AP STATISTICS · MIDTERM REVISION 2026 · VARIANT B

Describing Data

Displays, summaries and shape — a varied second set of Midterm-style questions with condensed solutions

12
questions

4
skills: centre, spread, shape, outliers

67
slides in total

A varied companion to Midterm Revision A — new wording, new numbers, new order

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.
Calculator	A dedicated section shows how to pull one-variable statistics and boxplots from a graphical calculator.

Today's roadmap

Four skills carry every Midterm data question

1

Centre and spread

Medians, IQR, five-number summaries and what each statistic tells you

2

Displays

Frequency tables, dotplots, histograms, stemplots and boxplots — reading them fast

3

Shape and skew

Bimodal, gaps, skew rules and the mean-versus-median comparison

4

Outliers

The 1.5 × IQR rule, applied with real numbers

Key topic · Describing Data

What this lesson covers, in one page

Displays	Frequency tables, dotplots, histograms and stemplots answer centre, spread and shape questions — read them with the same four questions.
Summaries	Median position (n+1)/2, IQR = Q3 − Q1, and the five-number summary: read Q1 and Q3 straight from tables and box edges.
Shape	Skew is named by the tail; mean above median signals right skew, and two clusters with an empty stretch are bimodal with a gap.
Outliers	The 1.5 × IQR fences decide — compute Q1 − 1.5×IQR and Q3 + 1.5×IQR before calling any value an outlier.

Vocabulary

Key terms with Chinese translations

frequency table	**频率表**	counts how often each category occurs
dotplot	**点图**	one dot per value; clusters and gaps show directly
histogram	**直方图**	bars give INTERVALS of values, never exact ones
stemplot	**茎叶图**	stems and leaves keep every individual value visible
boxplot	**箱线图**	the five-number summary drawn to scale
five-number summary	**五数概要**	Min · Q1 · Median · Q3 · Max
interquartile range	**四分位距**	IQR = Q3 − Q1, the spread of the middle 50%
outlier fence	**异常值界**	Q1 − 1.5×IQR and Q3 + 1.5×IQR
skewness	**偏度**	the tail names the skew — tail right, skew right
census	**普查**	every member of the population measured — no inference needed

S

Starter · Reading a summary table

What you must remember before the questions

- One quick recall question
- Median position, IQR and what the fences decide

Key points · reading a summary table

Centre, spread and the outlier rule in one table

1 **Median position** — Ordered data: the median sits at position (n+1)/2 — cumulate table frequencies until you cross half of n.

2 **IQR** — IQR = Q3 − Q1, the spread of the middle 50%. Quote Q1 and Q3 from the table, never recompute them.

3 **The fences** — Low fence Q1 − 1.5×IQR, high fence Q3 + 1.5×IQR — compute BOTH before flagging outliers.

4 **Shape in one look** — Mean above median → right skew; two clusters with an empty stretch between → bimodal with a gap.

Question S1

1 mark

Starter · Reading a summary table

THE QUESTION

The weekly rehearsal minutes of 84 members of a youth choir: Q1 = 18, Median = 33, Q3 = 52.

(i) State the interquartile range.
(ii) Between which two values do the middle 50% of rehearsal times lie, and how many members is that?

Question S1 — solution

1 mark

Starter · Reading a summary table

CONDENSED SOLUTION

- IQR = Q3 − Q1 = 52 − 18 = 34 minutes.
- The middle 50% of rehearsal times lie between Q1 = 18 and Q3 = 52 minutes.
- Half of 84 members is 42 — 42 choir members rehearsed between 18 and 52 minutes.
- Q1 and Q3 are quoted straight from the table; nothing else enters the subtraction.

TRAP The IQR is one subtraction of quoted quartiles. Re-deriving quartiles from the mean or inventing percentages is where marks go missing.

01

Reading displays

Frequency tables · dotplots · histograms · stemplots

- Questions 1 to 6
- Centre, spread and shape straight off the display

Key points · reading displays

Every display answers the same four questions in a different costume

1. **Centre before spread**

   Find the median first, then the IQR or range. The median is the (n+1)/2 position in ordered data — cumulate frequencies until you cross half of n.

2. **Frequency tables**

   Add the frequencies for any claim about totals or combinations. Compare frequencies directly — never invent percentages the table does not give.

