2022

AP CollegeBoard

AP® Statistics

Sample Student Responses and Scoring Commentary

Inside:

Free-Response Question 1

- ☑ **Scoring Guidelines**
- ☑ **Student Samples**
- ☑ **Scoring Commentary**

Question 1: Focus on Exploring Data — 4 points

General Scoring Notes

- Each part of the question (indicated by a letter) is initially scored by determining if it meets the criteria for essentially correct (E), partially correct (P), or incorrect (I). The response is then categorized based on the scores assigned to each letter part and awarded an integer score between 0 and 4 (see the table at the end of the question).
- The model solution represents an ideal response to each part of the question, and the scoring criteria identify the specific components of the model solution that are used to determine the score.

	Model Solution	Scoring
(a)	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.	**Essentially correct (E)** if the response provides a description that includes at least three of components 1-4 and component 5: 1. Direction of association (positive or increasing) 2. Strength of association (strong) 3. Form of association (linear or approximately linear) 4. Unusual features (no points with large discrepancies from the pattern (straight line) exhibited by most of the points on the plot) 5. Context (association between length and mass of bullfrogs) **Partially correct (P)** if the response satisfies only one or two components out of components 1-4 and component 5 *OR* if the response satisfies at least three out of components 1-4 but does not satisfy component 5. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- To satisfy component 4, it is sufficient to simply indicate that there are no unusual features.
- To satisfy component 5, it is minimally sufficient for the response to refer to the association or relationship between mass and length without explicitly mentioning bullfrogs.
- The strength of the response in part (a) may be considered if holistic scoring is needed.

	Model Solution	Scoring
(b)	The value of the slope of the least-squares regression line is 6.086. This value indicates that the predicted mass of a bullfrog increases by 6.086 grams for each additional millimeter of length.	**Essentially correct (E)** if the response satisfies the following three components: 1. Identifies the value of the slope as 6.086 2. Provides an interpretation that references an increase of a number of grams of mass for each one-millimeter increase in length 3. Indicates that the slope represents a change in a prediction using non-deterministic language such as "predicted," "estimated," "expected," or "average" **Partially correct (P)** if the response satisfies only two of the three components. **Incorrect (I)** if the response does not meet the criteria for E or P.

Additional Notes:

- The value of the slope, 6.086, may be rounded to 6.09 or 6.1, but not to 6, to satisfy the numerical requirement in component 1.
- A response that only contains 6.086 in the interpretation satisfies component 1.
- A calculation of slope may satisfy component 1, provided that two points from the line are used in the calculation.
- Units of measurements must be correctly specified for both mass and length to satisfy component 2.
- It is not required to refer specifically to the "least-squares regression line."

	Model Solution	Scoring
(c)	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 line.	**Essentially correct (E)** if the response provides a correct interpretation of r^2 in context. **Partially correct (P)** if the response provides a generic interpretation (no context) *OR* if the response provides a reasonable but incorrect interpretation of r^2 in context. **Incorrect (I)** if the response does not satisfy the criteria for E or P.

Additional Notes:

- Correct interpretations of r^2 include the concept that part of the variation in the response (dependent or *y*) variable is explained by the linear relationship with the explanatory (independent or *x*) variable. The response can take any of several equivalent forms, such as:
  - The proportion of the total variability in the dependent (response) variable *y* that is explained by the independent (explanatory) variable *x*.
  - The proportion of variation in *y* that is accounted for by the linear model.
  - The proportionate reduction of the total variation of the *y*-values that is associated with the use of the independent variable *x*.
  - The proportionate reduction in the sum of the squares of vertical deviations obtained by using the least-squares line instead of the sample mean to predict values of *y*.
- Correct interpretation of r^2 must explicitly relate to the dependent variable. Mention of the data, predicted values, or no mention of the dependent variable are incorrect interpretations. Common incorrect interpretations include:
  - The percent (or proportion or part of the total) variability in the predicted *y*-values that is explained by the linear relationship between *y* and *x*.
  - The percent (or proportion or part of the total) variability in the data that is explained by the linear relationship between *y* and *x*.
  - The percent (or proportion or part of the total) variability that is explained by the linear relationship between *y* and *x*.
  - The percent (or proportion or part of the total) variability in *y* that is on average explained by the linear relationship between *y* and *x*.
- A reasonable but incorrect interpretation of r^2 with context might include the following responses:
  - 81.9% of the variation in mass and length can be accounted for by the least-squares regression line.
  - 81.9% of the variability in predicted mass is accounted for by the length.
- For context, the response variable (*y*) must be identified as mass, and the explanatory variable (*x*) must be identified as length.
- An interpretation of the correlation between mass and length, $r = \sqrt{0.819} = 0.905$, is not considered a reasonable interpretation of r^2.
- The value of the percentage (81.9%) or proportion (0.819) of variation does not need to be specified, but if an incorrect value is specified, the score is lowered by one level, from E to P or from P to I.
- The strength of the response in part (c) may be considered if holistic scoring is needed.

