AP STATISTICS · UNIT 5 REVISION

Unit 5 Revisions b

Scatterplots, correlation and regression — from association to least-squares lines

10
multiple choice questions

3
free-response questions

69
slides in total

Built from the Unit 5 Revisions b past-paper booklet

How this lesson runs

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

Starter	One quick recall question on describing a scatterplot — 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

Reading plots	Describe every scatterplot with direction (positive or negative), form (linear or nonlinear) and strength.
Correlation	r measures the strength and direction of LINEAR association only, from −1 to +1 — never summarize a curved pattern with it.
Regression	The least-squares line $\hat{y} = a + bx$ predicts with the right units and the word PREDICTED; residuals check whether it fits.
Judging models	A residual plot with no pattern and a high r^2 support a linear model — the full justification the FRQ wants.

Vocabulary

Key terms with Chinese translations

scatterplot	**散点图**	displays two quantitative variables as points
association	**关联**	direction, form and strength of a relationship
positive association	**正相关**	as one variable increases, the other tends to increase
nonlinear pattern	**非线性模式**	a curved relationship — r cannot describe it
correlation r	**相关系数 r**	strength and direction of LINEAR association
coefficient of determination	**判定系数 r^2**	the proportion of variation in y explained by x
residual	**残差**	observed y minus predicted ŷ
residual plot	**残差图**	residuals vs x; no pattern supports a linear model
least-squares regression line	**最小二乘回归线**	ŷ = a + bx, the line minimizing squared residuals
extrapolation	**外推**	predicting outside the data range — not justified

S

Starter · Describing a scatterplot

What you must remember before the questions

- One quick recall question
- Direction, form and strength — the three-part description

Key points · describing scatterplots

Direction, form, strength — in that order

1. **Direction** — Positive: both variables rise together. Negative: one rises as the other falls. None: a shapeless cloud.
2. **Form** — Linear: points hug a straight line. Nonlinear: rising then levelling off, a curve, or a ”U“ shape.
3. **Strength** — How tightly the points cluster around the pattern — strong, moderate or weak.
4. **First move** — Every plot question starts the same way: describe it with direction, form and strength before touching the options.

Question S1

1 mark

Starter · Describing scatterplots

THE QUESTION

A scatterplot shows the mass, in grams, of chameleon eggs against their age, in days. The points rise steeply for the first 40 days, then the rise gradually levels off between about 40 and 190 days.
Describe the relationship between the mass and the age of the chameleon eggs in terms of direction, form and strength.

Question S1 — solution

1 mark

Starter · Describing scatterplots

CONDENSED SOLUTION

- Direction: POSITIVE — as age increases, mass increases.
- Form: NONLINEAR — the mass rises steeply at first, then the pattern bends and levels off.
- Strength: strong early on, but the flattening tail is the key feature to name.
- This exact plot is Question 3 — the same three-part description earns the marks there.

TRAP Rising does not automatically mean linear. “Levels off” is your cue for a nonlinear form — the single most-tested description in this unit.

01

Scatterplots and association

Questions 3, 4, 5 and 6

- Direction, form and strength from real plots
- Does the plot support the claim being made?

Key points · scatterplots and association

Describe the plot, then decide what it supports

1	**The description**	Direction (positive/negative/none), form (linear/nonlinear), strength. Lead with direction, then form.
2	**Curves are nonlinear**	Rising-then-levelling, ”U“ and dome shapes are nonlinear even when the trend is obvious.
3	**No trend = no association**	A horizontal cloud of points shows little to no association — knowing x then tells you nothing about y.
4	**Supporting a belief**	A belief is supported only when the association points the claimed way — a positive association supports “higher with higher”.

Question 3

1 mark

Scatterplots and association

THE QUESTION

A scatterplot shows the mass, in grams, of chameleon eggs against their age, in days. The points rise steeply from about 1 gram at 0 days to roughly 1.9 grams by 40 days, then continue rising more and more gently, levelling off around 2.0 to 2.3 grams by 190 days.

Which of the following best describes the relationship between the mass and the age of the chameleon eggs?
(A) The association is positive and linear.
(B) The association is positive and nonlinear.
(C) The association is negative and linear.
(D) The association is negative and nonlinear.