3. **Dotplots and gaps**

   Two clear peaks = bimodal. A peak is a cluster of tall dots. A gap is an interval with NO dots between two clusters.

4. **Histograms**

   Read the bar heights and widths carefully. The median is the value where half the total area lies to each side — work from the tallest bars first.

5. **Stemplots**

   Back-to-back stemplots share the stems: leaves read outward from the stem. Check both sides before comparing median and range.

Formulae · centre, spread and outliers

formulae

Everything you need for Questions 1 to 6

Median position	ordered data → position (n + 1) / 2; cumulate table frequencies to find the class
Range	range = maximum − minimum
Interquartile range	IQR = Q3 − Q1 (middle 50% spread)
Outlier fences	low: Q1 − 1.5 × IQR high: Q3 + 1.5 × IQR
Skew rule of thumb	mean > median → skewed right; mean < median → skewed left

Outliers are flagged by the fences, not by "looks far away". Compute both fences before deciding.

Question 1

1 mark

Reading displays

THE QUESTION

The back-to-back stemplot shows tail lengths (mm) of 63 squirrels, grey and red (key: 4|3 = 43 mm).

Which statement is true?

(A) The median and range are both greater for grey squirrels.

(B) The median and range are both less for grey squirrels.

(C) The median is greater for red squirrels, and the range is greater for grey.

(D) The median is greater for grey squirrels, and the range is less.

BACK-TO-BACK STEMPLOT

RED (left)		GREY (right)
9 8	1	4 8
9 8 7 6 5 4 3 2 2 1 1 0	2	1 3 5 6 7 8 9
9 7 5 4 3 1 0	3	0 1 2 2 2 2 3 4 5 7
8 4 1	4	0 4 8
7 2	5	3 9
3	6	6
8 0	7	4
9 4	8	3
2	9	1
	10	
	11	2
	12	
	13	1
	14	9

key: 4 | 3 = 43 mm (red leaves read outward from the stem)

Question 1 — solution

1 mark

Reading displays

CONDENSED SOLUTION

- Answer: (A).
- Median: 31 grey squirrels, so the (31+1)/2 = 16th value is 33 mm. Red: 32 squirrels, median = average of 16th and 17th = (31 + 33)/2 = 32 mm. Grey median greater.
- Range: grey 14–149 mm (135); red 18–92 mm (74). Grey range greater.

TRAP A back-to-back stemplot holds two whole distributions. Count the leaves on each side FIRST, then locate each median — never eyeball it.

Question 2

1 mark

Reading displays

THE QUESTION

The dotplot shows points earned in one class of a language app: two lone dots at 0 and 1, a cluster from 3 to 7 (peaking at 5), an empty stretch from 8 to 9, and a second cluster from 10 to 14 (peaking at 12).
Which description of the distribution of points is best?
(A) Approximately normal
(B) Uniform
(C) Bimodal with a gap
(D) Skewed to the left with an outlier

Question 2 — solution

1 mark

Reading displays

CONDENSED SOLUTION

- Answer: (C).
- Two distinct clusters of dots — one around 3–7, one around 10–14 — make the distribution bimodal.
- Between the clusters (8–9) there are no dots at all: a clear gap.

TRAP Name both features you can see. “Bimodal with a gap” needs gap evidence, and the empty 8–9 stretch is exactly that.

Question 3

1 mark

Reading displays

THE QUESTION

The histogram summarizes the delay times (min) of 260 trains at a city station last month: 0–5 → 34; 5–10 → 118; 10–15 → 61; 15–20 → 27; 20–25 → 12; 25–30 → 5; 30–35 → 2; 35–40 → 1.

Based on the histogram, which statement must be true?

(A) The minimum delay is 0 minutes.

(B) The maximum delay is 40 minutes.

(C) The median delay is not greater than 10 minutes.

(D) The mean delay is between 5 and 10 minutes.

Question 3 — solution

1 mark

Reading displays

CONDENSED SOLUTION

- Answer: (C).
- Half of 260 is 130. Cumulating: 34 + 118 = 152 ≥ 130, so the median lies in the 5–10 minute bar — not greater than 10 minutes.
- (A) and (B) fail: histogram bars give INTERVALS, not exact endpoints. (D) is not forced: the long right tail of late trains pulls the mean above the median.

