Past-paper walk-through

Residuals ▪ 2.7

eXercise

Questions in this set

- 2026 Q6
- 2023 Q5
- 2022 Q1
- 2017 Q1
- 2014 Q6

Reading the mark scheme

- **B1** a mark that stands on its own – the point is either made or not.
- **M1** a *method* mark: the right approach, even if the number is wrong.
- **A1** an *answer* mark, and it usually needs its M mark first.
- **C1** a working mark on the way to the answer.
- // separates answers the examiner accepts equally.
- **ecf** = error carried forward: a later part is marked on your own earlier answer.
- **owtte** = "or words to that effect" – the wording need not match.

Question 6

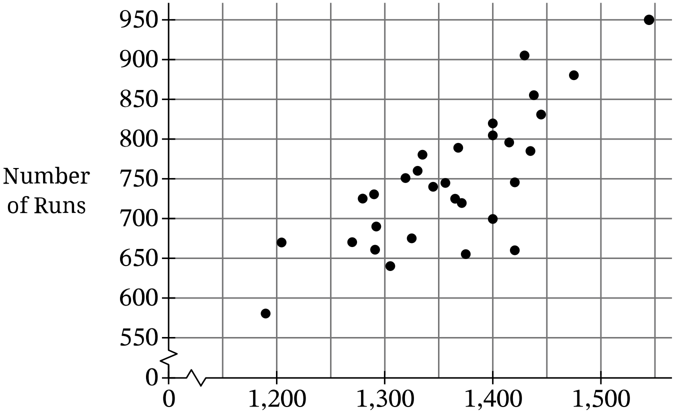
2026 ▪ Q6

The following information applies to parts A and B.

6. To win a game, a baseball team needs to score runs. To score runs, players need to get on base. The primary way players get on base is by hitting the ball. Figure 1 shows the relationship between the number of hits and the number of runs for 30 randomly selected professional baseball teams.

Figure 1. Scatterplot of Number of Runs by Number of Hits

Question 6


Number of Runs
950
900
850
800
750
700
650
600
550
0
0
1,200
1,300
1,400
1,500

Question 6

Number of Hits

Part A

Consider the scatterplot in Figure 1.

i. Describe the relationship between the number of hits and the number of runs for these baseball teams, in context.

ii. Based on the data from the sample of 30 teams, an equation of the least-squares regression line model for predicting the number of runs from the number of hits is as follows.

$$\text{Predicted number of runs} = -372.2 + 0.823\,(\text{number of hits})$$

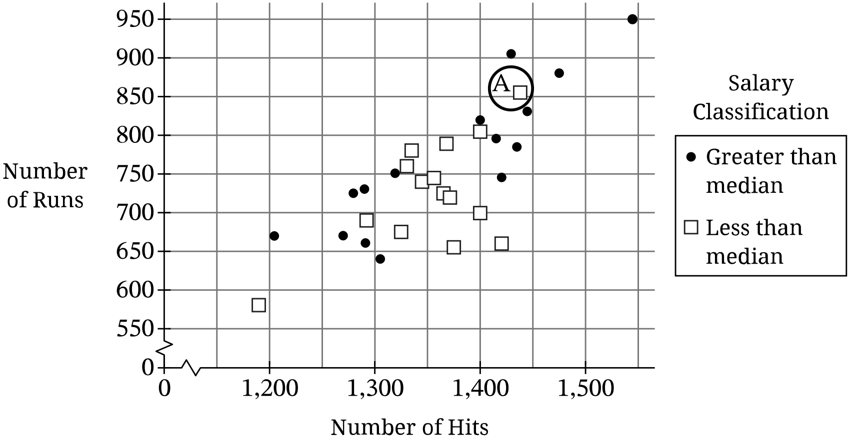
Using the given regression equation, calculate the point estimate for the predicted number of runs a team would score if they were to achieve their goal of 1,250 hits. Show your work.

Part B

Each team has a total salary equal to the sum of the salaries of all players on the team. The median total salary for the sample of 30 teams is $160 million. In Figure 2, points with dots represent teams with a total salary greater than the median, and points with squares represent teams with a total salary less than the median.

Figure 2. Scatterplot of Number of Runs by Number of Hits and Salary Classification

Question 6


Number of Runs
950
900
850
800
750
700
650
600
550
0
0
1,200
1,300
1,400
1,500
Number of Hits
A
Salary Classification
Greater than median
Less than median

Question 6

Consider the scatterplot in Figure 2.

i. Compare the team represented by the point that is circled and labeled A with the other teams that have the same total salary classification.

ii. For each salary classification, consider the linear relationship between the number of hits and the number of runs. For teams with a total salary greater than the median, is the strength of the linear relationship stronger than, weaker than, or similar to the strength of the linear relationship for teams with a total salary less than the median? Explain.

