AP STATISTICS · UNIT 5 REVISION

Unit 5 Revisions b

Hypotheses and test selection — paired vs two-sample, one-sided vs two-sided

9
multiple choice questions

2
free-response scenarios

71
slides in total

Built from the Unit 5 Revisions b past-paper booklet (second variant)

How this lesson runs

Key points, then questions to try — solution on the very next slide

Starter	One quick recall question on parameters and statistics — attempt it before the reveal.
Key points	Each section opens with a summary of the ideas you need, main point highlighted first.
Question	Every question appears on its own slide. Attempt each before turning the page.
Solution	The next slide gives the condensed solution, and every solution ends with a trap box.
True or false	The lesson ends with eight statements — one per slide, answer on the slide that follows.

Key topic · Unit 5 revision

What this lesson covers, in one page

Writing hypotheses	Hypotheses are written in parameters, never statistics: H_0: μ = value, and the direction of H_a comes from the words of the question.
Paired or two-sample?	The same unit measured twice (before and after, twice on each person) → matched pairs. Two independent groups → two-sample t.
Conducting the test	Compute the statistic, get the p-value, and compare it with α: reject H_0 only when $p < \alpha$.
Concluding in context	A conclusion names the parameter, the groups and the direction — and a large p-value is fail to reject, never proof of H_0.

Vocabulary

Key terms with Chinese translations

null hypothesis	**原假设**	H_0: the claim of no difference or no change being tested
alternative hypothesis	**备择假设**	H_a: what the evidence is trying to support
population mean	**总体均值**	μ — the parameter that hypotheses are written in
matched pairs	**配对设计**	the same unit measured twice, or naturally paired units
two-sample t-test	**两样本 t 检验**	compares the means of two independent groups
one-sided test	**单侧检验**	H_a uses $>$ or $<$ — a direction is claimed
two-sided test	**双侧检验**	H_a uses $\neq$ — any difference counts
parameter	**参数**	a number describing the whole population, such as μ
statistic	**统计量**	a number computed from the sample, such as $\bar{x}$
significance level	**显著性水平**	α — the p-value must beat it to reject H_0

S

Starter · Parameters and statistics

What you must remember before the questions

- One quick recall question
- Hypotheses are written in parameters — never in statistics

Key points · parameters and statistics

Which numbers describe populations, which describe samples

1	**Parameter**	A parameter describes the WHOLE population: μ, the population mean. Unknown — that is why we test for it.
2	**Statistic**	A statistic is computed from the SAMPLE: $\bar{x}$, the sample mean. Known, and it estimates the parameter.
3	**The rule**	Hypotheses are always written in parameters: H_0: μ = 38, never H_0: $\bar{x}$ = 38.
4	**Why it matters**	Every hypothesis question in this booklet tests whether you know which symbol to use.

Question S1

1 mark

Starter · Parameters and statistics

THE QUESTION

An airline claims that the mean flight time between City X and City Y is 38 minutes. After taking many flights, a local business group calculates a sample mean flight time of 41 minutes.
In one or two sentences: which of 38 and 41 is a parameter, which is a statistic, and which of the two should hypotheses be written about?

Question S1 — solution

1 mark

Starter · Parameters and statistics

CONDENSED SOLUTION

- 38 minutes is the PARAMETER: the claimed mean flight time for the whole population of flights, written μ.
- 41 minutes is the STATISTIC: the mean of the flights the group actually took, written $\bar{x}$.
- Hypotheses are written about the parameter: H0: μ = 38 versus Ha: μ > 38.
- The sample mean 41 is evidence, never part of the hypotheses.

TRAP **Writing** $H_a: \bar{x} > 38$ is the classic error. Hypotheses describe populations, not samples — the statistic appears only in the calculation.

01

Writing hypotheses

Questions 1, 2 and 3

- Hypotheses in μ, never in $\bar{x}$
- The direction of Ha comes from the words

Key points · writing hypotheses

Three questions decide every hypothesis answer

1. **Use μ**

   Hypotheses use the parameter μ, never the statistic $\bar{x}$. The safe threshold, not the sample mean, is the value tested.

2. **Read the direction**

   Greater, less, different: the words of the question fix H_a. “Greater than” → $H_a: \mu > \text{value}$; “different” → $H_a: \mu \neq \text{value}$.