TRAP Histograms bin the data. Exact minimums, maximums and means can NEVER be read off a histogram — but the median's bar can, by cumulating.

Question 4

1 mark

Reading displays

THE QUESTION

The frequency table shows the one school club each student in a year group joined this term.

Club	Frequency
Chess	21
Drama	42
Coding	9
Art	15
Debating	30

Which claim is supported by the table?

(A) More students chose chess than any other club.

(B) Twice as many students chose drama as chose chess.

(C) The total number of students in the year group's clubs is 105.

(D) The combined number who chose coding and art is greater than the number who chose debating.

Question 4 — solution

1 mark

Reading displays

CONDENSED SOLUTION

- Answer: (B).
- Drama 42 is exactly twice Chess 21.
- (A) fails: drama (42) beats chess (21). (C) fails: the total is 21+42+9+15+30 = 117. (D) fails: coding + art = 24, less than debating's 30.

TRAP Every option here is a quick sum. Do each sum — never answer a table question from impressions.

Question 5

1 mark

Reading displays

THE QUESTION

The histogram shows battery life (h) for 52 phones of the same model: 0–2 → 4; 2–4 → 3; 4–6 → 2; 6–8 → 5; 8–10 → 6; 10–12 → 11; 12–14 → 9; 14–16 → 8; 16–18 → 4. Low bars sit on the left; the tallest bars pile up on the right.

Which statement is correct?

(A) The distribution is skewed to the right.

(B) The distribution is skewed to the left.

(C) The distribution appears to be normal.

(D) The distribution appears to be uniform.

Question 5 — solution

1 mark

Reading displays

CONDENSED SOLUTION

- Answer: (B).
- The long tail of shorter battery lives stretches to the LEFT; the bulk of the phones cluster at 10–18 hours.
- A tail on the low side is the definition of left skew.

TRAP Skew names the TAIL, not the peak. Tall bars on the right with a tail of small values = skewed left.

Question 6

1 mark

Reading displays

THE QUESTION

For a sample of 38 bean seedlings, the mean height is 3.6 cm and the standard deviation is 2.1 cm. A height cannot be negative. Which is most likely true about the heights in this sample?

(A) The heights are approximately normal because the sample size is 38, so the central limit theorem applies.

(B) The heights are approximately normal because the standard deviation is less than the mean.

(C) The heights are skewed to the right because the least possible height is within 2 standard deviations of the mean.

(D) The heights are skewed to the left because the least possible height is within 2 standard deviations of the mean.

Question 6 — solution

1 mark

Reading displays

CONDENSED SOLUTION

- Answer: (C).
- Heights cannot be negative. The least possible height, 0, sits at 3.6 ÷ 2.1 ≈ 1.71 SDs below the mean — within 2 SDs.
- A distribution squeezed against a boundary at zero while its mean sits well above is typically right-skewed.
- (A) misuses the CLT: it says nothing about the POPULATION's shape. (B) is a fallacy — SD < mean proves nothing.

TRAP When a mean is small and close to a hard boundary (zero), suspect right skew. The CLT is about sample MEANS, never about the raw data.

02

Boxplots and summary statistics

Five-number summaries · IQR · outlier fences

- Questions 7 to 10
- From a table of statistics or a drawn boxplot, extract exactly what is certain

Key points · boxplots and summaries

A boxplot is the five-number summary drawn to scale — nothing more

1. **What a boxplot shows** — Minimum, Q1, median, Q3, maximum (or fences and flagged outliers). Each segment holds 25% of the data — but segment WIDTH shows spread, not count.
2. **Percentile claims** — “The top 25% sit above Q3”. Compare Q3s and medians across plots by reading the box edges.
3. **IQR from statistics** — IQR = Q3 − Q1. Read Q1 and Q3 straight from the table and subtract — nothing else.
4. **Outlier test with numbers** — Compute both fences. Anything beyond them is an outlier. Count outliers by looking at the plotted points beyond the fences.
5. **Statistics vs plots** — A table of statistics can mislead: mean 129 with median 50 screams right skew. Cross-check mean against median for shape.

