2017

AP CollegeBoard

AP Statistics

Sample Student Responses and Scoring Commentary

Inside:

- ☑ Free Response Question 1
- ☑ Scoring Guideline
- ☑ Student Samples
- ☑ Scoring Commentary

Question 1

Intent of Question

The primary goals of this question were to assess a student's ability to (1) explain statistical terms used when describing the relationship between two variables; (2) interpret the slope of a linear regression equation; and (3) calculate a value of y when given a regression equation, a value of x, and a residual.

Solution

Part (a):

In the context of a scatterplot in which y represents weight and x represents length, the following are defined.

A positive relationship means that wolves with higher values of length also tend to have higher weights.

A linear relationship means that as length increases by one meter, weight tends to change by a constant amount, on average.

A strong relationship means that the data points fall close to a line (or curve).

Part (b):

The slope of 35.02 indicates that two wolves that differ by one meter in length are predicted to differ by 35.02 kilograms in weight, with the longer wolf having the greater weight.

Part (c):

In general, a residual is equal to actual weight minus predicted weight, or equivalently,

$$\text{actual weight} = \text{predicted weight} + \text{residual}.$$

For the wolf with length 1.4 meters and residual of -9.67, the predicted weight is

$$-16.46 + 35.02(1.4) = 32.568 \text{ kilograms}.$$

Therefore, the actual weight of the wolf is $32.568 + (-9.67) = 22.898$ kilograms.

Scoring

Parts (a), (b), and (c) are scored as essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is scored as follows:

Essentially correct (E) if the response includes the following four components:

1. A reasonable definition of positive
2. A reasonable definition of linear
3. A reasonable definition of strong
4. At least one definition in context

Question 1 (continued)

Partially correct (P) if the response includes only three of the four components.

Incorrect (I) if the response does not meet the conditions for E or P.

Notes:

- The description of a positive relationship should clearly indicate that relatively low values of one variable tend to appear with relatively low values of the other variable, and relatively high values of the first variable tend to appear with relatively high values of the other variable.
    - Examples of acceptable responses:
        - As length increases, so does weight.
        - Longer wolves weigh more.
        - The points on the graph go up as you move from left to right.
    - Examples of unacceptable responses:
        - As length goes up, weight changes.
        - Both length and weight get bigger.
        - The correlation is greater than 0.
- The description of a linear relationship can take one of two approaches: the data pattern (data points exhibit the pattern of a line in the graph) or the constant rate of change (as the explanatory variable changes, the response variable exhibits a constant rate of change).
    - Examples of acceptable responses:
        - The points generally follow a straight line.
        - The relationship between x and y is straight.
        - Length and weight have a constant slope.
    - Examples of unacceptable responses:
        - The points all line up.
        - You can draw a straight line through the points.
        - There is a positive correlation.
        - Every increase in x yields a 35.02 increase in y.
- The description of strong should indicate how close points are to a line.
    - Examples of acceptable responses:
        - Observed values are close to predicted values.
        - Deviations from the least-squares regression line are small.
        - The correlation coefficient is close to 1.
    - Examples of unacceptable responses:
        - All the points are close together.
        - The scatterplots are clustered together.
        - There is a high positive correlation.
- Context can be shown by referring to length and weight or by using meters and kilograms.
- Sketches and graphs can be used to help clarify definitions, but a sketch alone cannot satisfy a definition component.

Part (b) is scored as follows:

Essentially correct (E) if the response includes the following three components:

1. The correct value of 35.02 for the slope.
2. An interpretation that includes an increase of a specified amount of weight for each unit increase in length.
3. An indication that the relationship is not exact by using words such as “on average” or “predicted weight.”

Question 1 (continued)

Partially correct (P) if the response includes only two of the three components.

Note: If the response identifies the slope as -16.46 (the intercept value), the second component is satisfied only if the response states that for each one-meter increase in length there is a *decrease* in predicted or average weight of 16.46 kilograms.

Incorrect (I) if the response does not meet the criteria for E or P.

Part (c) is scored as follows:

Essentially correct (E) if the response includes the following two components:

1. A correct computation for the predicted value 32.568 kilograms.
2. A correct computation for the actual weight 22.9 kilograms using the given residual and the predicted value.

Partially correct (P) if the response provides a correct computation for the predicted value but is not able to complete the correct calculation of the actual weight, including if the residual is defined in the wrong direction as (predicted weight) – (actual weight) to give an answer of 42.24 kilograms;

OR

if the response provides an incorrect value for the predicted weight, but then uses that value correctly to determine the actual weight as $\text{(predicted weight)} + \text{residual} = \text{(predicted weight)} + (-9.67)$;

OR

if the response provides a correct answer for the actual weight but does not give sufficient information to determine how it was calculated.