Part C

Different types of intervals can be used when working with linear regression models. One type of interval is a confidence interval for the slope of the least-squares regression line. Two other types of intervals in the context of this problem are as follows:

Question 6

- A confidence interval that estimates the mean number of runs for all teams with a specific number of hits.
- A prediction interval that predicts the number of runs for a single team with a specific number of hits.

All three intervals use the same basic formula:

$$\text{Point Estimate} \pm t^{*}(\text{standard error}),$$

where critical value t^{*} comes from a t-distribution that has $n-2$ degrees of freedom. The same critical value is used when finding each of these intervals with the same level of confidence.

Consider the point estimate for the predicted number of runs found in part A (ii) for the team whose goal is to achieve 1,250 hits next year. Recall that this value was calculated using the linear regression model based on the data from the sample of 30 teams.

i. What is the critical value for a 95% confidence level that would be used for the confidence interval for the mean number of runs as well as for the prediction interval for the number of runs? Indicate the answer to two decimal places.

ii. The standard error for the confidence interval for the mean number of runs is 17.48. Assuming the conditions for inference are met, calculate the 95% confidence interval for the mean number of runs for all teams with 1,250 hits. Show your work.

iii. The standard error for the prediction interval for the number of runs is 56.78. Assuming the conditions for inference are met, calculate the 95% prediction interval for the number of runs for a single team with 1,250 hits. Show your work.

Part D

i. Would a distribution of sample means be expected to have more variability or less variability than a distribution of individual observations? Explain.

Question 6

ii. The standard error used in the confidence interval that estimates the mean number of runs for all teams with x hits can be found using the following formula.

$$\sqrt{s^2\left(\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right)}$$

The standard error used in the prediction interval that predicts the number of runs for a single team with x hits can be found using the following formula.

$$\sqrt{s^2+s^2\left(\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right)}$$

In both standard error formulas, s is the standard deviation of the residuals, n is the sample size, and $\bar{x}$ is the sample mean number of hits. Based on the answer from part D (i) and the standard error formulas for the confidence interval and the prediction interval, explain why the prediction interval calculated in part C (iii) is wider than the confidence interval calculated in part C (ii).

STOP

END OF EXAM

Question 5

2023 ▪ Q5

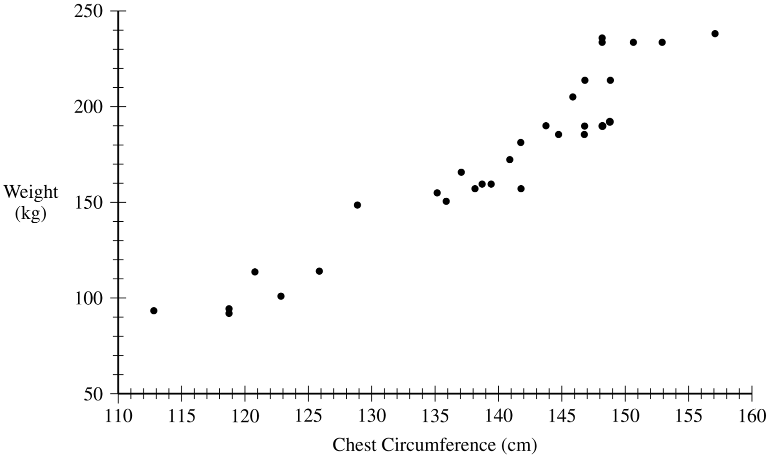
Wildlife biologists are interested in the health of tule elk, a species of deer found in California. An important measurement of tule elk health is their weight. The weight of a tule elk is difficult to measure in the wild. However, chest circumference, which is believed to be related to the weight of a tule elk, can easily be measured from a safe distance using a harmless laser. A study was done to investigate whether chest circumference, in centimeters (cm), could be used to accurately estimate the weight, in kilograms (kg), of male tule elk. For the study, wildlife biologists captured 30 male tule elk, measured their chest circumference and weight, and then released the elk. The data for the 30 male tule elk are shown in the scatterplot.
[Scatterplot titled with axes "Chest Circumference (cm)" (x-axis, 110 to 160) and "Weight (kg)" (y-axis, 50 to 250). The points show a strong, positive, roughly linear association: e.g. points near (113,90), (120,93), (121,98), (125,103), (130,145),

Question 5

(133,148), (135,150), (137,155), (138,160), (140,163), (141,175), (142,178), (143,180), (144,183), (145,204), (146,207), (147,210), (148,213), (149,233), (150,236), (152,238), (153,240), (155,241), continuing up to about (156,241), with the cloud of points trending upward from lower-left to upper-right.]

(a) Describe the relationship between chest circumference and weight of male tule elk in context.

