

AP[®] STATISTICS
2007 SCORING GUIDELINES (Form B)

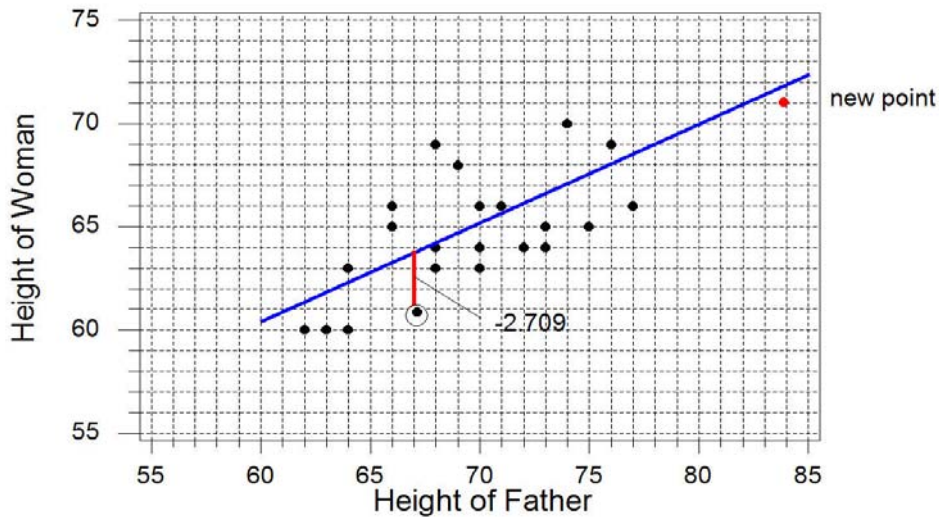
Question 4

Intent of Question

The goals of this question are to assess a student's ability to: (1) plot a least squares regression line; (2) examine a residual; and (3) discuss the effect of an additional observation on an estimated correlation coefficient and on the least squares estimate of the slope of a line.

Solution

Parts (a) and (b):



When $x = 67$, $\hat{y} = 35.1 + 0.427(67) = 63.709$
and the residual $= y - \hat{y} = 61 - 63.709 = -2.709$.

Part (c):

See the new point indicated in the plot above. The slope would remain about the same since the new point is consistent with the linear pattern in the original plot (i.e., close to the line).

The correlation coefficient would increase. We know that $b = r \frac{s_y}{s_x}$. The added point will increase s_x

more than it will increase s_y so $\frac{s_y}{s_x}$ will be less than 1. If the slope is to stay the same, r must increase.

OR

This point fits the pattern well and has an x value that is far from \bar{x} .

AP[®] STATISTICS
2007 SCORING GUIDELINES (Form B)

Question 4 (continued)

Scoring

This problem is scored in 4 sections. Section 1 consists of the graphical parts of (a) and (b) together. Section 2 consists of the numerical parts of (b). Section 3 consists of the first part of (c). Section 4 consists of the second part of (c).

Each section is scored as either essentially correct (E), partially correct (P), or incorrect (I).

Section 1 (graphical parts of a and b) is essentially correct (E) if:

1. the regression line is drawn correctly on the scatterplot;
2. the point (67, 61) is circled and the vertical segment corresponding to the residual is drawn on the scatterplot.

Section 1 is partially correct (P) if the response includes one of the above two elements.

Section 2 (numerical part of b) is essentially correct (E) if the residual is correctly computed as -2.709 ;

OR

the response states that the residual was approximated using the graph, a reasonable value for the residual is given, and the sign of the residual is correct.

Section 2 is partially correct (P) if the magnitude of the residual is correct but the sign is wrong.

Section 3 (first part of (c)) is essentially correct (E) if it:

1. states that the slope will remain about the same (or change slightly);
2. provides an explanation based on the new point fitting the pattern in the original plot.

Section 3 is partially correct (P) if it states that the slope will be about the same, but the explanation is missing or incorrect.

NOTE: If the line is drawn incorrectly in part (a), and the answer to this part is consistent with the line drawn, section 3 is essentially correct (E).

Section 4 (second part of (c)) is essentially correct (E) if it:

1. states that the value of the correlation coefficient will increase;
2. provides an explanation based on the relative changes in s_x and s_y ,

OR

based on the fact that the new point fits the pattern AND is far out in the x direction,

OR

because the linear pattern is stronger.

Section 4 is partially correct (P) if it states that the value of the correlation coefficient will increase, but the explanation is missing or incorrect.

NOTE: If the response just says that the correlation coefficient will increase because the point is close to the line, section 4 is partially correct.