3. **Safe threshold trap**

   When a level is called safe at 24.5 ppb, concern about unsafe amounts means $H_a: \mu > 24.5$ — never the sample mean 25.3.

4. **Paired designs**

   When each pair gives one difference, the parameter is μ_d, the mean difference: $H_0: \mu_d = 0$.

Question 1

1 mark

Writing hypotheses · the airline claim

THE QUESTION

An airline claims that the mean flight time between City X and City Y is 38 minutes. After taking many flights, a local business group believes that the claim is unrealistic and that the actual mean flight time is greater than 38 minutes. If the group conducts a study to investigate its belief, which of the following hypotheses should be tested?

(A) $H_0 : \bar{x} = 38$, versus $H_a : \bar{x} < 38$

(B) $H_0 : \bar{x} = 38$, versus $H_a : \bar{x} > 38$

(C) $H_0 : \mu = 38$, versus $H_a : \mu < 38$

(D) $H_0 : \mu = 38$, versus $H_a : \mu > 38$

Question 1 — solution

1 mark

Writing hypotheses · the airline claim

CONDENSED SOLUTION

- The claim is about the mean flight time for ALL flights — a population mean μ, not a sample mean.
- "Greater than 38 minutes" fixes the direction: Ha : μ > 38.
- Answer: (D) — H0 : μ = 38 versus Ha : μ > 38.
- (A) and (B) fail twice: wrong symbol AND the statistic. (C) has the right symbol but the wrong direction.

TRAP **Two checks every time: the symbol is μ, and the direction word ("greater" here) matches the inequality in H_a.**

Question 2

1 mark

Writing hypotheses · the safe threshold

THE QUESTION

Perchlorate is a chemical used in rocket fuel. People who live near a former rocket-testing site are concerned that perchlorate is present in unsafe amounts in their drinking water. Drinking water is considered safe when the average level of perchlorate is 24.5 parts per billion (ppb) or less. A random sample of 28 water sources in this area produces a mean perchlorate measure of 25.3 ppb. Which of the following is an appropriate alternative hypothesis that addresses their concern?

(A) $H_a : \mu < 25.3$

(B) $H_a : \mu > 25.3$

(C) $H_a : \mu < 24.5$

(D) $H_a : \mu > 24.5$

Question 2 — solution

1 mark

Writing hypotheses · the safe threshold

CONDENSED SOLUTION

- Water is safe when the AVERAGE level is 24.5 ppb or less — so 24.5 is the tested value, the threshold.
- “Unsafe amounts” means the true mean exceeds the safe level: Ha : $\mu > 24.5$.
- Answer: (D). The 25.3 is the sample mean — a statistic — and never appears in a hypothesis.
- (A), (B) and (C) each fail one way: sample mean as the value, or the wrong direction.

TRAP **The number in the hypothesis is the SAFE THRESHOLD from the question, not the sample mean. Any option built on 25.3 is a trap.**

Question 3

1 mark

Writing hypotheses · identical twins

THE QUESTION

A study was conducted using data collected on the birth weights of a random sample of 10 pairs of identical twins to determine whether the twin born first tends to weigh more than the twin born second. Let μ_F represent the average birth weight of all twins born first, μ_s represent the average birth weight of all twins born second, and μ_D represent the average difference in birth weight (weight of first minus weight of second) for all pairs of twins. Which of the following would be the null and alternative hypotheses for this study?

(A) $H_0 : \mu_F = \mu_s$ and $H_a : \mu_F \neq \mu_s$

(B) $H_0 : \mu_D = 0$ and $H_a : \mu_D > 0$

(C) $H_0 : \mu_F - \mu_s = \mu_D$ and $H_a : \mu_F - \mu_s > \mu_D$

(D) $H_0 : \mu_F - \mu_s = \mu_D$ and $H_a : \mu_F - \mu_s \neq \mu_D$

Question 3 — solution

1 mark

Writing hypotheses · identical twins

CONDENSED SOLUTION

- The data are PAIRS: each pair produces one difference, first minus second — so the natural parameter is μD, the mean difference.
- "Weighs more" gives the direction: the difference is positive, Ha : μD > 0.
- Answer: (B) — H0 : μD = 0 and Ha : μD > 0.
- (C) and (D) treat μD as a number to subtract, which it is not — it IS the parameter being tested.

TRAP **For paired data, the test is about** $\mu_D = 0$ or not — never about $\mu_F - \mu_s$ versus μ_D. And "first weighs more" is one-sided: >, not ≠.