Question 5


Weight (kg)
250
200
150
100
50
110
115
120
125
130
135
140
145
150
155
160
Chest Circumference (cm)

Question 5

2023 ▪ Q5

Wildlife biologists are interested in the health of tule elk, a species of deer found in California. An important measurement of tule elk health is their weight. The weight of a tule elk is difficult to measure in the wild. However, chest circumference, which is believed to be related to the weight of a tule elk, can easily be measured from a safe distance using a harmless laser. A study was done to investigate whether chest circumference, in centimeters (cm), could be used to accurately estimate the weight, in kilograms (kg), of male tule elk. For the study, wildlife biologists captured 30 male tule elk, measured their chest circumference and weight, and then released the elk. The data for the 30 male tule elk are shown in the scatterplot.

[Scatterplot titled with axes “Chest Circumference (cm)” (x-axis, 110 to 160) and “Weight (kg)” (y-axis, 50 to 250). The points show a strong, positive, roughly linear association: e.g. points near (113,90), (120,93), (121,98), (125,103), (130,145), (133,148), (135,150), (137,155), (138,160), (140,163), (141,175), (142,178), (143,180), (144,183), (145,204), (146,207), (147,210), (148,213), (149,233), (150,236), (152,238), (153,240), (155,241),

Question 5

continuing up to about (156,241), with the cloud of points trending upward from lower-left to upper-right.]

(b)(i) Following is the equation of the least-squares regression line relating chest circumference and weight for male tule elk.

$$\text{predicted weight} = -350.3 + 3.7455(\text{chest circumference})$$

The weight of one male tule elk with a chest circumference of 145.9 cm is 204.3 kg. Using the equation of the least-squares regression line, calculate the predicted weight for this male tule elk. Show your work.

(b)(ii) Calculate the residual for this male tule elk. Show your work.

Question 5

2023 ▪ Q5

Wildlife biologists are interested in the health of tule elk, a species of deer found in California. An important measurement of tule elk health is their weight. The weight of a tule elk is difficult to measure in the wild. However, chest circumference, which is believed to be related to the weight of a tule elk, can easily be measured from a safe distance using a harmless laser. A study was done to investigate whether chest circumference, in centimeters (cm), could be used to accurately estimate the weight, in kilograms (kg), of male tule elk. For the study, wildlife biologists captured 30 male tule elk, measured their chest circumference and weight, and then released the elk. The data for the 30 male tule elk are shown in the scatterplot.

[Scatterplot titled with axes "Chest Circumference (cm)" (x-axis, 110 to 160) and "Weight (kg)" (y-axis, 50 to 250). The points show a strong, positive, roughly linear association: e.g. points near (113,90), (120,93), (121,98), (125,103), (130,145), (133,148), (135,150), (137,155), (138,160), (140,163), (141,175), (142,178), (143,180), (144,183), (145,204), (146,207), (147,210), (148,213), (149,233), (150,236), (152,238), (153,240), (155,241),

Question 5

continuing up to about (156,241), with the cloud of points trending upward from lower-left to upper-right.]

(c) The equation of the least-squares regression line relating chest circumference and weight for male tule elk is repeated here.

$$\text{predicted weight} = -350.3 + 3.7455(\text{chest circumference})$$

Interpret the slope of the least-squares regression line in context.

Question 5

2023 ▪ Q5

Wildlife biologists are interested in the health of tule elk, a species of deer found in California. An important measurement of tule elk health is their weight. The weight of a tule elk is difficult to measure in the wild. However, chest circumference, which is believed to be related to the weight of a tule elk, can easily be measured from a safe distance using a harmless laser. A study was done to investigate whether chest circumference, in centimeters (cm), could be used to accurately estimate the weight, in kilograms (kg), of male tule elk. For the study, wildlife biologists captured 30 male tule elk, measured their chest circumference and weight, and then released the elk. The data for the 30 male tule elk are shown in the scatterplot.

[Scatterplot titled with axes "Chest Circumference (cm)" (x-axis, 110 to 160) and "Weight (kg)" (y-axis, 50 to 250). The points show a strong, positive, roughly linear association: e.g. points near (113,90), (120,93), (121,98), (125,103), (130,145), (133,148), (135,150), (137,155), (138,160), (140,163), (141,175), (142,178), (143,180), (144,183), (145,204), (146,207), (147,210), (148,213), (149,233), (150,236), (152,238), (153,240), (155,241),

Question 5

continuing up to about (156,241), with the cloud of points trending upward from lower-left to upper-right.]

(d)(i) The sambar, another species of deer, is similar in size to the tule elk. The slope of the population regression line relating chest circumference and weight for all male sambars is 4.5 kilograms per centimeter. A wildlife biologist wants to determine whether the slope of the population regression line for male tule elk is different than that for male sambars. Let β represent the slope of the population regression line for male tule elk. The wildlife biologist conducted a test of the following hypotheses using the sample of 30 tule elk.

$$H_0 : \beta = 4.5$$

$$H_a : \beta \neq 4.5$$

Question 5

The test statistic was calculated to be 3.408. Assume all conditions for inference were met.
Determine the *p*-value of the test.
(d)(ii) At a significance level of $\alpha = 0.05$, what conclusion should the wildlife biologist make regarding the slope of the population regression line for male tule elk? Justify your response.