Question 3 — solution

1 mark

Scatterplots and association

CONDENSED SOLUTION

- Direction: mass increases as age increases → POSITIVE.
- Form: steep rise, then the rise bends and levels off → NONLINEAR.
- Answer: (B) — the association is positive and nonlinear.
- Both words are needed: the positive trend alone does not make it linear.

TRAP The levelling tail is the giveaway. Students who only glance at “it goes up” choose (A) — name the form before the options.

Question 4

1 mark

Scatterplots and association · no association

THE QUESTION

Amal, a nutritionist, selected a random sample of 25 cold cereals and created a scatterplot of sugar content, in grams, against calorie content, in kilocalories. In the plot, the points form an approximately horizontal band — all cereals show roughly 350 to 450 kilocalories, whether they have 1 gram or 48 grams of sugar.

Based on the scatterplot, is there evidence of an association between sugar and calories in cold cereals such that the amount of sugar appears to be useful in predicting calorie content?

(A) No, because the dots show little to no trend and form an approximately horizontal pattern, which shows there is little to no association.

(B) No, because the sample size was less than 30, so the sampling distributions for each variable may not be normally distributed.

(C) Yes, because the dots form a relatively straight horizontal line, which shows a linear association.

(D) Yes, because sugar has a lot of calories, so more sugar must mean more calories.

Question 4 — solution

1 mark

Scatterplots and association · no association

CONDENSED SOLUTION

- The band is horizontal: calories sit near 350–450 no matter what the sugar content is.
- A horizontal pattern = little to no trend = little to no association.
- Without an association, sugar is NOT useful for predicting calories.
- Answer: (A). (D) is background knowledge, not data — the plot must decide.

TRAP A flat horizontal cloud is the picture of “no association”. And a real AP reader wants the DATA to decide — never nutrition common sense.

Question 5

1 mark

Scatterplots and association · supporting a belief

THE QUESTION

Courtney, an economist, suspects that regions whose residents spend more annually on housing and utilities also spend more on gasoline and other energy goods. She selected a random sample of 25 regions and plotted annual spending on gasoline and energy goods (dollars) against annual spending on housing and utilities (dollars). The plot shows a positive association: regions near 10,000–12,000 dollars of housing spending show roughly 1,000–2,000 dollars of energy spending, rising together.

Which of the following statements best explains whether the scatterplot supports Courtney's belief?
(A) The scatterplot supports Courtney's belief, because it depicts a positive association between the two spendings.
(B) The scatterplot supports Courtney's belief, because it depicts a negative association.
(C) The scatterplot does not support Courtney's belief, because it depicts a positive association.
(D) The scatterplot does not support Courtney's belief, because it depicts a negative association.

Question 5 — solution

1 mark

Scatterplots and association · supporting a belief

CONDENSED SOLUTION

- Courtney's belief: higher housing spending goes with higher energy spending — that is a POSITIVE claim.
- The plot shows a positive association: both spendings rise together.
- A positive association points exactly the way the belief claims → the belief IS supported.
- Answer: (A). The direction of the plot must match the direction of the claim.

TRAP Match the direction word to the claim. “Higher with higher” is positive — (C) and (D) twist that, and a negative plot could never support this belief.

Question 6

1 mark

Scatterplots and association · linear pattern

THE QUESTION

Jacob, an economist, selected a random sample of 30 regions and recorded each region's gross domestic product (GDP) and the total personal income of its residents, both in millions of Country A's currency. The scatterplot shows the points climbing in a fairly straight pattern from about 50,000/50,000 up to roughly 1,100,000/880,000, with a clear noticeable slope.

Based on the scatterplot, does there appear to be an association between GDP and personal income for the regions in Country A?

(A) No, because the points form a fairly straight linear pattern rather than being widely scattered.

(B) No, because the points trend upward at an angle rather than forming a vertical or horizontal line.

(C) Yes, because the sample size is at least 30, so the conditions for the central limit theorem have been met.

(D) Yes, because the points form a fairly straight linear pattern with a noticeable slope.

Question 6 — solution

1 mark

Scatterplots and association · linear pattern

CONDENSED SOLUTION

- The points climb together in a fairly straight line with a noticeable slope.
- That is a textbook positive linear association → yes, there is an association.
- Answer: (D) — the reason names the PATTERN, which is what the question asks about.
- (C) confuses association with the CLT: sample size has nothing to do with whether two variables move together.