02

Choosing the test

Questions 4, 5, 6, 7 and 9

- Paired or two independent samples?
- One-sided or two-sided, t or z?

Key points · choosing the test

Design first, direction second

1. **Same unit twice** — Before and after, twice on each person, sitting then standing → MATCHED PAIRS with μ_d. One sample of differences.

2. **Two independent groups** — Different individuals in each group, no natural pairing → two-sample t-test for $\mu_1 - \mu_2$.

3. **Direction from words** — “Lasts longer”, “will be reduced” → one-sided. “Is there a difference” → two-sided.

4. **Means, not proportions** — The data are measurements (hours, scores, seconds, pulse rates) → t-tests for means. z-tests are for proportions; chi-square is for counts in categories.

Question 4

1 mark

Choosing the test · Brand A vs Brand B batteries

THE QUESTION

A manufacturer claims its Brand A battery lasts longer than its competitor's Brand B battery. Nine batteries of each brand are tested independently, and the hours of battery life are recorded for each battery. Provided that the assumptions for inference are met, which of the following tests should be conducted to determine if Brand A batteries do, in fact, last longer than Brand B batteries?

(A) A one-sided, paired t-test
(B) A one-sided, two-sample t-test
(C) A two-sided, two-sample t-test
(D) A one-sided, two-sample z-test

Question 4 — solution

1 mark

Choosing the test · Brand A vs Brand B batteries

CONDENSED SOLUTION

- Nine Brand A batteries and nine Brand B batteries are DIFFERENT batteries — two independent samples.
- No battery is tested twice → not paired. Independent samples → two-sample t-test.
- “Lasts longer” claims a direction → one-sided.
- Answer: (B) — a one-sided, two-sample t-test. (D) is wrong because the data are means of measurements, not proportions.

TRAP **Independent groups can never be a paired test, however tempting. And “longer” in the question is your signal for one-sided.**

Question 5

1 mark

Choosing the test · talkative vs quiet children

THE QUESTION

A study compared the language skills of two groups of 24-month-old children. One group consisted of children identified as talkative, and the other group consisted of children identified as quiet. Scores on a language test:
Talkative: 75 70 70 65 85 85 80 90 90 60
Quiet: 80 75 65 70 90 90 75 85 75 80
Assuming that it is reasonable to regard the groups as simple random samples and that the other conditions for inference are met, what statistical test should be used to determine if there is a significant difference in the average test score of talkative and quiet children at the age of 24 months?
(A) A chi-square goodness-of-fit test
(B) A chi-square test of independence
(C) A matched-pairs t-test for means
(D) A two-sample t-test for means

Question 5 — solution

1 mark

Choosing the test · talkative vs quiet children

CONDENSED SOLUTION

- Two DIFFERENT sets of children: the talkative group and the quiet group — independent samples.
- The parameter of interest is a difference between two MEANS of test scores.
- Answer: (D) — a two-sample t-test for means.
- Chi-square tests count how many fall in categories; scores are measurements, so (A) and (B) do not apply. (C) would need each child in both groups.

TRAP **Match the test to the DATA TYPE: measurements of a score → t-test for means. Counts in categories → chi-square. Proportions → z.**

Question 6

1 mark

Choosing the test · swimmers shaving heads

THE QUESTION

An amateur men's swimming association is trying to decide whether times in the 100-meter breaststroke will be reduced if the men shave their heads. From the population of swimmers, six were selected at random and agreed to swim two races — one before and one after shaving their heads. Times (seconds), swimmers I to VI:

Before (x_1): 66.2 67.5 64.3 68.0 67.2 63.9

After (x_2): 65.9 67.0 66.3 67.1 67.5 64.2

Let μ_1 represent the population mean of swimming times before shaving and μ_2 the population mean after shaving. Let μ_d represent the population mean of the differences, $x_1 - x_2$. Which of the following would be the most appropriate test and alternative hypothesis to use in testing this theory?

(A) A two-sample t-test with $H_a : \mu_1 > \mu_2$

(B) A two-sample t-test with $H_a : \mu_1 < \mu_2$

(C) A paired t-test with $H_a : \mu_d > 0$

(D) A paired t-test with $H_a : \mu_d \neq 0$

Question 6 — solution

1 mark

Choosing the test · swimmers shaving heads

CONDENSED SOLUTION

- Each swimmer races TWICE — before and after — so the samples are not independent: matched pairs.
- The parameter is μd, the mean of (before − after). Times REDUCED means before > after, so Ha : μd > 0.
- Answer: (C) — a paired t-test with Ha : μd > 0.
- (D) is two-sided, but “will be reduced” names a direction.