Solution 5

(a) Mark scheme

- The scatterplot reveals a strong, positive, roughly linear association between chest circumference and weight of male tule elk. There are no points that seriously deviate from the straight-line pattern of the points in the plot. Full credit needs a description that includes at least 3 of components 1-4 AND component 5: (1) direction of association (positive/increasing), (2) strength of association (strong), (3) form of association (linear or approximately linear), (4) unusual features (no points with large discrepancies from the straight-line pattern exhibited by most points - it is sufficient to simply state there are no unusual features/outliers), (5) context - association between chest circumference and weight of tule elk (minimally sufficient to refer to the association between chest

Solution 5

circumference/measurement and weight without explicitly naming tule elk). Partial credit if the response satisfies only 1-2 of components 1-4 AND component 5, OR at least 2 of components 1-4 but not component 5.

Solution 5

(b)(i) Mark scheme

- Using the least-squares regression equation predicted weight = -350.3 + 3.7455(chest circumference), the predicted weight of a male tule elk with a chest circumference of 145.9 cm is -350.3 + 3.7455(145.9) $\approx$ 196.17 kg. NOTE: parts (b)(i) and (b)(ii) are scored TOGETHER as ONE holistic unit in the official scoring guidelines, so this leaf carries this part's full mark. Full credit for the combined part (b) needs both: (1) provides the correct predicted value 196.17 kg in part (b)(i) with work shown, (2) provides the correct residual 8.13 kg in part (b)(ii) [graded together with leaf 5(b)(ii)], or a value consistent with the predicted weight calculated here, with work shown. Partial credit if the response satisfies only one of the two components, OR gives both correct values with no work

Solution 5

shown. A response using an undefined or incorrect variable to identify the work shown (e.g. x or p-hat instead of predicted weight) does not satisfy component 1. An arithmetic or transcription error can be ignored if correct work is shown.

Solution 5

(b)(ii) Mark scheme

- The residual for the male tule elk with chest circumference 145.9 cm and actual weight 204.3 kg is residual $=$ actual $-$ predicted $= 204.3 - 196.17 \approx 8.13$ kg. This leaf's credit is bundled with 5(b)(i) in the official scoring guidelines (Part (b) of the SG scores both sub-parts together as one holistic unit), so it carries 0 of this question's marks on its own - but a correct residual of 8.13 kg (or a value consistent with the predicted weight from part (b)(i)), with work shown, is still expected to complete credit for part (b). An arithmetic or transcription error can be ignored if correct work is shown.

Solution 5

(c) Mark scheme

- The value of the slope of the least-squares regression line is 3.7455 (may be rounded to 3.746, 3.75, or 3.7, but not to 3 or 4). This value indicates that the predicted weight of a tule elk increases by 3.7455 kilograms for each additional centimeter of chest circumference. Full credit needs all 3 components: (1) identifies the value of the slope as 3.7455 (a response that only contains 3.7455 in the interpretation still satisfies this), (2) provides an interpretation that references an increase of a number of kilograms of weight for each one-centimeter increase in chest circumference (units must be correctly specified for both weight and length), (3) indicates the slope represents a change in a PREDICTION using non-deterministic language such as 'predicted,' 'estimated,' 'expected,' or

Solution 5

'average' (not a deterministic claim that weight definitely increases). Partial credit if only 2 of the 3 components are satisfied. It is not required to refer specifically to the 'least-squares regression line.'

Solution 5

(d)(i) Mark scheme

- The degrees of freedom for the test of the population slope (relating chest circumference and weight for male tule elk, testing H0: β=4.5 against a two-sided alternative, versus the sambar species' known slope of 4.5 kg/cm) are n − 2 = 30 − 2 = 28. The t-table shows that for 28 degrees of freedom, the p-value for a one-sided test would be 0.001. Because this is a two-sided test, the p-value is (2)(0.001) = 0.002. NOTE: parts (d)(i) and (d)(ii) are scored TOGETHER as ONE holistic unit in the official scoring guidelines, so this leaf carries this part's full mark. Full credit for the combined part (d) needs all 3: (1) gives a correct p-value of 0.002, consistent with the two-sided alternative hypothesis, (2) provides a correct comparison of the p-value to alpha AND a correct decision about the

Solution 5

null/alternative hypothesis [graded together with leaf 5(d)(ii)], (3) provides a conclusion statement in context, consistent with and in terms of the alternative hypothesis, using non-deterministic language [also from part (d)(ii)]. Partial credit if only 2 of the 3 components are satisfied.