Formulae · five-number summaries

formulae

Everything you need for Questions 7 to 10

Five-number summary	Min · Q1 · Median · Q3 · Max — each box section holds 25% of the data
IQR	IQR = Q3 − Q1
Outlier fences	Q1 − 1.5 × IQR and Q3 + 1.5 × IQR
Mean vs median	mean much greater than median → right skew; mean much less → left skew
What is NOT shown	a boxplot cannot give the sample size, the individual values, or the mean

Q1 and Q3 are quoted straight from the summary table — never recompute them from the mean or SD.

Question 7

1 mark

Boxplots and summaries

THE QUESTION

Three flute ensembles (27 players each) took the same 40-point sight-reading test; the boxplots show their scores. Ensemble 1: whiskers 13–21, Q1 ≈ 15, median ≈ 17, Q3 ≈ 19, no outliers. Ensemble 2: whisker low 20, Q1 ≈ 22, median ≈ 24, Q3 ≈ 26, whisker top 27, outlier at ≈ 39. Ensemble 3: whiskers 24–33, Q1 ≈ 26.5, median ≈ 28.5, Q3 ≈ 31, outlier at ≈ 37.

Which statement is true?

(A) Ensemble 1 has the smallest interquartile range.

(B) Ensemble 2 has the smallest range of scores.

(C) The top 25% of scores in ensemble 2 are lower than the highest score in ensemble 1.

(D) At least 50% of the scores in ensemble 3 are higher than all of the scores in ensemble 1.

Question 7 — solution

1 mark

Boxplots and summaries

CONDENSED SOLUTION

- Answer: (D).
- Ensemble 3's median (≈28.5) is above Ensemble 1's maximum (≈21), so at least half of Ensemble 3's scores beat every Ensemble 1 score.
- (A) fails: Ensemble 1's IQR (≈4) TIES Ensemble 2's — not strictly the smallest. (B) fails: Ensemble 2 spans 20–39, the largest range of the three. (C) fails: Ensemble 2's top 25% sit above Q3 ≈ 26, which is higher than Ensemble 1's maximum of ≈ 21.

TRAP Compare medians and box edges ACROSS plots. “At least 50% above X” means: is the median of that plot above X?

Question 8

1 mark

Boxplots and summaries

THE QUESTION

Hours of streaming watched last week by a random sample of subscribers: N = 132, Mean = 41.7, Median = 26.0, StDev = 63.9, SE Mean = 5.6, Minimum = 0.0, Maximum = 2400.0, Q1 = 12.0, Q3 = 54.0.

Which statement is true?

(A) Seventy-five percent of the subscribers watch more than 54 hours.

(B) Fifty percent of the subscribers watch between 0 and 41.7 hours.

(C) The interquartile range of the hours watched is 42.

(D) The distribution of hours watched contains outliers on both the low side and the high side.

Question 8 — solution

1 mark

Boxplots and summaries

CONDENSED SOLUTION

- Answer: (C).
- IQR = Q3 − Q1 = 54.0 − 12.0 = 42.
- (A) fails: 75% watch more than Q1 = 12, not more than Q3 = 54 — only 25% exceed 54. (B) fails: 50% sit between Q1 = 12 and Q3 = 54; the median 26, not the mean 41.7, splits the sample.
- (D) fails: fences are 12 − 1.5·42 = −51 (no low outlier possible, hours ≥ 0) and 54 + 63 = 117. The 2400 is a high outlier only.

TRAP Q1 and Q3 carry percentile meaning: below Q1 = 25%, between Q1 and Q3 = 50%, above Q3 = 25%. The mean never splits the data.

Question 9

1 mark

Boxplots and summaries

THE QUESTION

Times (min) 41 students spent on a phone app in one 75-minute revision session: Mean = 7.912, StDev = 7.015, Q1 = 2.5, Q3 = 10.5. The histogram shows most students between 0 and 14 minutes, a few values out to 35.

Using the 1.5 × IQR rule, which conclusion about the median and outliers is correct?

(A) The median is greater than 7.912, and there are exactly two outliers.

(B) The median is greater than 7.912, and there are exactly four outliers.

(C) The median is less than 7.912, and there are exactly two outliers.

(D) The median is less than 7.912, and there are exactly four outliers.