TRAP Association is about the PATTERN of points, not the sample size. The CLT is an inference idea — it never decides whether a scatterplot shows association.

02

Residual plots and model appropriateness

Questions 7, 8 and 12

- What a residual plot must look like
- When can a linear model be used?

Key points · residuals and model appropriateness

The residual plot is the judge of a linear model

1	**Residual**	residual = observed − predicted. A point above the line was under-predicted; below, over-predicted.
2	**The good plot**	Random scatter around 0 with no pattern → a linear model is appropriate.
3	**The bad plot**	A curve, a ”U“, a fan or a trend in the residuals → the linear model is missing something.
4	**r^2 as a check**	A high r^2 supports the fit, but it is the PLOT pattern that rules out curvature — quote both.

Question 7

1 mark

Residual plots · model appropriateness

THE QUESTION

A residual plot from a regression analysis using x as the independent variable is shown. The residuals scatter randomly above and below the 0 line for all x from 0 to 8, ranging between about −0.8 and +0.5, with no curve, trend or fan shape anywhere in the plot.

Based on the residual plot, is it reasonable to conclude that a linear regression model is appropriate?
(A) Yes, because the plot shows no apparent pattern.
(B) Yes, because the points in the plot display less variation overall as x increases.
(C) No, because the plot shows no apparent pattern.
(D) No, because the points in the plot display more variation overall as x increases.

Question 7 — solution

1 mark

Residual plots · model appropriateness

CONDENSED SOLUTION

- What a residual plot should look like: random scatter around the 0 line.
- This plot shows exactly that — no curve, no trend, no fan.
- No apparent pattern → a linear model IS appropriate.
- Answer: (A). “No pattern” is the evidence FOR the model, never against it.

TRAP “No apparent pattern” is good news for a linear model. Reading it as a flaw and choosing (C) reverses the whole rule.

Question 8

1 mark

Correlation · what r can and cannot say

THE QUESTION

Two scatterplots are shown. Plot A: the points form a ”U“ shape — they start near y = 3, dip to about y = 1.5 around x = 2, then climb back to about y = 3.5 by x = 4. Plot B: the points fall in an almost perfectly straight line from (0.5, 4.5) down to (4.8, 0.5).

Based on the scatterplots, which statement CANNOT be true?
(A) Plot A indicates a nonlinear relationship between the variables.
(B) Plot B indicates a linear relationship between the variables.
(C) The correlation coefficient for the data in plot B is close to 1.
(D) The correlation coefficient for the data in plot A is close to 0.

Question 8 — solution

1 mark

Correlation · what r can and cannot say

CONDENSED SOLUTION

- Plot A: a ”U“ → nonlinear, so (A) is true, and r near 0 for a symmetric ”U“ is plausible → (D) is true.
- Plot B: a straight, downward line → (B) is true — it is a LINEAR relationship.
- But the line falls, so r for plot B is close to −1, never close to +1.
- Answer: (C) — r close to 1 for plot B cannot be true: r = 1 demands an UPWARD straight line.

TRAP r = 1 is an upward line only. A perfect DOWNWARD line has r = −1. Check the sign of the slope before believing any claim about r.

Question 12

2 marks

Model appropriateness · the full justification

THE QUESTION

The scatterplot of annual income against amount spent on vacations for 8 families shows the points rising in a fairly straight pattern from about (27,400, 1,700) up to (98,100, 5,100), with no curve or bending anywhere. The regression analysis gives $r^2 = 0.8498$.

Determine whether a linear model is appropriate for describing the relationship between income and vacation costs. Justify your answer.

Question 12 — solution

2 marks

Model appropriateness · the full justification

CONDENSED SOLUTION

- Evidence 1 — the scatterplot: the points rise in a fairly straight pattern with no curvature → the form is linear.
- Evidence 2 — the strength measure: $r^2 = 0.8498$, so about 85% of the variation in vacation spending is explained by income.
- A linear model IS appropriate, justified by BOTH the plot's linear pattern and the high r^2.
- One piece of evidence alone is half an answer — the rubric pays for the pair.

TRAP "Yes" with only r^2 quoted ignores curvature; "yes" with only the plot ignores strength. The complete justification always cites the pattern AND the measure.