TRAP **Define the difference BEFORE choosing the sign: here d =** $x_1 - x_2$ (before minus after), so “times reduced” → $H_a : \mu_d > 0$. Keep the definition and the sign consistent.

Question 7

1 mark

Choosing the test · typing while walking

THE QUESTION

A team of psychologists studied the effect of multitasking on the completion of cognitive tasks. A group of 40 women participated in the study. Each woman owned a smartphone equipped with the same type of keyboard. The women typed a text passage on the phone twice, one time while sitting in a quiet room (a single task) and the other time while walking (a multitask). The order of the single task and the multitask was randomly determined for each woman. The psychologists recorded the time it took each woman to type the text for both tasks. If the conditions of inference are met, which of the following tests is most appropriate to analyze the data?

(A) A two-sample t-test for a difference between means

(B) A matched-pairs t-test for a mean difference

(C) A one-sample z-test for a proportion

(D) A two-sample z-test for a difference between proportions

Question 7 — solution

1 mark

Choosing the test · typing while walking

CONDENSED SOLUTION

- Every woman types the passage TWICE — sitting and walking — so the two time measurements come from the same person.
- Same unit measured twice → matched pairs, and the recorded data are times (means, not proportions).
- Answer: (B) — a matched-pairs t-test for a mean difference.
- The random ORDER is a design detail, not a clue for a two-sample test.

TRAP **“Twice”, “before and after”, “each person does both” — each phrase means paired. The coin flip protects against order effects, it does not create two groups.**

Question 9

1 mark

Choosing the test · sitting vs standing pulse

THE QUESTION

From a random sample of 50 people, sitting pulse rates and standing pulse rates were measured for each person. A coin was flipped to determine whether the sitting or the standing pulse rate would be measured first. Let μ_s represent the mean sitting pulse rate in the population, μ_t the mean standing pulse rate, and μ_d the mean of the differences between the sitting and standing (sitting − standing) pulse rates in the population. Which of the following represents an appropriate test and hypotheses to determine if there is a difference in mean pulse rates between sitting and standing in the population?

(A) A two-sample t-test with $H_0 : \mu_s = \mu_t$ and $H_a : \mu_s \neq \mu_t$

(B) A two-sample t-test with $H_0 : \mu_s = \mu_t$ and $H_a : \mu_s < \mu_t$

(C) A two-sample z-test with $H_0 : \mu_s = \mu_t$ and $H_a : \mu_s \neq \mu_t$

(D) A matched-pairs t-test with $H_0 : \mu_d = 0$ and $H_a : \mu_d \neq 0$

Question 9 — solution

1 mark

Choosing the test · sitting vs standing pulse

CONDENSED SOLUTION

- The SAME 50 people give both measurements — sitting and standing — so the data are paired.
- With defined differences (sitting − standing), the hypotheses are about μd: H0 : μd = 0.
- “Is there a difference” gives no direction → two-sided: Ha : μd ≠ 0.
- Answer: (D) — a matched-pairs t-test with H0 : μd = 0 and Ha : μd ≠ 0.

TRAP **Same people, two measurements, randomized order → matched pairs every time. Two-sample options (A), (B), (C) ignore that the samples are the same people.**

03

Conducting and concluding

Question 8

- What belongs in a correctly conducted test
- Reject only when $p < \alpha$

Key points · conducting the test

The p-value is compared with α — in that order

1	**Check conditions**	A correctly conducted test assumes the population is approximately normal (or n is large) and the data have no outliers.
2	**Degrees of freedom**	For a one-sample t-test with n = 10, df = n − 1 = 9 — using 9 is part of doing it right.
3	**The decision rule**	$p < \alpha \rightarrow$ reject H_0. $p > \alpha \rightarrow$ FAIL to reject H_0. Rejecting when $p > 0.05$ is simply wrong.
4	**Never “prove”**	Failing to reject does not prove H_0 — it only means the data are not convincing against it.