AP[®] STATISTICS
2007 SCORING GUIDELINES (Form B)

Question 4 (continued)

4 Complete Response

All four sections essentially correct

3 Substantial Response

Three sections essentially correct and no sections partially correct

OR

Two sections essentially correct and two sections partially correct

2 Developing Response

Two sections essentially correct and no sections partially correct

OR

One section essentially correct and two sections partially correct

OR

Four parts partially correct

1 Minimal Response

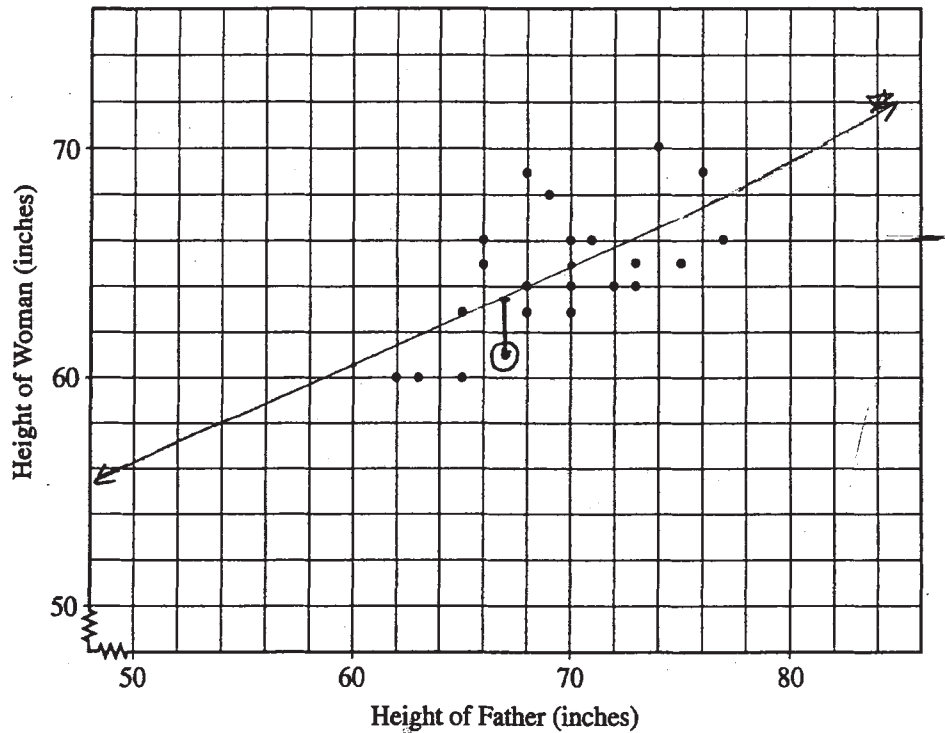
One section essentially correct and no sections partially correct

OR

No sections essentially correct and two sections partially correct

If a response is between two scores (for example, 2½ points), use a holistic approach to determine whether to score up or down depending on the strength of the response and communication.

4. Each of 25 adult women was asked to provide her own height (y), in inches, and the height (x), in inches, of her father. The scatterplot below displays the results. Only 22 of the 25 pairs are distinguishable because some of the (x, y) pairs were the same. The equation of the least squares regression line is $\hat{y} = 35.1 + 0.427x$.



- (a) Draw the least squares regression line on the scatterplot above.

- (b) One father's height was $x = 67$ inches and his daughter's height was $y = 61$ inches. Circle the point on the scatterplot on the preceding page that represents this pair and draw the segment on the scatterplot that corresponds to the residual for it. Give a numerical value for the residual.

$$\text{residual} = y - \hat{y} = 61 - (35.1 + .427(67)) = 61 - 63.709 = -2.709 \text{ inches}$$

- (c) Suppose the point $x = 84$, $y = 71$ is added to the data set. Would the slope of the least squares regression line increase, decrease, or remain about the same? Explain.

(Note: No calculations are necessary to answer this question.)