03

Correlation and slope interpretation

Questions 9, 2 and 1

- Reading r from a plot, reading slope from an equation
- And picking the outlier on a scatterplot

Key points · correlation and slope

Two readings, two different objects

1	**r = ±1**	r = ± means every point lies exactly on a straight line: upward for +1, downward for −1.
2	**Slope units**	In $\hat{y} = a + bx$, b carries the units of y per unit of x — say them both when you interpret.
3	**The word PREDICTED**	A regression slope never causes anything: the sales price is PREDICTED to increase, not guaranteed to.
4	**Outliers on plots**	For “which point is farthest from the line”, compare each marked point's vertical gap to the line — the biggest gap wins.

Question 9

1 mark

Correlation · reading r = −1

THE QUESTION

Which of the following scatterplots could represent a data set with a correlation coefficient of r = −1?

(A) Points falling from (8, 4) down to (89, 1), lying very close to a straight downward line.

(B) Points dipping from about 3.2 down to 1.3 and then climbing back up to 3.6 — a ”U“ shape.

(C) Points climbing steadily from (9, 0.7) up to (91, 3.7) in a rising pattern.

(D) Points scattered in a round cloud between 20 and 80 on x and 1 and 5 on y, with no pattern.

Question 9 — solution

1 mark

Correlation · reading r = −1

CONDENSED SOLUTION

- r = −1 requires EVERY point to lie exactly on a straight downward line.
- (A) is the only option whose points lie on a straight downward line — exactly what r = −1 requires. (B) is a ”U“, (C) rises, (D) is a shapeless cloud.
- Only the option with points exactly on a downward straight line can give r = −1.
- Answer: (A). Flag: the four mini-plots could not be reproduced here — verify against the printed booklet that (A) is the one on the straight descending line.

TRAP Any bend or scatter, however slight, pulls r off ±1. Perfection is the requirement — inspect all four mini-plots for a flawless descending line.

Question 2

1 mark

Slope interpretation

THE QUESTION

A real-estate analyst fits the least-squares regression line $\hat{\hat{y}} = 0.5 + 1.1x$, where x = listing price, in hundreds of thousands of dollars, and $\hat{y}$ = sales price, in hundreds of thousands of dollars. The units matter: the slope 1.1 is in hundreds of thousands of dollars of sales price per hundred thousand dollars of listing price.

Which of the following best interprets the slope of the regression line?
(A) For each hundred-thousand-dollar increase in listing price, the sales price is predicted to increase by $1,100.
(B) For each dollar increase in listing price, the sales price is predicted to increase by $1.10.
(C) For each hundred-thousand-dollar increase in listing price, the sales price increases by exactly $110,000, with no variation.
(D) For each hundred-thousand-dollar increase in listing price, the sales price is predicted to increase by $110,000.

Question 2 — solution

1 mark

Slope interpretation

CONDENSED SOLUTION

- Slope 1.1 = hundreds of thousands of sales dollars per hundred thousand of listing price.
- 1.1 × $100,000 = $110,000 of predicted sales price per $100,000 of listing price.
- Regression PREDICTS: the word “predicted” must appear — it is not an exact rule.
- Answer: (D). (A) botches the units, (B) botches the scale, (C) drops “predicted”.

TRAP A slope answer needs all three: the right number, the right units, and “predicted”. Miss any one and the interpretation loses its mark.

Question 1

1 mark

Outliers on a scatterplot

THE QUESTION

A scatterplot shows the heights, in feet, of roller coasters against their top speeds. A regression line is drawn through the points. Four points are labelled with their heights: 60, 105, 150 and 200 feet sit close to the line, while one coaster at a height of 215 feet lies far below the line — it has a top speed much lower than the line predicts for a coaster that tall.

One of the points is an outlier because it is far from the regression line. Which point is it?
(A) The point at height 60 feet
(B) The point at height 105 feet
(C) The point at height 150 feet
(D) The point at height 215 feet

Question 1 — solution

1 mark

Outliers on a scatterplot · roller coasters

CONDENSED SOLUTION

- An outlier here means far from the REGRESSION line — not merely far from the other points.
- The labelled points at 60, 105, 150 and 200 feet sit close to the line.
- The 215-foot coaster's top speed is far below what the line predicts for its height → the biggest vertical gap.
- Answer: (D) — the point at height 215. Read the plot at full size to confirm which labelled value sits farthest off the line.