Question 8

1 mark

Conducting the test · the birthday lunch

THE QUESTION

On their birthdays, employees at a large company are permitted to take a 60-minute lunch break instead of the usual 30 minutes. Data were obtained from 10 randomly selected company employees on the amount of time that each actually took for lunch on his or her birthday. The company wishes to investigate whether these data provide convincing evidence that the mean time is greater than 60 minutes. Of the following, which information would NOT be expected to be a part of the process of correctly conducting a hypothesis test to investigate the question, at the 0.05 level of significance?

(A) Being willing to assume that the distribution of actual birthday lunch times for all employees at the company is approximately normal

(B) Knowing that there are no outliers in the data as indicated by the normal probability plot and boxplot

(C) Using 9 for the number of degrees of freedom

(D) Given that the p-value is greater than 0.05, rejecting the null hypothesis and concluding that the mean time was not greater than 60 minutes

Question 8 — solution

1 mark

Conducting the test · the birthday lunch

CONDENSED SOLUTION

- (A) and (B) are condition checks — normality and no outliers — part of conducting the test correctly.
- (C) is correct practice: n = 10 gives df = 9 for the one-sample t-test.
- (D) reverses the decision rule: p > 0.05 means FAIL to reject H0, never reject.
- Answer: (D) — rejecting when the p-value exceeds α is NOT part of a correct test.

TRAP **The decision rule is absolute: reject only when p < α. An answer that rejects with p > α — or that "proves"** H_0 with a large p — is automatically wrong.

04

The bank FRQ

Questions 10 to 13 · drive-up versus walk-in customers

- Variables, procedure, calculation, conclusion
- A complete two-sample t-test at $\alpha = 0.01$

Key points · the bank FRQ

Four parts, one complete inference

1	**Part A · variables**	Name each variable and classify it: quantitative or categorical. Visits per year = quantitative; customer group = categorical.
2	**Part B · procedure**	Two independent random samples of different customers → two-sample t-test for a difference between two means.
3	**Part C · calculation**	State the hypotheses, compute $t = (\bar{x}_1 - \bar{x}_2)/SE$, find df and the p-value — show every step.
4	**Part D · conclusion**	Compare p with α, decide, then conclude IN CONTEXT with the parameter, the groups and the units.

Scenario · Drive-up versus walk-in customers

read first

The data used by Questions 10 to 13

THE STORY

A bank in a large city gives customers two options for completing transactions: drive up to a teller window or walk into the bank to speak with a teller (teller: bank employee). Out of the bank's total number of customers, about 3,000 customers typically use the drive-up option and 1,000 customers typically use the walk-in option.

A manager wanted to investigate whether there is a difference in the average number of bank visits per year between customers who typically use the drive-up option and those who typically use the walk-in option.

A random sample of 50 customers from each group was selected, and the number of times each customer visited the bank during the past year was recorded. The summary statistics are in the table on the right. Complete the appropriate inference procedure at the 0.01 significance level.

SUMMARY STATISTICS

Group	n	Mean	Std. dev.
Drive Up	50	48	7
Walk In	50	54	10

n = 50 in each group; the samples are independent (different customers). The significance level is $\alpha = 0.01$, and the question asks about a DIFFERENCE, so the test is two-sided.

Question 10 A

1 mark

Bank FRQ · Part A · variables

THE QUESTION

Respond to parts A, B, C, and D.
Part A
Identify and classify the variable(s) in the study.

Question 10 A — solution

1 mark

Bank FRQ · Part A · variables

CONDENSED SOLUTION

- Variable 1: the number of bank visits per year — QUANTITATIVE, because it is a numerical count and an average of it is meaningful.
- Variable 2: the customer group (drive-up or walk-in) — CATEGORICAL, because it places each customer into one of two groups.
- One quantitative response variable and one categorical explanatory variable — that structure is what makes this a comparison of two means.
- Both the name and the classification are needed for the mark.

TRAP Every variable gets a classification, not just the first one you spot. “Number of visits” is quantitative even though it is a count — averages of counts are meaningful.

Question 13 B

1 mark

Bank FRQ · Part B · procedure

THE QUESTION

Respond to parts A, B, C, and D.
Part B
Identify the appropriate inference procedure.

Question 13 B — solution

1 mark

Bank FRQ · Part B · procedure

CONDENSED SOLUTION

- Two random samples of 50 customers each, from DIFFERENT customers — the samples are independent.
- The parameter of interest is the difference between two population means: $\mu_1 - \mu_2$.
- Appropriate procedure: a TWO-SAMPLE t-TEST for a difference between two means, two-sided, at $\alpha = 0.01$.
- Name the procedure and say why: independent samples, means of a quantitative variable.