Solution 5

(d)(ii) Mark scheme

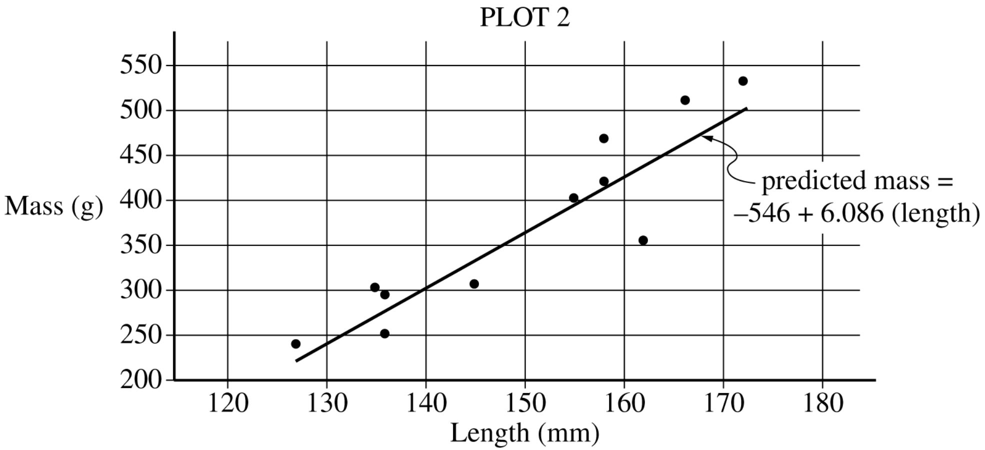
- Because the p-value (0.002) is less than $\alpha = 0.05$, reject the null hypothesis. There is sufficient statistical evidence that the population slope for the linear regression of weight vs. chest circumference for male tule elk is different from 4.5 kg/cm (i.e. different from the population slope for male sambars). This leaf's credit is bundled with 5(d)(i) in the official scoring guidelines (Part (d) of the SG scores both sub-parts together as one holistic unit), so it carries 0 of this question's marks on its own - but a correct decision (reject H0) with a correctly-contextualized, non-deterministic conclusion is still expected to complete credit for part (d). An explicit decision statement is not strictly required if the decision is clearly implied within the conclusion.

Question 1

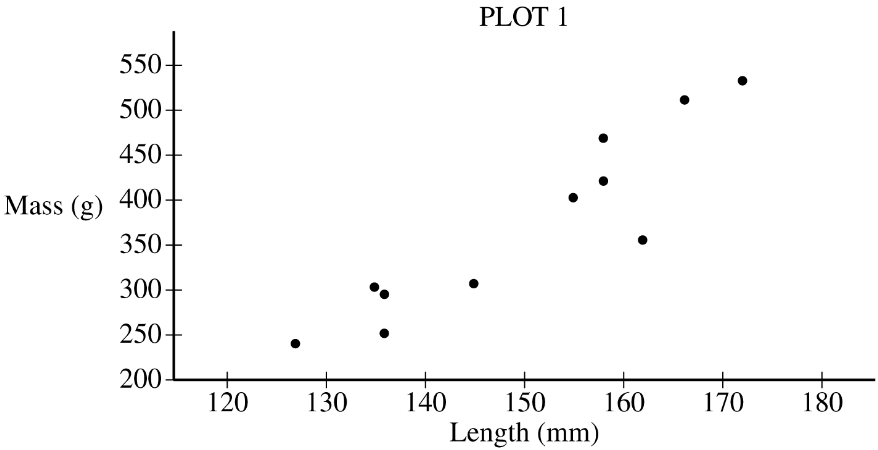
2022 ▪ Q1

A biologist gathered data on the length, in millimeters (mm), and the mass, in grams (g), for 11 bullfrogs. The data are shown in Plot 1.
[Plot 1: Scatterplot titled “PLOT 1”; x-axis “Length (mm)” from 120 to 180; y-axis “Mass (g)” from 200 to 550. Approximate points (length, mass): (127, 240), (135, 295), (136, 300), (136, 255), (145, 310), (155, 400), (157, 470), (158, 420), (162, 355), (166, 510), (172, 530).]
(a) Based on the scatterplot, describe the relationship between mass and length, in context.

Question 1


PLOT 2
Mass (g)
550
500
450
400
350
300
250
200
120
130
140
150
160
170
180
Length (mm)
predicted mass =
–546 + 6.086 (length)

Question 1


PLOT 1
Mass (g)
550
500
450
400
350
300
250
200
120
130
140
150
160
170
180
Length (mm)

Question 1

2022 ▪ Q1

A biologist gathered data on the length, in millimeters (mm), and the mass, in grams (g), for 11 bullfrogs. The data are shown in Plot 1.