TRAP "Outlier" is always relative to the model named in the question. Compare gaps to the LINE, not distances to the crowd of points.

04

The height and foot length FRQ

Question 10 · Part A, parts i and ii

- Read the regression equation from computer output
- Substitute, compute, round — and show the work

Key points · the regression prediction FRQ

From output to prediction in three moves

1	**Read the output**	Constant = intercept a; the predictor row = slope b. Here $\hat{y} = 83.056 + 3.483x$.
2	**Substitute**	$\hat{y} = 83.056 + 3.483 \times 23 = 163.165 \approx 163.2$ cm. Show the substituted line before the answer.
3	**Second prediction**	$\hat{y} = 83.056 + 3.483 \times 27 = 177.097 \approx 177.1$ cm — same line, new x, same discipline.
4	**Stay in range**	Both 23 cm and 27 cm lie inside the data, so the predictions are justified — extrapolation would not be.

Scenario · Height and foot length

read first

The computer output used by both parts

THE STORY

A random sample of 65 high school seniors was selected from a certain high school. The scatterplot displays the relationship between height, in centimetres (cm), and foot length, in cm, for each student.

The least-squares regression line is drawn on the plot: the points climb steadily from about (22, 155) up to (31, 191), hugging the line closely — a strong positive linear pattern.

The corresponding computer output is shown on the right, and it also reports S = 6.548 and r^2 = 60.61%.

COMPUTER OUTPUT

Predictor	Coef	SE Coef
Constant	83.056	9.187
Foot length	3.483	0.354

Also from the output: S = 6.548 and r^2 = 60.61%.

So the regression line is:
$\hat{y} = 83.056 + 3.483x$,
with x = foot length in cm and $\hat{y}$ = height in cm.

Question 10 i

2 marks

Height FRQ · Part A (i)

THE QUESTION

Using the computer output, respond to the following. Show the work that leads to your answers.

i. Calculate the predicted height of a high school senior with a foot length of 23 cm.

Question 10 i — solution

2 marks

Height FRQ · Part A (i)

CONDENSED SOLUTION

- From the output: $\hat{y} = 83.056 + 3.483x$, where x is foot length in cm.
- Substitute x = 23: $\hat{y} = 83.056 + 3.483 \times 23$.
- Compute: 83.056 + 80.109 = 163.165.
- The predicted height is approximately 163.2 cm — with the substituted line shown, as the question demands.

TRAP Show the substitution $\hat{y} = 83.056 + 3.483(23)$ before the answer. A bare "163.2" with no work loses the mark in a "show the work" part.

Question 10 ii

2 marks

Height FRQ · Part A (ii)

THE QUESTION

Using the computer output, respond to the following. Show the work that leads to your answers.

ii. Calculate the predicted height of a high school senior with a foot length of 27 cm.

Question 10 ii — solution

2 marks

Height FRQ · Part A (ii)

CONDENSED SOLUTION

- Same line: $\hat{y} = 83.056 + 3.483x$.
- Substitute x = 27: $\hat{y} = 83.056 + 3.483 \times 27$.
- Compute: 83.056 + 94.041 = 177.097.
- The predicted height is approximately 177.1 cm. Both 23 cm and 27 cm lie within the observed foot lengths, so both predictions are justified.

TRAP Recompute cleanly for each x — carrying 163.2 forward and adjusting by hand is where arithmetic slips creep in. Substitution first, rounding last.

05

The income and vacation FRQ

Questions 11 and 12 · correlation, r^2 and appropriateness

- Compute r and r^2 from the 8-family table
- Then build the full appropriateness justification

Key points · correlation, r^2 and justification

Calculate, interpret, then justify the model

1	**The formula**	$r = \Sigma(x-\bar{x})(y-\bar{y}) / \sqrt{[\Sigma(x-\bar{x})^2 \cdot \Sigma(y-\bar{y})^2]}$ — technology does the arithmetic; you interpret it.
2	**The result**	$r = 0.9219$: strong, positive, linear association between income and vacation spending.
3	**Then r^2**	$r^2 = 0.8498$ — about 85% of the variation in vacation spending is explained by income.
4	**The justification**	Linear model appropriate: straight pattern, no curvature, and $r^2 = 0.8498$ is high. Quote both.