TRAP **"Two-sample t-test" alone is half the answer — tie it to WHY: independent random samples and a quantitative response measured in each group.**

Question 11 C

2 marks

Bank FRQ · Part C · the procedure

THE QUESTION

Respond to parts A, B, C, and D.

Part C

Complete the inference procedure, including calculating the appropriate statistics.

Question 11 C — solution

2 marks

Bank FRQ · Part C · the procedure

CONDENSED SOLUTION

- Hypotheses: H0 : $\mu 1 - \mu 2 = 0$ (drive-up minus walk-in) versus Ha : $\mu 1 - \mu 2 \neq 0$, at $\alpha = 0.01$.
- Difference in means: $\bar{x}1 - \bar{x}2 = 48 - 54 = -6$ visits.
- Standard error: $SE = \sqrt{(7^2/50 + 10^2/50)} = \sqrt{(0.98 + 2.00)} = \sqrt{2.98} \approx 1.726$.
- Test statistic: $t = -6 / 1.726 \approx -3.476$.
- Welch degrees of freedom: df ≈ 87.7. Two-sided p-value: p ≈ 0.0008.

TRAP **Show the SE before t. The most common slips: dividing by pooled SD out of habit, or forgetting that "difference" makes the test two-sided.**

Question 12 D

1 mark

Bank FRQ · Part D · conclusion

THE QUESTION

Respond to parts A, B, C, and D.
Part D
Justify a conclusion, in context.

Question 12 D — solution

1 mark

Bank FRQ · Part D · conclusion

CONDENSED SOLUTION

- Compare: $p \approx 0.0008 < \alpha = 0.01 \rightarrow$ REJECT H0.
- Conclusion in context: the data give convincing evidence of a DIFFERENCE in the mean number of visits per year for all drive-up and all walk-in customers.
- Direction: walk-in customers averaged about 6 more visits per year (54 − 48) in the samples.
- No causation claimed and no “prove” language — evidence for a difference, described in the study's units.

TRAP **Reject** H_0 because $p < \alpha$, then finish in context: the parameter (mean visits per year), the two groups, and what the data suggest. “The test is significant” alone earns nothing.

05

The car auction FRQ

Questions 14 and 15 · 2024 versus 2026 selling prices

- Variables and study type
- Then the 98% confidence interval, worked in full

Key points · the car auction FRQ

From study design to a two-mean interval

1	**Part A · variables**	Selling price = quantitative; auction year (2024 or 2026) = categorical. Name AND classify each.
2	**Part B · study type**	Samples of vehicles ALREADY sold, no treatment assigned → an observational study.
3	**The interval**	$\bar{x}_1 - \bar{x}_2 \pm t^* \times$ SE with Welch df — estimate the difference, do not test it.
4	**Reading the interval**	0 inside the interval → the data are consistent with no difference in the population means.

Scenario · Vintage car auction prices

read first

The data used by Questions 14, 15 and the confidence interval

THE STORY

A city's annual car auction sells vintage vehicles (vintage: older). A car collector was curious to know whether the mean selling price of the vehicles was different between the auction years 2024 and 2026.

Approximately 2,000 vehicles were sold at each auction, and none of the vehicles were sold at both auctions.

The collector took a random sample of 36 vehicles sold at the 2024 auction and a random sample of 42 vehicles sold at the 2026 auction, and calculated the mean and standard deviation of the selling prices for each sample (table on the right).

Complete the appropriate inference procedure at the 98% confidence level to estimate the difference between the population means of the vehicle selling prices for 2024 and 2026.

SUMMARY STATISTICS

Year	n	Mean price	Std. dev.
2024	36	$106,400	$32,000
2026	42	$96,500	$22,000

Confidence level: 98%. The samples are independent (no vehicle was sold at both auctions). This procedure ESTIMATES the difference — a confidence interval, not a hypothesis test.

Question 14 A

1 mark

Cars FRQ · Part A · variables

THE QUESTION

Respond to parts A, B, C, D, and E.
Part A
Identify and classify the variable(s) in the study.