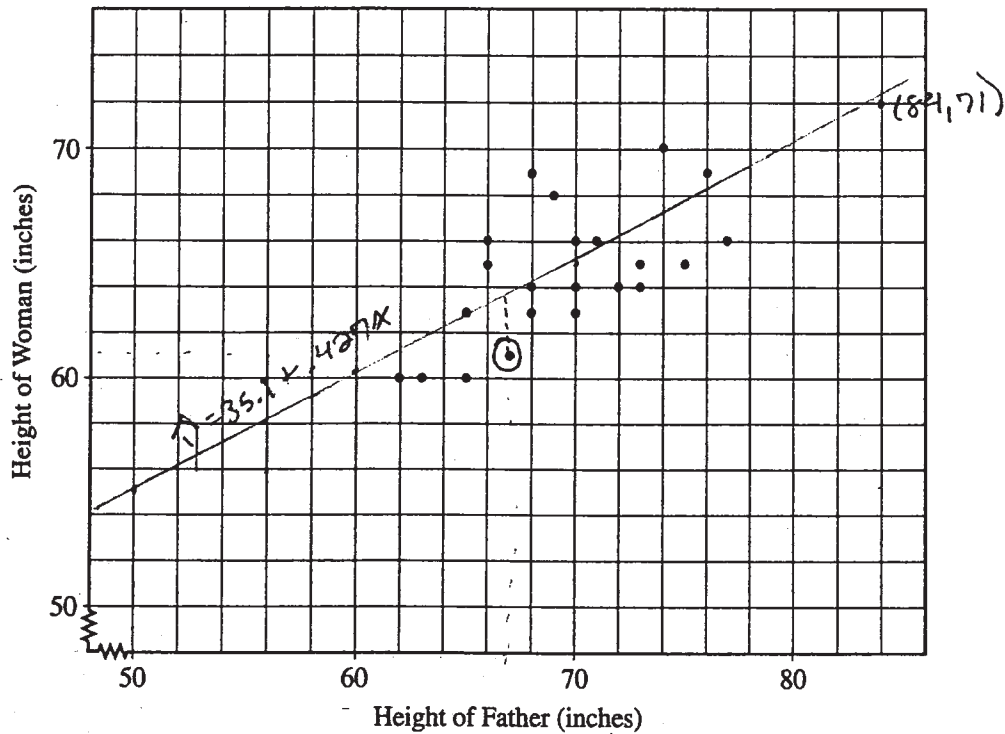
The slope would remain the same because the residual here is very insignificant and the point $(84, 71)$ almost falls on the existing line, thus the line would not have to accommodate it by moving.

Would the correlation increase, decrease, or remain about the same? Explain.

(Note: No calculations are necessary to answer this question.)

The correlation would increase because the point is influential. It increases the "scope" of the least squares regression line by providing a point near the high end of the domain for which the prediction is accurate, meaning that this line can reasonably be applied for a larger range of values. This underlying principle "strengthens" the correlation and makes it closer to 1.

4. Each of 25 adult women was asked to provide her own height (y), in inches, and the height (x), in inches, of her father. The scatterplot below displays the results. Only 22 of the 25 pairs are distinguishable because some of the (x, y) pairs were the same. The equation of the least squares regression line is $\hat{y} = 35.1 + 0.427x$.



- (a) Draw the least squares regression line on the scatterplot above.

Unauthorized copying or reuse of
any part of this page is illegal.

GO ON TO THE NEXT PAGE.

- (b) One father's height was $x = 67$ inches and his daughter's height was $y = 61$ inches. Circle the point on the scatterplot on the preceding page that represents this pair and draw the segment on the scatterplot that corresponds to the residual for it. Give a numerical value for the residual.

$$61 - (63.79) = -2.79$$

- (c) Suppose the point $x = 84$, $y = 71$ is added to the data set. Would the slope of the least squares regression line increase, decrease, or remain about the same? Explain.

(Note: No calculations are necessary to answer this question.)

The slope of the least squares regression line would remain about the same because the point $(84, 71)$ is very close to the least squares regression line.

Would the correlation increase, decrease, or remain about the same? Explain.