Scenario · Income and vacation cost

read first

The 8-family data used by Questions 11 and 12

THE STORY

A travel agency wants to market vacation packages to families based on their income. To inform its marketing strategy, the agency obtained data on 8 families, recording each family's income, in dollars, and the amount spent on vacations, in dollars, annually.

The scatterplot of the data shows the points rising steadily from about (27,400, 1,700) to (98,100, 5,100) in a fairly straight pattern, with no curve or bend.

The data are shown in the table on the right.

INCOME AND VACATION COST (DOLLARS)

Income	Vacation
41,100	2,700
53,000	2,400
27,400	1,700
34,400	2,500
65,800	2,800
98,100	5,100
72,000	4,200
56,700	3,200

8 families, n = 8. Each row is one family: income and annual vacation spending, both in dollars.

Question 11 i

2 marks

Income FRQ · Part A (i)

THE QUESTION

Using the table of income and vacation spending for the 8 families, respond to the following. Show the work that leads to your answers.

i. Calculate the correlation coefficient.

Question 11 i — solution

2 marks

Income FRQ · Part A (i)

CONDENSED SOLUTION

- $r = \Sigma(x-\bar{x})(y-\bar{y}) / \sqrt{[\Sigma(x-\bar{x})^2 \cdot \Sigma(y-\bar{y})^2]}$, computed from the 8 pairs (technology allowed).
- Substituting the 8 income and vacation values: $r = 0.9219$.
- Interpretation: a strong, positive, linear association — higher incomes go with higher vacation spending.
- The value is close to 1 and positive, which matches the straight upward pattern in the plot.

TRAP A bare 0.9219 shows the number but not the understanding — pair it with “strong, positive, linear association between income and vacation spending”.

Question 11 ii

2 marks

Income FRQ · Part A (ii)

THE QUESTION

Using the table of income and vacation spending for the 8 families, respond to the following. Show the work that leads to your answers.

ii. Calculate the coefficient of determination.

Question 11 ii — solution

2 marks

Income FRQ · Part A (ii)

CONDENSED SOLUTION

- $r^2 = (0.9219)^2 = 0.8498$.
- As a percentage: about 85% of the variation in vacation spending is explained by the linear relationship with income.
- The remaining 15% is due to other factors and natural variation.
- Interpretation in context earns the second mark — say what explains what.

TRAP r^2 must be interpreted, not just squared: “85% of the VARIATION in vacation cost is explained by income”. Swapping the variables loses the mark.

Question 12b

2 marks

Income FRQ · Part B · revisited

THE QUESTION

Question 12, met in Section 2, completes this FRQ. The scatterplot of income against vacation spending shows a fairly straight, rising pattern with no curvature, and the analysis gives $r^2 = 0.8498$.

Determine whether a linear model is appropriate for describing the relationship between income and vacation costs. Justify your answer.

Question 12b — solution

2 marks

Income FRQ · Part B · revisited

CONDENSED SOLUTION

- State the decision: a linear model IS appropriate.
- Evidence 1: the scatterplot shows a linear pattern — points rising steadily with no curvature.
- Evidence 2: $r^2 = 0.8498$ is high — about 85% of the variation in vacation spending is explained by income.
- Both the plot pattern and the strength measure, in context — that is the complete justification.

TRAP The exam rubric pays for BOTH pieces of evidence. “The r^2 is high” alone cannot rule out curvature — the plot is what shows the form.

Mark scheme

Marks by question, with what earns each mark

S1	**1**	Direction positive · form nonlinear · strength named
3	**1**	Positive and nonlinear → (B)
4	**1**	Horizontal pattern, little to no association → (A)
5	**1**	Positive association supports the belief → (A)
6	**1**	Fairly straight linear pattern, noticeable slope → (D)
7	**1**	No apparent pattern → linear model appropriate → (A)
8	**1**	Plot B is linear DOWNWARD: r near −1, not near 1 → (C) cannot be true
9	**1**	Points exactly on a downward straight line (flag: verify against the printed mini-plots)
2	**1**	Per $100,000 of listing price, predicted +$110,000 → (D)
1	**1**	Farthest from the regression line → 215 ft → (D)
10 i	**2**	1 for substituting 83.056 + 3.483(23) · 1 for 163.165 ≈ 163.2 cm
10 ii	**2**	1 for substituting 83.056 + 3.483(27) · 1 for 177.097 ≈ 177.1 cm
11 i	**2**	1 for r = 0.9219 · 1 for the strong positive linear association interpretation
11 ii	**2**	1 for $r^2 = 0.8498$ · 1 for interpreting ≈ 85% of variation explained
12	**2**	1 for the linear, no-curvature plot pattern · 1 for the high $r^2 = 0.8498$, both in context

MCQ total: 10 marks. Free-response total: 10 marks.