Model Solution	Scoring
(d) (i) The largest residual in absolute value belongs to the bullfrog with length 162 millimeters and mass 356 grams. (ii) The least-squares regression line overestimates the mass of the bullfrog with length 162 millimeters. Plot 2 shows that the point for the bullfrog with length 162 millimeters is below the least-squares regression line.	**Essentially correct (E)** if the response satisfies the following two components: 1. The response to part (d-i) identifies the correct bullfrog (length between 160 and 165 millimeters, mass between 350 and 375 grams) 2. The response to part (d-ii) explicitly indicates whether the linear model overestimates or underestimates mass for the bullfrog identified in part (d-i) and provides a correct justification based on a comparison of the identified observation to the least-squares regression line **Partially correct (P)** if the response satisfies only one of the two components. **Incorrect (I)** if the response does not satisfy the criteria for E or P.

Additional Notes:

- The comparison of the observation to the regression line in the response to part (d-ii) is satisfied if the response does one of the following:
  - Correctly indicates if the observation is below (above) the least-squares regression line in Plot 2.
  - Notes that observed mass is smaller (larger) than the mass predicted by the least-squares regression line.
  - Marks the observation selected in part (d-i) on Plot 2 with an indication of the vertical distance from the least-squares regression line.
  - Notes the correct sign of the residual.
- Numerical values are not required in the response to part (d-ii). If a numerical value is given for the predicted mass, however, it must be reasonable. A numerical value for the predicted mass could be computed with the formula given in the stem, e.g., $-546 + (6.086)(162) = 439.9$ grams, for a bullfrog of length 162 millimeters, or a value can be read from the line shown in Plot 2. Any value between 425 and 450 should be considered a reasonable value. Showing work is not required.
- The word overestimate with the calculated predicted value of mass is enough to satisfy component 2.
- If the wrong observation is identified in part (d-i), the response to part (d) may be scored P if the response to part (d-ii) correctly compares that observation to the least-squares regression line and states the correct conclusion about overestimating or underestimating mass with justification.
- It is not required to refer specifically to the "least-squares regression line."

Scoring for Question 1 Each essentially correct (E) part counts as 1 point, and each partially correct (P) part counts as ½ point.	**Score**
Complete Response	**4**
Substantial Response	**3**
Developing Response	**2**
Minimal Response	**1**
If a response is between two scores (for example, 2 ½ points), use a holistic approach to decide whether to score up or down, depending on the strength of the response and quality of the communication.	

Question 1

Begin your response to **QUESTION 1** on this page.

STATISTICS

SECTION II

Total Time—1 hour and 30 minutes

6 Questions

Part A

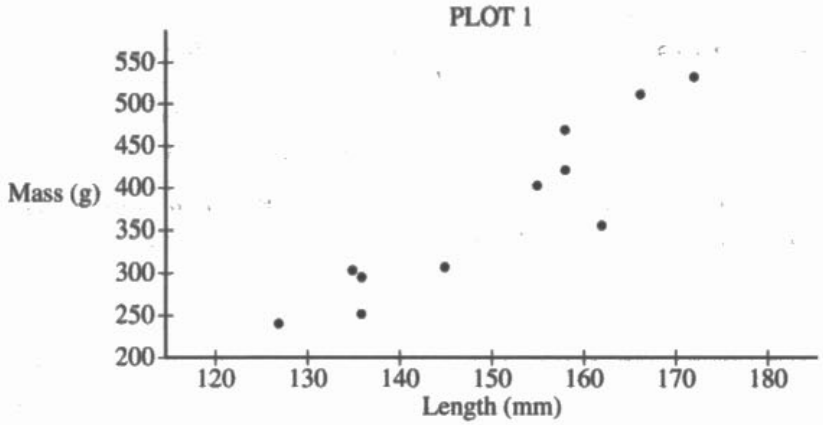
Suggested Time—1 hour and 5 minutes

5 Questions

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. 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



(a) Based on the scatterplot, describe the relationship between mass and length, in context.