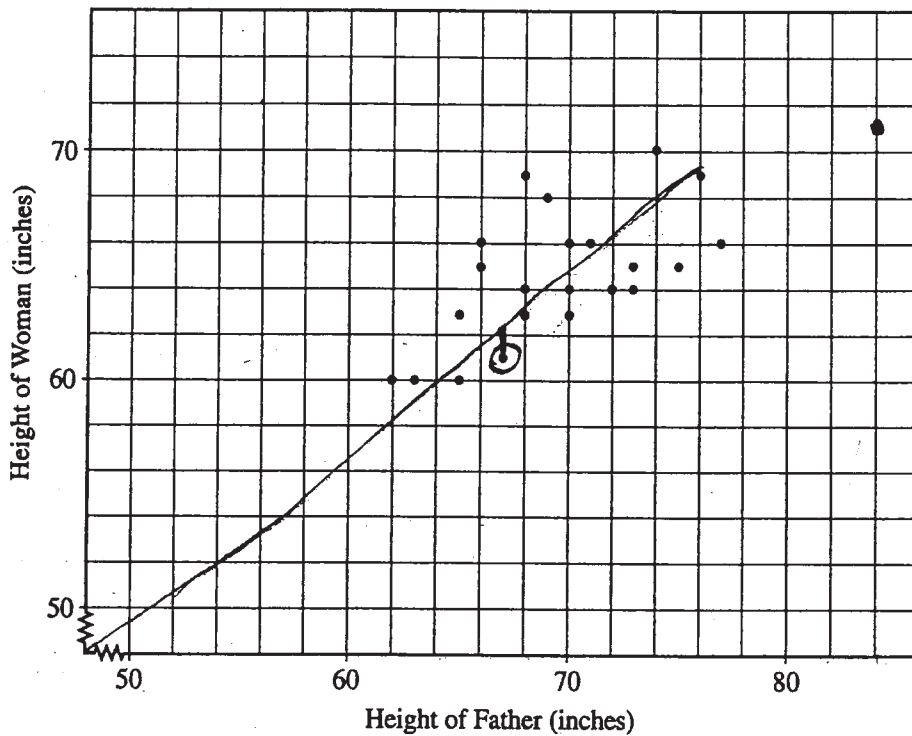
(Note: No calculations are necessary to answer this question.)

The correlation would also remain about the same because correlation & the slope of the least squares regression line are exactly the same.

$y = \text{hers}$
 $x = \text{fathers}$

4C1

4. Each of 25 adult women was asked to provide her own height (y), in inches, and the height (x), in inches, of her father. The scatterplot below displays the results. Only 22 of the 25 pairs are distinguishable because some of the (x, y) pairs were the same. The equation of the least squares regression line is $\hat{y} = 35.1 + 0.427x$.



- (a) Draw the least squares regression line on the scatterplot above.

Unauthorized copying or reuse of
any part of this page is illegal.

GO ON TO THE NEXT PAGE.

(67, 61)

- (b) One father's height was $x = 67$ inches and his daughter's height was $y = 61$ inches. Circle the point on the scatterplot on the preceding page that represents this pair and draw the segment on the scatterplot that corresponds to the residual for it. Give a numerical value for the residual.

$$\text{residual} = \text{Actual} - \text{Predicted}$$

$$\text{residual} \approx 1$$

(84, 71)

- (c) Suppose the point $x = 84$, $y = 71$ is added to the data set. Would the slope of the least squares regression line increase, decrease, or remain about the same? Explain.

(Note: No calculations are necessary to answer this question.)

This point may be influential, which would change the slope, however this point seems to follow the increasing linear pattern, so I don't think the slope will change too much.

- Y Would the correlation increase, decrease, or remain about the same? Explain.

(Note: No calculations are necessary to answer this question.)

Correlation will increase because r is non-resistant to extreme values. However this point may not be considered an extreme value to some.

AP[®] STATISTICS
2007 SCORING COMMENTARY (Form B)

Question 4

Sample: 4A

Score: 4

This is a complete response that accurately presents the least squares estimate of the regression line on the plot, circles the point corresponding to a woman of height 61 inches who has a father of height 67 inches, and inserts a vertical line segment corresponding to the residual for that point. The residual is correctly calculated in part (b). The response to part (c) correctly explains that the additional point would have little effect on the least squares estimate of the regression line because it is close to the existing regression line. The response to part (d) provides a reasonable support for the conclusion that the additional point would increase the estimated correlation.

Sample: 4B

Score: 3

This is a substantial response that accurately presents the least squares estimate of the regression line on the plot, circles the point corresponding to a woman of height 61 inches who has a father of height 67 inches, and inserts a vertical line segment corresponding to the residual for that point. The residual is computed correctly in part (b). The response to part (c) explains that the additional point would have little effect on the least squares estimate of the regression line because it is close to the existing regression line. The response to part (d) incorrectly argues that the correlation would stay about the same because the additional point would have little effect on the slope of the regression line.

Sample: 4C

Score: 2

This is a developing response that reflects less than complete understanding in several parts. The least squares estimate of the regression line is not presented accurately on the plot. The point corresponding to a woman of height 61 inches who has a father of height 67 inches is circled, and a vertical line segment is used to indicate the residual for that point. The magnitude of the residual reported in part (b) is consistent with the line displayed in part (a), but the numerical value should be negative. The response to part (c) begins with a general statement, but it eventually indicates that the additional point would have little effect on slope of the least squares estimate of the regression line because it is close to the existing regression line. The response to part (d) indicates that the correlation would increase, but the explanation is poor and somewhat contradictory.