Incorrect (I) if the response does not meet the criteria for E or P.

Notes:

- The expression $-16.46 \quad 35.02(1.4)$ is enough to satisfy the first component.
- The equation $-16.46 + 35.02(1.4) - 9.67 = 22.9$ satisfies both components.
- Arithmetic mistakes are overlooked if they do not lead to an unreasonable answer (such as a negative value). For example, $32.568 + (-9.67) = 21.9$ satisfies the second component.

Question 1 (continued)

4 Complete Response

Three parts essentially correct

3 Substantial Response

Two parts essentially correct and one part partially correct

2 Developing Response

Two parts essentially correct and no parts partially correct

OR

One part essentially correct and one or two parts partially correct

OR

Three parts partially correct

1 Minimal Response

One part essentially correct

OR

No parts essentially correct and two parts partially correct

STATISTICS
SECTION II
Part A
Questions 1-5
Spend about 65 minutes on this part of the exam.
Percent of Section II score—75

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

   (i) Positive:

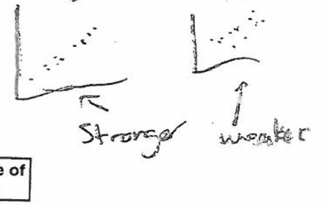
   In general, as the length of the gray wolves increases the general trend of the data suggests that weight will increase as well. A positive association means that the data is in general going upwords from left to right when graphing length, x, vs weight, y.

   (ii) Linear:

   Linear means that the data appears to be clustered around a line, and the data is not better followed by another curve. the graph between length and weight is best described by a line.

   (iii) Strong:

   This means that the data is well modelled by a line The residuals are generally fairly small, and the line is a good fit.



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

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

As the length of a gray wolf increases by one meter, the least squares regression line predicts that the weight in kilograms of a wolf would increase by approximately 35.02.

(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?

$y - \hat{y} = -9.67$ from the residual

$x = 1.4$ m, the actual length of the wolf.

$\hat{y} = -16.46 + 35.02 \cdot 1.4 = 32.57$ kg

$y = \hat{y} - 9.67 = 32.57 - 9.67 = \boxed{22.90 \text{ kg}}$

STATISTICS
SECTION II
Part A
Questions 1-5
Spend about 65 minutes on this part of the exam.
Percent of Section II score—75

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

   (i) Positive:

   Positive means there is a positive correlation between length, in meters, of the wolves and weight, in kilograms, of the wolves. As the length of the wolf increases it is expected that the weight will also increase.

   (ii) Linear:

   The data follows a linear pattern. As length increases, weight increases. The data points follow approximately the same pattern. Because it is linear we can also assume the residual plot follows no clear pattern.

   (iii) Strong:

   Strong implies that the data points for each wolf are closely related to each other. The points in the scatterplot follow fairly closely to the least squares regression line, also implying that the residuals for the points are small.

GO ON TO THE NEXT PAGE.

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

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

As the length of a wolf increases by one meter (x variable), the weight of the wolf is predicted to increase by 35.02 kilograms (y variable).

(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?

$-16.46 + 35.02(1.4 \text{ meters}) = 32.568$ expected kilograms

$W - 32.568 = -9.67$

$+32.568 \quad +32.568$

$\boxed{W = 22.898 \text{ kilograms}}$

GO ON TO THE NEXT PAGE.

STATISTICS
SECTION II
Part A
Questions 1-5

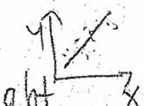
Spend about 65 minutes on this part of the exam.
Percent of Section II score—75

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

   (i) Positive:



When the data is plotted, there is an upward-right trend between the length and weight of the wolves

   (ii) Linear:

when the data is plotted on the scatterplot, there is a strong relationship between the length and weight of the wolves with r close to 1 or −1 (in this case r is close to 1).

   (iii) Strong:

When the data is plotted, the relationship between the length and weight of the wolves is strong which can be seen when drawing a line of best fit, the data should be near the regression line (no outliers).

1C 2

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

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

For every increase of the length of the wolves, the predicted weight increases by 35.02.

(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?

Residual = observed − predicted

$\hat{y} = -16.46 + 35.02(1.4)$

$\hat{y} = 65.488$

−9.67 = observed − 65.488

observed = 55.818

The weight of the wolf is 55.818 kilograms

GO ON TO THE NEXT PAGE.

Question 1

Overview

The primary goals of this question were to assess a student's ability to (1) explain statistical terms used when describing the relationship between two variables; (2) interpret the slope of a linear regression equation; and (3) calculate a value of y when given a regression equation, a value of x, and a residual.