Question 14 A — solution

1 mark

Cars FRQ · Part A · variables

CONDENSED SOLUTION

- Variable 1: the selling price of the vehicle — QUANTITATIVE, a dollar amount with meaningful averages.
- Variable 2: the auction year (2024 or 2026) — CATEGORICAL, two groups.
- Same structure as the bank FRQ: quantitative response, categorical explanatory → comparing two means.
- Name each variable first, then classify it — the classification is the graded word.

TRAP **"Price" alone does not earn the mark — SELLING PRICE, in dollars, classified as quantitative. Year is categorical even though it is written with numbers.**

Question 15 B

1 mark

Cars FRQ · Part B · study type

THE QUESTION

Respond to parts A, B, C, D, and E.
Part B
Identify the type of study being conducted.

Question 15 B — solution

1 mark

Cars FRQ · Part B · study type

CONDENSED SOLUTION

- The collector took samples of vehicles that were ALREADY sold and recorded their prices.
- No treatment was assigned and nothing was manipulated — the year a vehicle sold at was not chosen by the collector.
- Type of study: an OBSERVATIONAL STUDY.
- Comparisons from observational studies can describe differences but cannot establish cause and effect.

TRAP **No assignment of treatments = observational. An experiment would need the collector to decide which auction each vehicle sold at, which is impossible here.**

Question Ext

2 marks

Cars FRQ · the 98% confidence interval

THE QUESTION

The booklet's remaining parts ask you to complete the inference procedure. Using the summary statistics for the two auction years, construct and interpret a 98% confidence interval for the difference between the population mean selling prices (μ_{2024} − μ_{2026}). Show the work that leads to your answers.

Question Ext — solution

2 marks

Cars FRQ · the 98% confidence interval

CONDENSED SOLUTION

- Point estimate: $\bar{x}_1 - \bar{x}_2 = 106{,}400 - 96{,}500 = \$9{,}900$.
- SE = $\sqrt{(32{,}000^2/36 + 22{,}000^2/42)} = \sqrt{(28{,}444{,}444 + 11{,}523{,}810)} \approx 6{,}322$.
- Welch df ≈ 60.6 → $t^* = 2.390$ for a 98% level.
- Interval: $9{,}900 \pm 2.390 \times 6{,}322 = 9{,}900 \pm 15{,}110 \approx (-\$5{,}210, \$25{,}010)$.
- Interpretation: we are 98% confident that the difference in mean selling prices ($\mu_{2024} - \mu_{2026}$) is between −$5,210 and $25,010. Because 0 is inside the interval, the data give no convincing evidence of a difference in mean selling prices between the two auction years.

TRAP 0 inside the interval means “no convincing difference”, not “the means are equal”. And df comes from the Welch formula — never add the sample sizes.

Mark scheme

Marks by question, with what earns each mark

S1	**1**	38 = parameter (μ), 41 = statistic ($\bar{x}$); hypotheses are written in parameters
1	**1**	$H_0 : \mu = 38$, $H_a : \mu > 38$ → (D)
2	**1**	$H_a : \mu > 24.5$ — the safe threshold, not the sample mean → (D)
3	**1**	$H_0 : \mu_D = 0$ and $H_a : \mu_D > 0$ → (B)
4	**1**	Independent samples, one-sided → one-sided two-sample t-test → (B)
5	**1**	Two independent groups of means → two-sample t-test → (D)
6	**1**	Same swimmers twice, times reduced → paired t-test, $H_a : \mu_d > 0$ → (C)
7	**1**	Each woman types twice → matched-pairs t-test → (B)
9	**1**	Same 50 people, difference defined → matched-pairs t-test, two-sided → (D)
8	**1**	Rejecting H_0 when $p > 0.05$ is NOT part of a correct test → (D)
Bank A	**1**	Visits per year = quantitative; customer group = categorical
Bank B	**1**	Two-sample t-test for a difference between two means, two-sided
Bank C	**2**	1 for hypotheses and $t = -6/1.726 \approx -3.476$ · 1 for $df \approx 87.7$ and $p \approx 0.0008$
Bank D	**1**	$p \approx 0.0008 < 0.01$ → reject; convincing evidence of a difference, in context
Cars A	**1**	Selling price = quantitative; auction year = categorical
Cars B	**1**	Observational study — samples of vehicles already sold
Cars CI	**2**	1 for $9{,}900 \pm 2.390 \times 6{,}322$ · 1 for (−$5,210, $25,010) with 0 inside → no convincing difference

MCQ total: 9 marks (plus the 1-mark starter). Free-response total: 10 marks — bank 5, cars 5.