Question 9 — solution

1 mark

Boxplots and summaries

CONDENSED SOLUTION

- Answer: (D).
- Median: the histogram piles up at small times, so the median is pulled below the mean of 7.912.
- Outliers: IQR = 10.5 − 2.5 = 8. Upper fence = 10.5 + 1.5·8 = 22.5. The histogram shows exactly four plotted values beyond 22.5 (at 25, 27, 31 and 35); nothing lies below the lower fence of −9.5. So exactly four outliers.

TRAP Compute the fence FIRST (10.5 + 1.5×8 = 22.5), then count the plotted values beyond it. And right-skew puts the median BELOW the mean.

Question 10

1 mark

Boxplots and summaries

THE QUESTION

The bar chart shows the relative frequency of households supporting a council food-waste collection plan in five zones: Zone 1 ≈ 0.36; Zone 2 ≈ 0.68; Zone 3 ≈ 0.52; Zone 4 ≈ 0.75; Zone 5 ≈ 0.43.

Which statement about the zones is supported by the data?

(A) Zone 1 has the fewest number of households that support the plan.

(B) Zone 4 has the greatest percentage of households that support the plan.

(C) In three zones, fewer than half of the households support the plan.

(D) All zones show at least 60% of households in support of the plan.

Question 10 — solution

1 mark

Boxplots and summaries

CONDENSED SOLUTION

- Answer: (B).
- Zone 4's bar (0.75) is the tallest — the greatest percentage in support.
- (A) confuses a PERCENTAGE with a COUNT: Zone 1 has the lowest rate, but without zone sizes we cannot know the fewest households. (C) fails: only Zones 1 and 5 sit below 0.5. (D) fails: Zones 1 and 5 are below 60%.

TRAP Relative frequency is a rate, not a count. Any option claiming “fewest number” from a percentage chart is a trap.

03

Census and variability

Questions 11 and 12

- What a census changes
- How sample size moves the variance of a sampling distribution

Key points · census and variability

Who needs inference — and what sample size touches

1. **Census**

   Every member of the population is measured. The proportion is known exactly — no inference procedure is needed.

2. **Why inference exists**

   Samples leave sampling uncertainty. Confidence intervals and tests quantify that uncertainty — a census has none.

3. **Sample size lever**

   Variance of the sampling distribution of $\hat{p}$ is $p(1-p)/n$: smaller n → MORE variance in the sampling distribution.

4. **What never changes**

   The population and the individual samples are untouched by n — only the sampling distribution's spread moves.

Question 11

1 mark

Census and variability

THE QUESTION

A clinic surveyed all 48 of its staff members to determine the proportion who walk or cycle to work. Which statement is true?

(A) The clinic should not use the data because this is an observational study.

(B) The clinic did not select a random sample, so the survey will not provide useful information.

(C) The clinic must construct a confidence interval to estimate the proportion of staff who walk or cycle to work.

(D) The clinic needs no inference procedure to determine the proportion, because the survey was a census of all staff.

Question 11 — solution

1 mark

Census and variability

CONDENSED SOLUTION

- Answer: (D).
- All 48 staff were surveyed — a census. The sample IS the whole population, so the computed proportion needs no inference.
- (A) is nonsense: observational studies still produce valid data. (B) fails: a census gives complete information. (C) fails: confidence intervals estimate unknown population values — nothing is unknown here.

TRAP Inference exists because samples leave uncertainty. A census has no sampling error, so no inference procedure is needed.

Question 12

1 mark

Census and variability

THE QUESTION

To estimate the proportion of subscribers who watch documentaries weekly, a streaming service reduces its planned random sample from 3,200 to 1,900 subscribers.

What is the effect of this reduction?

(A) The variance of the sample will increase.

(B) The variance of the population will decrease.

(C) The variance of the sampling distribution of the estimator will increase.

(D) The variance of the sampling distribution of the estimator will decrease.

Question 12 — solution

1 mark

Census and variability

CONDENSED SOLUTION

- Answer: (C).
- A smaller sample carries more sampling uncertainty: the variance of the sampling distribution of the estimator grows (for $\hat{p}$: $p(1-p)/n$ gets bigger when n shrinks).
- (A) and (B) are about the sample and the population themselves — neither changes with the sample size.

TRAP Sample size only ever touches the SAMPLING DISTRIBUTION: smaller n → wider sampling distribution → more variance.