Answer key

One line per question

ANSWERS

- S1 positive · nonlinear · strength named
- 1 (D) — the 215 ft coaster, farthest from the line
- 2 (D) — predicted increase of $110,000 per $100,000
- 3 (B) — positive and nonlinear
- 4 (A) — horizontal pattern, no association
- 5 (A) — positive association supports the belief
- 6 (D) — fairly straight linear pattern, noticeable slope

FREE RESPONSE

- 7 (A) — no apparent pattern → linear model appropriate
- 8 (C) — plot B falls: r near −1, cannot be near +1
- 9 perfect downward line — verify against printed plots (A)
- 10 i–ii 163.165 ≈ 163.2 cm · 177.097 ≈ 177.1 cm
- 11 i–ii $r = 0.9219$ · $r^2 = 0.8498$ (≈ 85% explained)
- 12 yes — straight pattern with no curvature and $r^2 = 0.8498$ is high

True or false · 1

T / F

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

Correlation measures the strength of any relationship between two quantitative variables.

True or false · 1 — answer

FALSE

FALSE

r measures LINEAR association only. In Question 3's chameleon plot the association is strong but curved — a curved pattern gets a modest r even when the link is obvious.

True or false · 2

T / F

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

A residual equals the observed value minus the predicted value.

True or false · 2 — answer

TRUE

TRUE

residual = observed − predicted. Points above the line have positive residuals (under-predicted); below, negative (over-predicted).

True or false · 3

T / F

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

Predicting a value outside the range of the data is justified as long as the regression line is extended.

True or false · 3 — answer

FALSE

FALSE

That is extrapolation and it is never justified by the line. The model was fitted to the observed range only — beyond it, nothing guarantees the pattern continues.

True or false · 4

T / F

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

Interpreting a slope correctly requires the word “predicted” and the units of both variables.

True or false · 4 — answer

TRUE

TRUE

Question 2's answer: for each hundred-thousand-dollar increase in listing price, the sales price is PREDICTED to increase by $110,000. Number, units and “predicted” — all three.

True or false · 5

T / F

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

A correlation coefficient near ±1 proves the relationship between the variables is linear.

True or false · 5 — answer

FALSE

FALSE

r is computed as if the relationship were linear. In plot A of Question 8 the ”U“ pattern could still give a large |r| — always look at the scatterplot first.

True or false · 6

T / F

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

The least-squares regression line always passes through the point $(\bar{x}, \bar{y})$.

True or false · 6 — answer

TRUE

TRUE

The line of means: $\bar{y} = a + b \times \bar{x}$ is always true, which is why the line pivots through the centre of the cloud of points.

True or false · 7

T / F

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

An r^2 of 0.8498 means about 85% of the variation in vacation spending is explained by income.

True or false · 7 — answer

TRUE

TRUE

That is exactly Question 11(ii): $r^2 = 0.9219^2 \approx 0.8498$. The interpretation names y's variation explained by x, in context.

True or false · 8

T / F

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

A residual plot showing no apparent pattern supports the use of a linear model.

True or false · 8 — answer

TRUE

TRUE

Question 7's logic: random scatter around 0 is the evidence that a straight line fits. A curve or trend in the residuals would say the opposite.

FIVE THINGS TO DO IN THE EXAM

Do these and Unit 5 will not cost you marks

- Describe every scatterplot with direction, form and strength — “levels off” means nonlinear.
- Read the slope with its units and the word PREDICTED — and never drop either one.
- Judge a model from the residual plot: no pattern supports linear; a curve or fan does not.
- Quote BOTH pieces in an appropriateness justification: the plot pattern AND the high r^2.
- Interpret r^2 as variation in y explained by x, in context — and check predictions stay in range.

Built from the Unit 5 Revisions b past-paper booklet.