Sample: 1A
Score: 4

Each of the two sentences in the response to part (a-i) provides a reasonable definition of a positive relationship. They both indicate that the value of one variable tends to increase as the value of the other variable increases. The first sentence also satisfies the context requirement for part (a) because it describes the relationship in terms of length and weight. The first sentence of the response to part (a-ii) provides a reasonable definition of a linear relationship by indicating that the points on the scatterplot exhibit a straight line pattern. The second sentence provides context. The response to part (a-iii) provides a reasonable definition of the strength of a relationship. It indicates that a relationship is strong when the points on the scatterplot are close to a line with the phrase "residuals are generally fairly small." The two graphs displayed in the response provide an alternative acceptable explanation because one graph illustrates a strong relationship, and the other illustrates a weaker relationship. Because reasonable definitions are given for all three components, and at least one is given in context, part (a) was scored as essentially correct. In part (b) the response provides an acceptable interpretation of the slope of the least-squares regression line by indicating that the "least squares regression line predicts that the weight in kilograms of a wolf would increase by approximately 35.02" for each 1-meter increase in length. It uses the correct value for the slope, and it clearly indicates that the least-squares regression line does not describe changes in the actual weights of wolves. The response is presented in the context of lengths and weights of wolves. Part (b) was scored as essentially correct. In part (c) the response provides a correct formula for a residual as a difference between an actual weight and a predicted weight. It displays the correct formula and reports the correct value for the predicted weight. It shows how to use the residual to compute the actual weight of the wolf from the predicted weight. Part (c) was scored as essentially correct. Because three parts were scored as essentially correct, the response earned a score of 4.

Sample: 1B
Score: 3

The statement about "positive correlation" in the first sentence of the response to part (a-i) essentially defines a positive relationship as a positive relationship. It does not explain the meaning of a linear relationship. The sentence is extraneous and does not affect the scoring. The second sentence provides a reasonable definition of a positive relationship, and it is presented in the context of the wolf study. The first sentence of the response to part (a-ii) merely introduces the concept that the response addresses. The second and third sentences describe a positive relationship. This was considered extraneous and not a parallel response. The fourth sentence describes a consequence of a straight line relationship but does not explain what is meant by a linear relationship. This response is not a reasonable definition of a linear relationship, but it does satisfy the context requirement for part (a). The first sentence of the response to part (a-iii) is too vague to provide a reasonable definition of the strength of a relationship. The first part of the second sentence is a reasonable definition of strength because it indicates that the points on the scatterplot are close to a straight line, in this case, the least-squares regression line. The rest of the second sentence enhances the explanation. Because reasonable definitions are given for two of the three components, and at least one is given in context, part (a) was scored as partially correct.

Question 1 (continued)

In part (b) the response correctly interprets the slope as a 35.02 kg increase in y for each 1-meter increase in x. The correct value of the slope is used, and the response clearly indicates that the slope reflects a change in a predicted response. The response is presented in context. Part (b) was scored as essentially correct. In part (c) the response provides a correct formula and value for the predicted weight. It displays a correct formula for the residual and correctly calculates the actual weight of the wolf. Consequently, part (c) was scored as essentially correct. Context is provided in the form of units of length and weight, but context is not needed for an essentially correct response to part (c). Because two parts were scored as essentially correct, and one part was scored as partially correct, the response earned a score of 3.

Sample: 1C
Score: 2

The reference to "an upward-right trend" in the response to part (a-i) is not sufficiently precise to qualify as a reasonable definition of a positive relationship. The inclusion of the graph, however, provides the additional explanation needed for an acceptable response. This is a reasonable definition of a positive relationship, and it satisfies the context requirement for part (a). The response to part (a-ii) does not address the meaning of a linear relationship. Instead it uses a value of a correlation coefficient close to 1 or -1 to describe a strong relationship. Although it is not a reasonable definition of a linear relationship, the response does satisfy the context requirement for part (a). In the response to part (a-iii), indicating that the "data should be near the regression line" provides a reasonable definition of a strong relationship. The response also satisfies the context requirement for part (a). Because reasonable definitions are given for two of the three components, and at least one is given in context, part (a) was scored as partially correct. The response to part (b) correctly identifies the value of the slope, but the interpretation is incorrect because it indicates that the predicted weight will increase by 35.02 kg for "every increase" in length instead of a 1-meter increase in length. Because the response refers to predicted weights and is presented in context, it was scored as partially correct. In part (c) the response provides a correct formula for the predicted weight, but the value for the predicted weight is incorrect. This is an arithmetic error that does not affect the score. A correct formula is presented for computing the actual weight from the residual and the incorrect predicted weight. Part (c) was scored as essentially correct. Because one part was scored as essentially correct, and two parts were scored as partially correct, the response earned a score of 2.