Question 13

6 marks

Data description FRQ · consolidated

THE QUESTION

Workout session lengths (min) for 45 gym members in one week: Min = 1, Q1 = 6, Median = 10, Q3 = 17, Max = 38, Mean = 13.4.

(a) Find the interquartile range and interpret it in context.
(b) Using the 1.5 × IQR rule, is the maximum of 38 an outlier?
(c) Justify the claim that the distribution is skewed to the right.
(d) Which measures of centre and spread best describe this distribution? Justify.

Question 13 — solution

6 marks

Data description FRQ · consolidated

CONDENSED SOLUTION

- (a) IQR = 17 − 6 = 11 (1). Interpretation: the middle 50% of sessions lasted between 6 and 17 minutes, a spread of 11 minutes (1).
- (b) Upper fence = Q3 + 1.5×IQR = 17 + 16.5 = 33.5. Since 38 > 33.5, the maximum IS an outlier (2).
- (c) Mean 13.4 > median 10, and the maximum (38) lies much further from the median than the minimum (1): a long tail of long sessions — right skew (1).
- (d) Median and IQR: the outlier and the skew make the mean and SD unreliable summaries here (1).

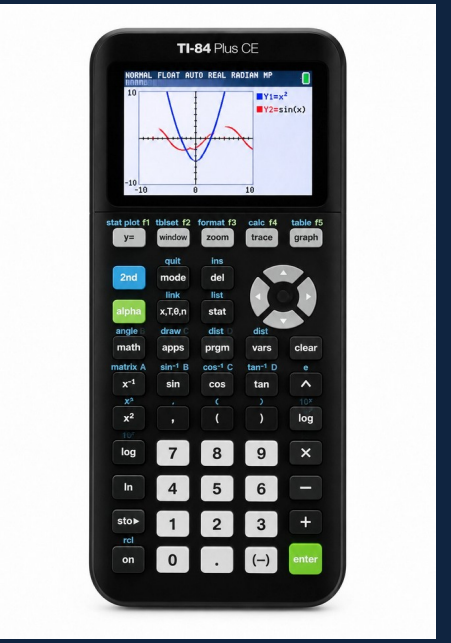
TRAP Quote both fences before declaring an outlier, and justify skew with mean-versus-median or tail distance — never with a bare “it looks skewed”.

04

Using your graphical calculator

One-variable statistics and boxplots from a data list

Why you need it: questions in this lesson hand you summary statistics — the calculator produces them from the raw data list, and the same list gives you an instant boxplot for outlier checks.



Calculator · 1-Var Stats and boxplots

TI-84 family

When you need it, and the exact key presses

WHEN YOU NEED 1-VAR STATS

- A question gives raw data and asks for centre or spread: $\bar{x}$, Sx, the median, Q1 and Q3 all come from one command.
- 1-Var Stats returns the whole five-number summary: minX, Q1, Med, Q3, maxX — plus $\bar{x}$ and Sx.
- Outlier checks: read Q1 and Q3 off the screen, compute the 1.5 × IQR fences by hand.
- Boxplot: StatPlot with the modified box icon plots the summary and flags outliers beyond the fences automatically.

THE KEY PRESSES

STAT	then ENTER on 1: Edit
L_1	type the data values
STAT	then ▶ CALC
1: 1-Var Stats	then ENTER
L_1	the list to summarize
ENTER	$\bar{x}$, Sx, Med, Q1, Q3 appear

1-Var Stats

Calculator · worked example

The gym FRQ — summary statistics from the raw list

WORKED EXAMPLE

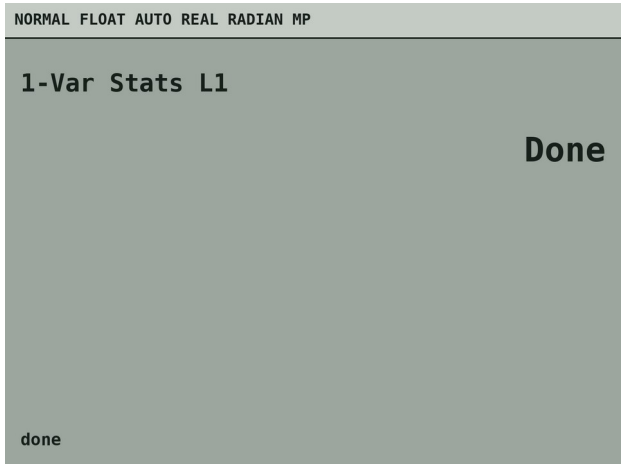
- Question 13's gym recorded the workout minutes of 45 members (list L_1).
- • STAT → ENTER → type the 45 values into L_1
- • STAT → ▶ CALC → 1: 1-Var Stats → L_1