There appears to be a moderately strong linear relationship between mass and length for the 11 bullfrogs. There are no obvious outliers.

GO ON TO THE NEXT PAGE.

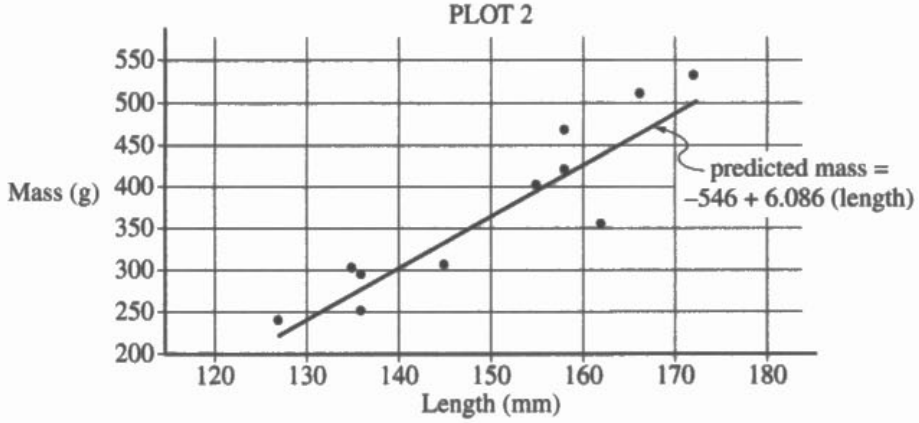
Use a pencil or a pen with black or dark blue ink. Do NOT write your name. Do NOT write outside the box.

Question 1

Continue your response to **QUESTION 1** on this page.

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



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

The predicted mass in grams of a bullfrog is expected to increase by 6.086 for every 1 mm increase in the length of a bullfrog

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

81.9% of the variation in the mass(g) of a bullfrog is accounted for by the least squares regression line on length(mm) of a bullfrog.

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

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

The bullfrog with the largest absolute value residual has a length of approximately 162 mm and a mass of approximately 355 g.

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

The least squares regression line would overestimate the mass of the bullfrog because the bullfrog has a negative residual with the observed value being less than the expected.

GO ON TO THE NEXT PAGE.

Use a pencil or a pen with black or dark blue ink. Do NOT write your name. Do NOT write outside the box.

0123552

Q5149/3

Question 1

Begin your response to **QUESTION 1** on this page.

STATISTICS

SECTION II

Total Time—1 hour and 30 minutes

6 Questions

Part A

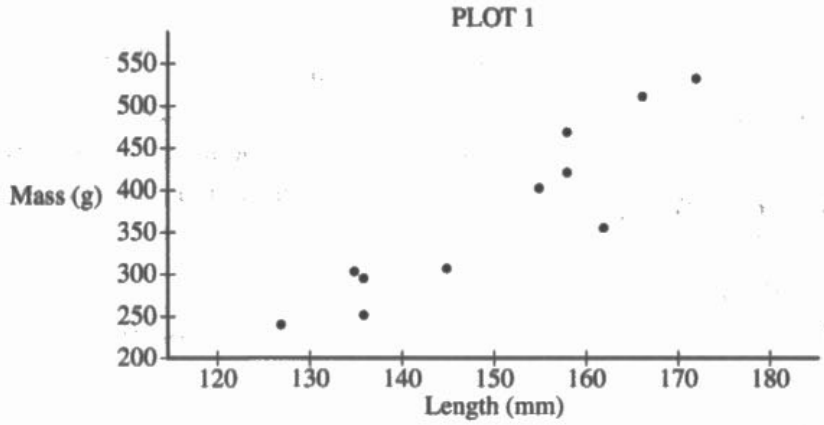
Suggested Time—1 hour and 5 minutes

5 Questions

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. 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



(a) Based on the scatterplot, describe the relationship between mass and length, in context.

The mass and length of bullfrogs have a linear positive relationship. That is, as length in mm increases, mass in g increase as well. There seem to be no extreme outliers.

GO ON TO THE NEXT PAGE.

