42) > median (32), so the mean is dragged up by unusually long commutes (1 mark).
- Upper half is much wider than the lower half: Q3 − median = 68 − 32 = 36, versus median − Q1 = 32 − 24 = 8, and the long right tail confirms it (1 mark).

TRAP You must give at least one numerical comparison. A bare "right skewed" earns only partial credit, and calling it symmetric is wrong.

Question 22

2 marks

The commuting survey

THE QUESTION

Part E

Manuel wants to perform a study about the average daily commute time to work for Area Three and Area Six in Region Y. He believes there is a difference in the commute times for the two areas. Determine a valid investigative question for Manuel's study. Ensure all three components of a valid investigative question are included in your response.

Question 22 — solution

2 marks

The commuting survey

CONDENSED SOLUTION

- Population: adults who commute to work in Area Three and Area Six of Region Y.
- Variable: average daily commute time to work, in minutes.
- Comparison: a difference between the two areas.
- Model answer: “For all adults who commute to work in Area Three and all adults who commute to work in Area Six of Region Y, is there a difference in the average daily commute time to work, in minutes, between the two areas?”
- Marking: 1 mark for two components, 2 marks for all three.

TRAP No credit if you name only one area, omit the units, or ask about individuals instead of the population average.

Mark scheme · multiple choice

One mark each, no partial credit, no penalty for a wrong answer

Q1	D	Counting and percentages from a table
Q2	D	Block on the variable most strongly related to the response
Q3	B	Same blocking rule, different wording
Q4	D	Stratified sampling reduces variability
Q5	B	A fixed constant is not a variable
Q6	D	Quantitative: measurement and count
Q7	A	Median class from cumulative percentages
Q8	C	Stratified: sample every stratum
Q9	A	A survey cannot establish causation
Q10	B	Cluster: sample whole groups
Q11	C	Cluster sampling lowers fieldwork cost
Q12	C	Treatments assigned means experiment
Q13	A	Repeat of Q7

Total: 13 marks. Suggested weighting: MCQ 1 mark each; free response as shown on the next slide.

Mark scheme · free response

Mark by part, award for reasoning

Q14	**4**	1 state (hypotheses and α) · 1 plan (test named, conditions checked) · 1 do (z and p-value) · 1 conclude (decision plus context)
Q15	**2**	1 mark for Plan One = experiment with a reason · 1 mark for Plan Two = observational study with a reason
Q16	**2**	1 mark for saying no, one fitness centre is not representative · 1 mark for random assignment allowing cause and effect
Q17	**2**	1 mark for two of the three components · 2 marks for population, variable with units and comparison all present
Q18	**2**	1 mark for census · 1 mark for cluster sample (accept two-stage cluster)
Q19	**4**	1 mark areas are similar · 1 mark random selection · 1 mark undercoverage or self-selection named · 1 mark consequence for representation
Q20	**2**	1 mark for side-by-side boxplots · 1 mark for the five-number-summary justification
Q21	**2**	1 mark for right skew · 1 mark for a numerical comparison (mean vs median, or the quartile half-widths)
Q22	**2**	1 mark for two of the three components · 2 marks for population, variable with units and comparison all present

Free response total: 20 marks. Lesson total: 33 marks (13 MCQ + 20 FRQ).

Answer key

Fast self-check across all 22 questions

MULTIPLE CHOICE

- Q1 D
- Q2 D
- Q3 B
- Q4 D
- Q5 B
- Q6 D
- Q7 A
- Q8 C
- Q9 A
- Q10 B
- Q11 C
- Q12 C
- Q13 A

FREE RESPONSE

- Q14 two-proportion z-test, $z = 2.91$, $p \approx 0.0036$, reject H_0
- Q15 Plan One experiment; Plan Two observational
- Q16 i. no, one fitness centre only; ii. random assignment gives causation
- Q17 three-component investigative question
- Q18 i. census; ii. cluster sample
- Q19 i. similarity + random selection; ii. undercoverage and self-selection
- Q20 i. side-by-side boxplots; ii. built from the five-number summary
- Q21 right-skewed (mean 42 > median 32)
- Q22 three-component investigative question

FIVE THINGS TO DO IN THE EXAM

Do these and you will not lose Unit 1 marks

- Name the method, not the activity. "Cluster sample", not "survey".
- Separate random selection from random assignment — generalisation versus causation.
- On a grouped table, cumulate the percentages to locate the median class.
- Write state, plan, do, conclude for every inference question, conditions included.
- Answer in context: the population, the variable, the units and the comparison.

Questions and condensed solutions drawn from the Unit 1 past-paper booklet.