Answer key

One line per question

ANSWERS

- S1 38 = parameter (μ) · 41 = statistic ($\bar{x}$) · hypotheses in μ
- 1 (D) — H0 : μ = 38, Ha : μ > 38
- 2 (D) — Ha : μ > 24.5, the safe threshold
- 3 (B) — H0 : μD = 0, Ha : μD > 0
- 4 (B) — one-sided, two-sample t-test
- 5 (D) — two-sample t-test for means
- 6 (C) — paired t-test, Ha : μd > 0
- 7 (B) — matched-pairs t-test

FREE RESPONSE

- Bank A visits per year = quantitative · group = categorical
- Bank B two-sample t-test for a difference of means
- Bank C t = −3.476 · df ≈ 87.7 · p ≈ 0.0008
- Bank D p ≈ 0.0008 < 0.01 → reject — a real difference; walk-in ≈ 6 more visits/year
- Cars A selling price = quantitative · auction year = categorical
- Cars B observational study
- Cars CI 9,900 ± 2.390 × 6,322 ≈ (−$5,210, $25,010) · 0 inside → no convincing difference

True or false · 1

T / F

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

Hypotheses are written in statistics such as the sample mean $\bar{x}$.

True or false · 1 — answer

FALSE

FALSE

Hypotheses are always written in PARAMETERS. Questions 1 and 2 of this booklet hinge on it: $H_0 : \mu = 38$ and $H_a : \mu > 24.5$ — the threshold, never the sample mean 25.3.

True or false · 2

T / F

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

A one-sided alternative hypothesis is used when the question asks whether the mean is greater (or less) than a value.

True or false · 2 — answer

TRUE

TRUE

"Greater than 38 minutes" → $H_a: \mu > 38$ (Question 1); "times will be reduced" → one-sided as well (Question 6). "Is there a difference" is the cue for two-sided.

True or false · 3

T / F

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

A matched-pairs design means the same unit is measured twice, or the units come in natural pairs.

True or false · 3 — answer

TRUE

TRUE

Questions 6, 7 and 9: each swimmer races before and after, each woman types twice, the same 50 people give both pulse rates. One sample of differences.

True or false · 4

T / F

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

Two samples of different, unrelated individuals call for a two-sample t-test.

True or false · 4 — answer

TRUE

TRUE

Questions 4 and 5: nine batteries of each brand, two separate groups of children. Independent samples → two-sample t for $\mu_1 - \mu_2$.

True or false · 5

T / F

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

When the p-value is greater than α, reject H_0 and accept the alternative hypothesis.

True or false · 5 — answer

FALSE

FALSE

$p > \alpha \rightarrow$ FAIL to reject H_0. Question 8's answer (D) is wrong for exactly this reason: rejecting with $p > 0.05$ reverses the decision rule.

True or false · 6

T / F

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

When the difference is defined as d = before − after, the paired test is about μ_d, the mean of the differences.

True or false · 6 — answer

TRUE

TRUE

Question 6: times reduced means before > after, so $H_a: \mu_d > 0$. Define the difference first, then set the sign of the alternative from it.

True or false · 7

T / F

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

Randomly deciding which measurement comes first protects the study against order effects.

True or false · 7 — answer

TRUE

TRUE

Question 9's coin flip (and Question 7's randomized order) stop systematic order effects — such as practice or fatigue — from distorting the paired differences.

True or false · 8

T / F

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

A large p-value proves that the null hypothesis is true.

True or false · 8 — answer

FALSE

FALSE

A large p-value means only FAIL TO REJECT: the data are not convincing against H_0. Absence of evidence is not proof — the bank FRQ's conclusion works the other way, $p < \alpha$.

FIVE THINGS TO DO IN THE EXAM

Do these and hypotheses will not cost you marks

- Write hypotheses in parameters: H0 : μ = value — and let the question's words fix the direction of Ha.
- Same unit measured twice → matched pairs with μd; two independent groups → two-sample t.
- Watch for the safe threshold: the tested value comes from the question's standard, not from the sample mean.
- Conclude only by comparing p with α: p < α → reject; p > α → fail to reject, never “accept” or “prove”.
- Finish every conclusion in context — the parameter, the two groups and the units.

Built from the Unit 5 Revisions b past-paper booklet (second variant).