[Plot 1: Scatterplot titled "PLOT 1"; x-axis "Length (mm)" from 120 to 180; y-axis "Mass (g)" from 200 to 550. Approximate points (length, mass): (127, 240), (135, 295), (136, 300), (136, 255), (145, 310), (155, 400), (157, 470), (158, 420), (162, 355), (166, 510), (172, 530).]

(b) From the data, the biologist calculated the least-squares regression line for predicting mass from length. The least-squares regression line is shown in Plot 2.

[Plot 2: Same scatterplot as Plot 1 with a least-squares regression line drawn through the points, labeled "predicted mass = -546 + 6.086 (length)"; x-axis "Length (mm)" from 120 to 180, y-axis "Mass (g)" from 200 to 550.]

Identify and interpret the slope of the least-squares regression line in context.

Question 1

2022 ▪ Q1

A biologist gathered data on the length, in millimeters (mm), and the mass, in grams (g), for 11 bullfrogs. The data are shown in Plot 1.

[Plot 1: Scatterplot titled "PLOT 1"; x-axis "Length (mm)" from 120 to 180; y-axis "Mass (g)" from 200 to 550. Approximate points (length, mass): (127, 240), (135, 295), (136, 300), (136, 255), (145, 310), (155, 400), (157, 470), (158, 420), (162, 355), (166, 510), (172, 530).]

(c) Interpret the coefficient of determination of the least-squares regression line, $r^2 \approx 0.819$, in context.

Question 1

2022 ▪ Q1

A biologist gathered data on the length, in millimeters (mm), and the mass, in grams (g), for 11 bullfrogs. The data are shown in Plot 1.

[Plot 1: Scatterplot titled "PLOT 1"; x-axis "Length (mm)" from 120 to 180; y-axis "Mass (g)" from 200 to 550. Approximate points (length, mass): (127, 240), (135, 295), (136, 300), (136, 255), (145, 310), (155, 400), (157, 470), (158, 420), (162, 355), (166, 510), (172, 530).]

(d)(i) From Plot 2, consider the residuals of the 11 bullfrogs.

Based on the plot, approximately what is the length and mass of the bullfrog with the largest absolute value residual?

(d)(ii) Does the least-squares regression line overestimate or underestimate the mass of the bullfrog identified in part (d-i)? Explain your answer.

Solution 1

(a) Mark scheme

- The scatterplot reveals a strong, positive, roughly linear association between the mass and length of bullfrogs. There are no points that seriously deviate from the straight-line pattern of the points in the plot. Full credit needs a description including at least three of: (1) direction of association (positive/increasing), (2) strength of association (strong), (3) form (linear or approximately linear), (4) no unusual features (no points with large discrepancies from the pattern), plus (5) context (association between length and mass of bullfrogs, or between length and mass in general).

Solution 1

(b) Mark scheme

- The slope of the least-squares regression line is 6.086 (may be rounded to 6.09 or 6.1, but not to 6). This indicates the predicted mass of a bullfrog increases by 6.086 grams for each additional millimeter of length. Full credit needs all three: (1) identifies the slope value as 6.086, (2) an interpretation referencing an increase of a number of grams of mass for each one-millimeter increase in length (correct units for both mass and length), (3) uses non-deterministic language such as 'predicted', 'estimated', 'expected', or 'average' rather than stating the change as certain.

Solution 1

(c) Mark scheme

- The coefficient of determination is $r^2 \approx 0.819$. This value indicates that 81.9% of the variation in bullfrog mass can be explained by variation in bullfrog length as described by the least-squares regression line. The interpretation must explicitly relate to the dependent (response) variable, mass, being explained by the independent (explanatory) variable, length – a generic definition with no context, an interpretation of r (not r^2) instead, or an interpretation that references the predicted y-values or the data (rather than mass explained by length) earns only partial credit.

Solution 1

(d)(i)

Mark scheme

- From Plot 2, the largest residual in absolute value belongs to the bullfrog with length approximately 162 millimeters and mass approximately 356 grams (any length between 160 and 165 mm and mass between 350 and 375 g, read from the plot, is accepted as reasonable). This identification is the first of two components that together earn the single point shared with part (d-ii): full credit for that shared point additionally requires correctly stating in (d-ii) whether the regression line overestimates or underestimates this bullfrog's mass.

Solution 1

(d)(ii) Mark scheme

- The least-squares regression line overestimates the mass of the bullfrog identified in part (d-i): Plot 2 shows the point for the bullfrog with length 162 mm lies below the least-squares regression line (the predicted mass at length 162, using $-546 + (6.086)(162) = 439.9$ grams, is larger than the bullfrog's actual mass of about 356 grams). This single point is shared between (d-i) and (d-ii) and requires both: (1) correctly identifying the bullfrog in (d-i) (length between 160-165 mm, mass between 350-375 g), and (2) explicitly stating in (d-ii) whether the line overestimates or underestimates the mass, with a justification comparing the identified point to the regression line (a numerical predicted-mass value is not required, but if given must be reasonable, e.g. between 425 and 450 g).