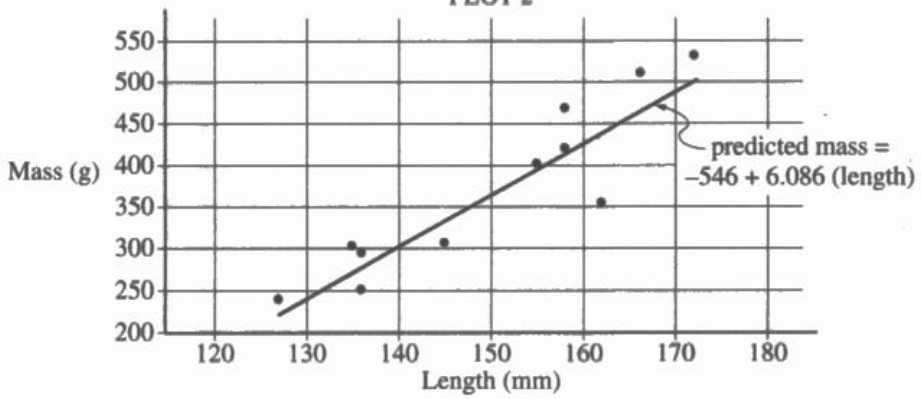
Use a pencil or a pen with black or dark blue ink. Do NOT write your name. Do NOT write outside the box.

Question 1

Continue your response to **QUESTION 1** on this page.

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



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

The slope is 6.086. This means that for every one mm increase in length, mass increases by 6.086 g on average.

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

81.9% of the variability in predicted mass is accounted for by the least-squares regression line.

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

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

The bullfrog with the largest absolute value residual has approximately a length of 162 mm and a mass of 351 g.

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

The least-squares regression line overestimates the mass of the bullfrog because, with a length of about 162 mm, it predicted the mass to be about 440 g. 440 g is greater than the observed value of about 351 g.

GO ON TO THE NEXT PAGE.

Use a pencil or a pen with black or dark blue ink. Do NOT write your name. Do NOT write outside the box.

0095082

Q6149/3

Question 1

Begin your response to **QUESTION 1** on this page.

STATISTICS

SECTION II

Total Time—1 hour and 30 minutes

6 Questions

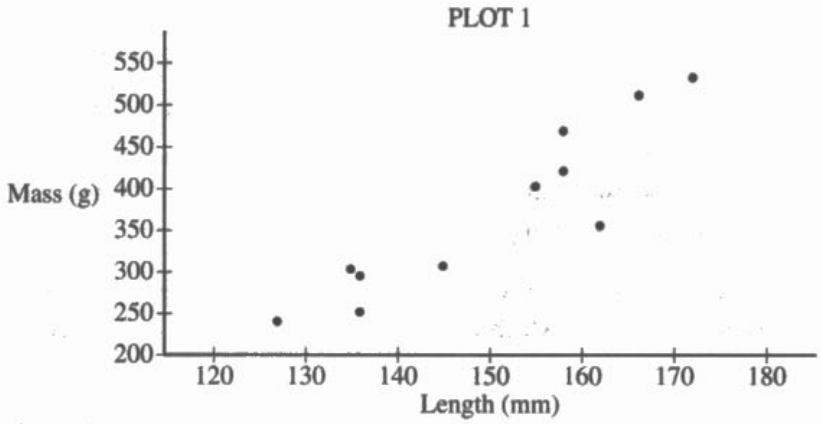
Part A

Suggested Time—1 hour and 5 minutes

5 Questions

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. 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.



(a) Based on the scatterplot, describe the relationship between mass and length, in context.

There is a strong, positive relationship between length and mass. As the length of a bullfrog increases, the mass also increases.

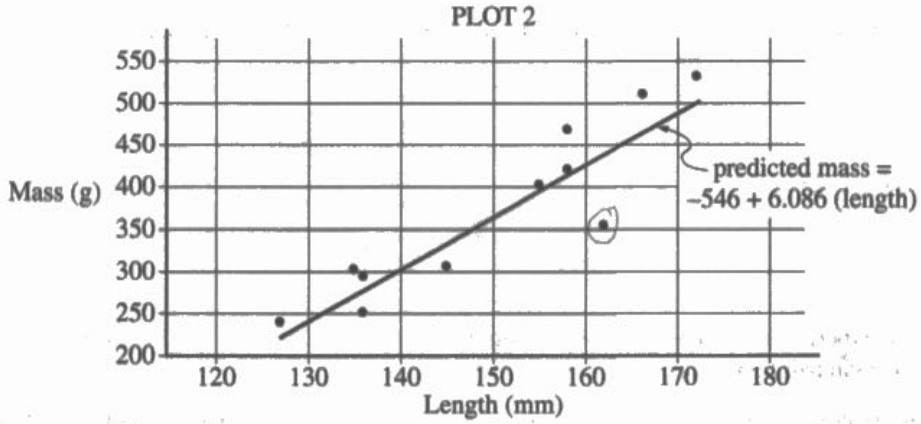
GO ON TO THE NEXT PAGE.