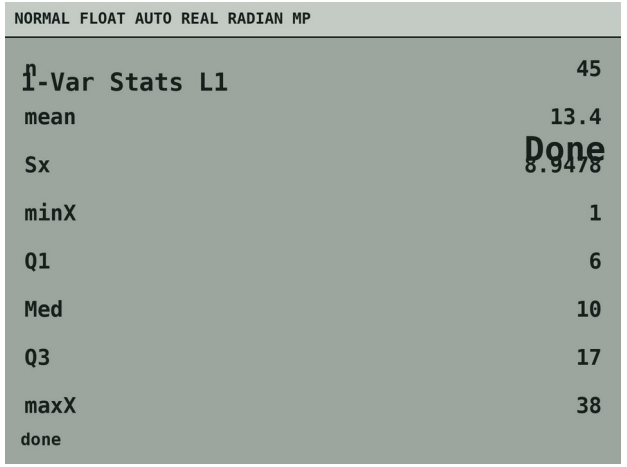
Result: $\bar{x}$ = 13.4, Q1 = 6, Med = 10, Q3 = 17, maxX = 38 — the exact statistics the question quotes.

Outlier check by hand: IQR = 17 − 6 = 11, upper fence = 17 + 1.5 × 11 = 33.5, and 38 > 33.5 — an outlier.

StatPlot with the modified-box icon draws the same conclusion: the plotted 38 sits beyond the fence whisker.

Casio fx-CG50: Stat mode → 1-Var over the list column — same outputs.





Calculator · key points for 1-Var Stats

Four habits that keep the summary right

1. **One command, everything** — 1-Var Stats gives $\bar{x}$, Sx, n, minX, Q1, Med, Q3 and maxX in one screen — quote them, do not recompute.
2. **Sx versus σx** — Sx is the SAMPLE standard deviation (divide by n − 1) — the one AP Statistics uses.
3. **Fences by hand** — The calculator flags outliers on its boxplot, but the written fence computation Q3 + 1.5×IQR is what earns the mark.
4. **Median is on the screen** — Med comes straight from 1-Var Stats — never average anything to find it when the command is available.

Mark scheme · multiple choice

1 mark each

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

Q1	A	Grey median 33 > red 32; grey range 135 > red 74
Q2	C	Two clusters (3–7, 10–14) with empty 8–9 between
Q3	C	Cumulate 34 + 118 = 152 ≥ 130 → median bar ≤ 10 min
Q4	B	Drama 42 = 2 × Chess 21 — exactly twice
Q5	B	Tail of short battery lives on the left → skewed left
Q6	C	Zero is within 2 SD of mean 3.6, SD 2.1 → right skew
Q7	D	Ensemble 3 median ≈28.5 above Ensemble 1 max ≈21
Q8	C	IQR = 54 − 12 = 42
Q9	D	Fence 10.5 + 1.5·8 = 22.5 → four high outliers
Q10	B	Zone 4's 0.75 is the greatest percentage
Q11	D	Census of all 48 staff — no inference needed
Q12	C	Smaller n → more variance in the sampling distribution

Multiple choice total: 12 marks. FRQ Q13: 6 marks.

Mark scheme · free response

6 marks

Q13 — mark by part, award for reasoning

Q13a	**2**	1 for IQR = 11 · 1 for the in-context interpretation (middle 50% between 6 and 17 minutes)
Q13b	**2**	1 for upper fence 17 + 1.5×11 = 33.5 · 1 for the outlier decision (38 > 33.5, yes)
Q13c	**1**	1 for a valid skew justification: mean 13.4 > median 10, or tail distance comparison
Q13d	**1**	1 for median and IQR, justified by the skew/outlier

Free response total: 6 marks. Lesson total: 18 marks (12 MCQ + 6 FRQ).