Question 1

2017 ▪ Q1

Researchers studying a pack of gray wolves in North America collected data on the length x, in meters, from nose to tip of tail, and the weight y, in kilograms, of the wolves. A scatterplot of weight versus length revealed a relationship between the two variables described as positive, linear, and strong.

(a)(i) For the situation described above, explain what is meant by each of the following words.

Positive:

(a)(ii) Linear:

(a)(iii) Strong:

(b) The data collected from the wolves were used to create the least-squares equation $\hat{y} = -16.46 + 35.02x$.

Interpret the meaning of the slope of the least-squares regression line in context.

Question 1

(c) One wolf in the pack with a length of 1.4 meters had a residual of -9.67 kilograms. What was the weight of the wolf?

Solution 1

(a)(i) Mark scheme

- 'Positive' means that wolves with higher values of length also tend to have higher values of weight – low length goes with low weight, and high length goes with high weight (as length increases, weight tends to increase). Acceptable phrasing: "As length increases, so does weight"; "Longer wolves weigh more"; "The points on the graph go up as you move left to right." NOT acceptable: "As length goes up, weight changes"; "the correlation is greater than 0" (must describe the low-low/high-high pattern, not just restate the term). Full part (a) credit needs a definition like this for positive, one for linear, one for strong, AND at least one of the three stated in context (using length/weight or meters/kilograms).

Solution 1

(a)(ii) Mark scheme

- 'Linear' means that as length increases by one meter, weight tends to change by a constant amount, on average – the data points follow the pattern of a straight line (equivalently, the rate of change of weight with respect to length is constant). Acceptable: "The points generally follow a straight line"; "the relationship between x and y is straight"; "length and weight have a constant slope." NOT acceptable: "the points all line up"; "you can draw a straight line through the points"; "there is a positive correlation"; merely restating that every increase in x yields a 35.02 increase in y. See the context/four-component requirement noted under 1ai for part (a) as a whole.

Solution 1

(a)(iii) Mark scheme

- 'Strong' means the data points fall close to the line (or curve) – observed weights are close to predicted weights, i.e. residuals (deviations) from the least-squares line are small, or equivalently the correlation coefficient is close to 1 in magnitude. Acceptable: "Observed values are close to predicted values"; "Deviations from the least-squares regression line are small"; "The correlation coefficient is close to 1." NOT acceptable: "the points are close together"; "the scatterplots are clustered together"; "there is a high positive correlation" (these describe density of points, not closeness to a line). See the context/four-component requirement noted under 1ai for part (a) as a whole.

Solution 1

(b) Mark scheme

- The slope of 35.02 means: two wolves whose lengths differ by one meter are predicted to differ in weight by 35.02 kilograms, with the longer wolf having the greater weight. Full credit requires three components: (1) correctly cites the slope value 35.02; (2) interprets it as a specified increase in weight for each unit increase in length; (3) indicates the relationship is not exact/deterministic, using language such as "predicted" or "on average" (e.g. "...predicted to differ by 35.02 kg" or "...weight increases by 35.02 kg on average"). A response that only says "weight increases by 35.02 kg per meter" with no averaging/predicted language misses component 3. Note: if the response instead identifies -16.46 (the

Solution 1

intercept) as the slope, component 2 is satisfied only if it states weight DECREASES by 16.46 kg on average for each meter increase.

Solution 1

(c) Mark scheme

- In general, residual = actual weight - predicted weight, so actual weight = predicted weight + residual. For the wolf with length 1.4 m: predicted weight $= -16.46 + 35.02(1.4) = 32.568$ kilograms. Actual weight $= 32.568 + (-9.67) = 22.898$ kilograms (≈ 22.9 kg). Full credit needs both: (1) the correct predicted-weight computation (32.568 kg), and (2) the correct actual-weight computation obtained by ADDING the (negative) residual to the predicted value, giving 22.9 kg. A common error is subtracting instead of adding the residual (giving 42.24 kg), which does not earn full credit. Minor arithmetic slips that don't lead to an unreasonable value (e.g. a negative weight) can still be treated as showing the correct method.

Question 6

2014 ▪ Q6

Jamal is researching the characteristics of a car that might be useful in predicting the fuel consumption rate (FCR); that is, the number of gallons of gasoline that the car requires to travel 100 miles under conditions of typical city driving. The length of a car is one explanatory variable that can be used to predict FCR. Length of a car is measured in inches, and FCR is measured in gallons per 100 miles.
Jamal examined the scatterplot and determined that a linear model would be a reasonable way to express the relationship between FCR and length. A computer output from a linear regression is shown below.