Use a pencil or a pen with black or dark blue ink. Do NOT write your name. Do NOT write outside the box.

Question 1

Continue your response to **QUESTION 1** on this page.

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.



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

Slope = 6.086 g/mm

For every 1 mm increase in length of a bullfrog, there is a 6.086g increase in mass of the bullfrog.

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

The typical discrepency between length and mass of a bullfrog is approximately 0.819 g/mm.

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

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

length ≈ 162 mm

mass ≈ 352 g

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

The least-squares regression line overestimates the mass of the bullfrog because the line is above the point on plot 2.

GO ON TO THE NEXT PAGE.

Use a pencil or a pen with black or dark blue ink. Do NOT write your name. Do NOT write outside the box.

0196992

Q5149/3

Question 1

Note: Student samples are quoted verbatim and may contain spelling and grammatical errors.

Overview

The primary goals of this question were to access a student's ability to (1) use data presented as a scatterplot to describe a relationship between two variables within the context of a study; (2) identify and interpret the slope of a least-squares regression line; (3) interpret the coefficient of determination with respect to the proportion of variation in values of the response variable that can be explained by variation in the values of the explanatory variable; (4) identify the observation with the largest absolute residual using a scatterplot of the data with the least-squares regression line inserted; and (5) determine if the least-squares regression line overestimates or underestimates the value of the response for the identified observation and provide a justification based on a comparison of the identified observation to the least squares regression line.

This question primarily assesses skills in skill category 2: Data Analysis, and skill category 4: Statistical Argumentation. Skills required for responding to this question include (2.A) Describe data presented numerically or graphically, (2.C) Calculate summary statistics, relative positions of points within a distribution, correlation, and predicted response, and (4.B) Interpret statistical calculations and findings to assign meaning or assess a claim.

This question covers content from Unit 2: Exploring Two-Variable Data of the course framework in the AP Statistics Course and Exam Description. Refer to topics 2.4, 2.6, 2.7, and 2.8, and learning objectives DAT-1.A, DAT-1.D, DAT-1.F, DAT-1.G, and DAT-1.H.

Sample: 1A
Score: 4

The response earned the following: Part (a) – E; Part (b) – E; Part (c) – E; Part (d) – E.

In part (a) the response includes strength, form, and unusual features, satisfying three of the first four components, and satisfies component 5, context. Part (a) was scored essentially correct (E).

In part (b) the response identifies the value of the slope by providing interpretation with context and includes "expected" for non-deterministic language, satisfying both components. Part (b) was scored essentially correct (E).

In part (c) the response includes a correct interpretation of r^2 in context. Part (c) was scored essentially correct (E).

In part (d) the response identifies the correct bullfrog and indicates that it will be an overestimate with "negative residual" for justification, satisfying both components. Part (d) was scored essentially correct (E).

Sample: 1B
Score: 3

The response earned the following: Part (a) – E; Part (b) – E; Part (c) – I; Part (d) – E.

In part (a) the response includes direction, form, and unusual features, satisfying three of the first four components, and satisfies component 5, context. The response does not satisfy component 2, strength. Part (a) was scored essentially correct (E).

Question 1 (continued)

In part (b) the response identifies the value of the slope, provides an interpretation with context, and includes "on average" for non-deterministic language. All three components are satisfied. Part (b) was scored essentially correct (E).

In part (c) the response does not include a correct interpretation of r^2 because it refers to the "predicted mass." The response contains mass but does not contain length; therefore, the context is not met. Part (c) was scored incorrect (I).

In part (d) the response identifies the correct bullfrog and indicates that it will be an overestimate with justification by calculating the predicted point. All three components are met. Part (d) was scored essentially correct (E).

Sample: 1C
Score: 2

The response earned the following: Part (a) – P; Part (b) – P; Part (c) – I; Part (d) – E.

In part (a) the response includes direction and strength, satisfying components 1 and 2. The response includes context, satisfying component 5. Part (a) was scored partially correct (P).

In part (b) the response identifies the value of the slope, satisfying component 1. The response provides a correct interpretation with units, satisfying component 2. However, the response does not include any non-deterministic language, so it does not satisfy component 3. Part (b) was scored partially correct (P).

In part (c) the response does not include a correct interpretation of r^2 and no context is included. Part (c) was scored incorrect (I).

In part (d) the response identifies the correct bullfrog and indicates that it will be an overestimate with justification as "the line is above the point." All three components are satisfied. Part (d) was scored essentially correct (E).