Answer key

One line per question

ANSWERS

- Q1 (A) — grey median 33 > red 32; grey range 135 > red 74
- Q2 (C) — two clusters with an empty 8–9 between
- Q3 (C) — 34 + 118 = 152 ≥ 130, median bar ≤ 10 min
- Q4 (B) — drama 42 is exactly twice chess 21
- Q5 (B) — tail of short battery lives on the left
- Q6 (C) — zero within 2 SD of mean 3.6 (SD 2.1) → right skew
- Q7 (D) — Ensemble 3 median above Ensemble 1 maximum

FREE RESPONSE

- Q8 (C) — IQR = 54 − 12 = 42
- Q9 (D) — fence 22.5 → exactly four high outliers
- Q10 (B) — Zone 4's 0.75 is the greatest percentage
- Q11 (D) — census of all 48 staff, no inference needed
- Q12 (C) — smaller n, more variance in the sampling distribution
- Q13 IQR = 11 · fence 33.5 → 38 is an outlier · mean 13.4 > median 10, right skew · median and IQR

True or false · 1

T / F

Decide, then justify your decision — answer on the next slide

In a frequency table, “twice as many chose drama as chess” can be checked by dividing one frequency by the other.

True or false · 1 — answer

TRUE

TRUE

42 ÷ 21 = 2, so the claim is supported. Frequency comparisons are simple divisions — check them numerically, never by impression.

True or false · 2

T / F

Decide, then justify your decision — answer on the next slide

A histogram can be used to read off the exact minimum value of a data set.

True or false · 2 — answer

FALSE

FALSE

Histogram bars only give intervals. The exact minimum, maximum and mean cannot be recovered from a histogram — only the median's bin can be located.

True or false · 3

T / F

Decide, then justify your decision — answer on the next slide

A distribution with two clear clusters of dots separated by an empty interval is described as bimodal with a gap.

True or false · 3 — answer

TRUE

TRUE

Two peaks = bimodal; the empty stretch between them is the gap — exactly the app-points dotplot in the lesson.

True or false · 4

T / F

Decide, then justify your decision — answer on the next slide

If the median of a data set is greater than its mean, the distribution is most likely skewed to the right.

True or false · 4 — answer

FALSE

FALSE

Median above the mean signals a long tail of SMALL values — left skew. Right skew pulls the median below the mean.

True or false · 5

T / F

Decide, then justify your decision — answer on the next slide

The IQR is Q3 − Q1; for Q1 = 12 and Q3 = 54 it equals 42.

True or false · 5 — answer

TRUE

TRUE

54 − 12 = 42. The IQR is the spread of the middle 50% — no other statistic enters the subtraction.

True or false · 6

T / F

Decide, then justify your decision — answer on the next slide

Using the 1.5 × IQR rule with Q1 = 2.5 and Q3 = 10.5, values above 22.5 are outliers.

True or false · 6 — answer

TRUE

TRUE

Upper fence = 10.5 + 1.5×8 = 22.5. The revision-session data put exactly four observations beyond it.

True or false · 7

T / F

Decide, then justify your decision — answer on the next slide

If a clinic surveys all 48 of its staff, it must construct a confidence interval to estimate the proportion who cycle to work.

True or false · 7 — answer

FALSE

FALSE

A census measures the whole population, so the proportion is known exactly — inference procedures exist only to handle sampling uncertainty.

True or false · 8

T / F

Decide, then justify your decision — answer on the next slide

Reducing a planned sample from 3,200 subscribers to 1,900 increases the variance of the sampling distribution of the estimator.

True or false · 8 — answer

TRUE

TRUE

Variance of $\hat{p}$ is $p(1-p)/n$: smaller n means larger variance. The population and the individual samples never change.

FIVE THINGS TO DO IN THE EXAM

Do these and you will not lose data-description marks

- Cumulate frequencies to (n+1)/2 before locating any median.
- Histograms give intervals, never exact maximums or means.
- Skew is named by the tail: tail right = right skew; mean above median = right skew.
- Compute both fences, Q1 − 1.5·IQR and Q3 + 1.5·IQR, before calling anything an outlier.
- A census needs no inference; smaller samples mean more sampling variance.

Varied questions in the style of Midterm 2026 versions A and B.