Question 6

[Scatterplot titled “Graph I” plotting FCR (y-axis, 0 to 8) versus Length (inches) (x-axis, 150 to 210). Points are scattered showing a positive association between length and FCR, roughly increasing from about 3-4 at length 150-160 up to about 6-7 at length 200-210, with considerable scatter. One point is labeled A, located at approximately (175, 5.88).]

Linear Fit FCR = -1.595789 + 0.0372614 * Length

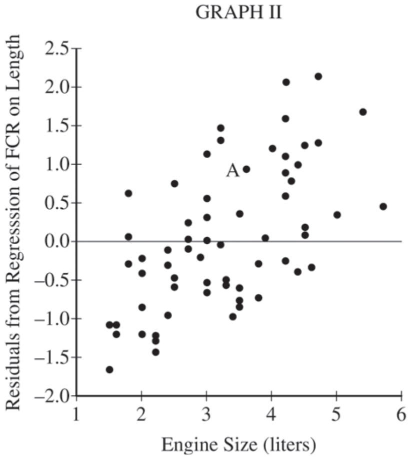
Summary of Fit

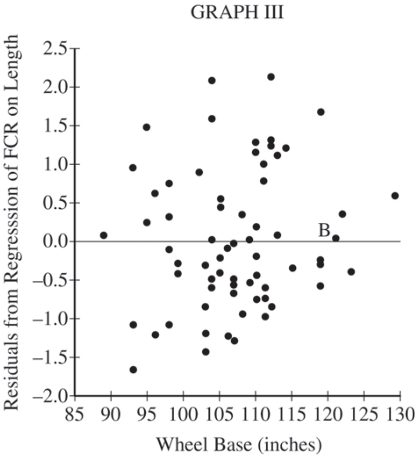
Statistic	Value
RSquare	0.250401
Root Mean Square Error	0.902382
Observations	66

Question 6

(a) The point the graph labeled A represents one car of length 175 inches and an FCR of 5.88. Calculate and interpret the residual for the car relative to the least squares regression line.

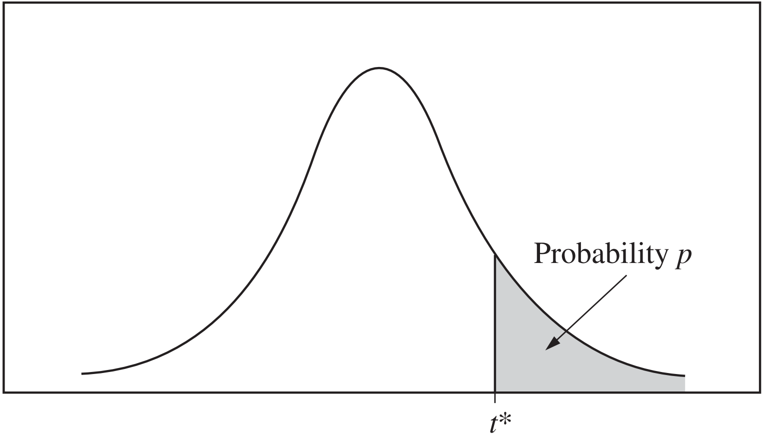
Question 6


GRAPH II
Residuals from Regresssion of FCR on Length
2.5
2.0
1.5
1.0
0.5
0.0
–0.5
–1.0
–1.5
–2.0
1
2
3
4
5
6
A
Engine Size (liters)


GRAPH III
Residuals from Regression of FCR on Length
2.5
2.0
1.5
1.0
0.5
0.0
–0.5
–1.0
–1.5
–2.0
85
90
95
100
105
110
115
120
125
130
B
Wheel Base (inches)

Question 6

Table entry for p and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^*.



Question 6

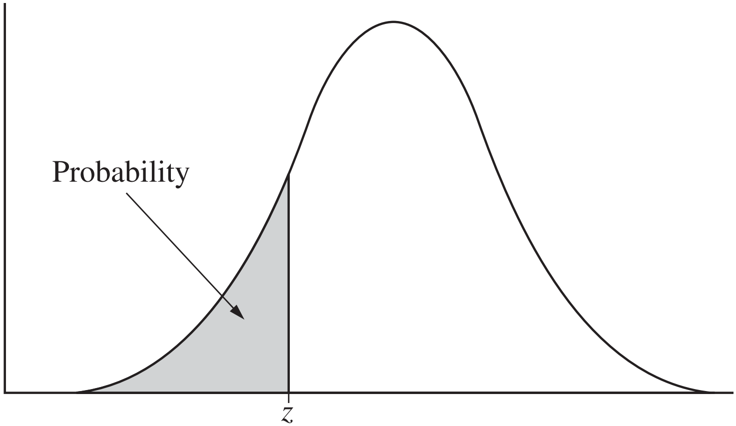


Table entry for z is the probability lying below z.