

**Unit 2 Progress Check: FRQ**

1. 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.

The following table shows the annual income, in dollars, and amount spent on vacation, in dollars, for a sample of 8 families.

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

- (a) Create a scatterplot of the data in the table.
- (b) Describe the association shown in the scatterplot created in part (a).
- (c) Calculate the coefficient of determination for the data, and interpret the value in context.

**Part A, B, and C**

The primary goals of this question are to assess a student's ability to (1) appropriately graph a relationship between two quantitative variables; (2) describe the relationship shown in a scatterplot; and (3) calculate and interpret the coefficient of determination.

## Scoring

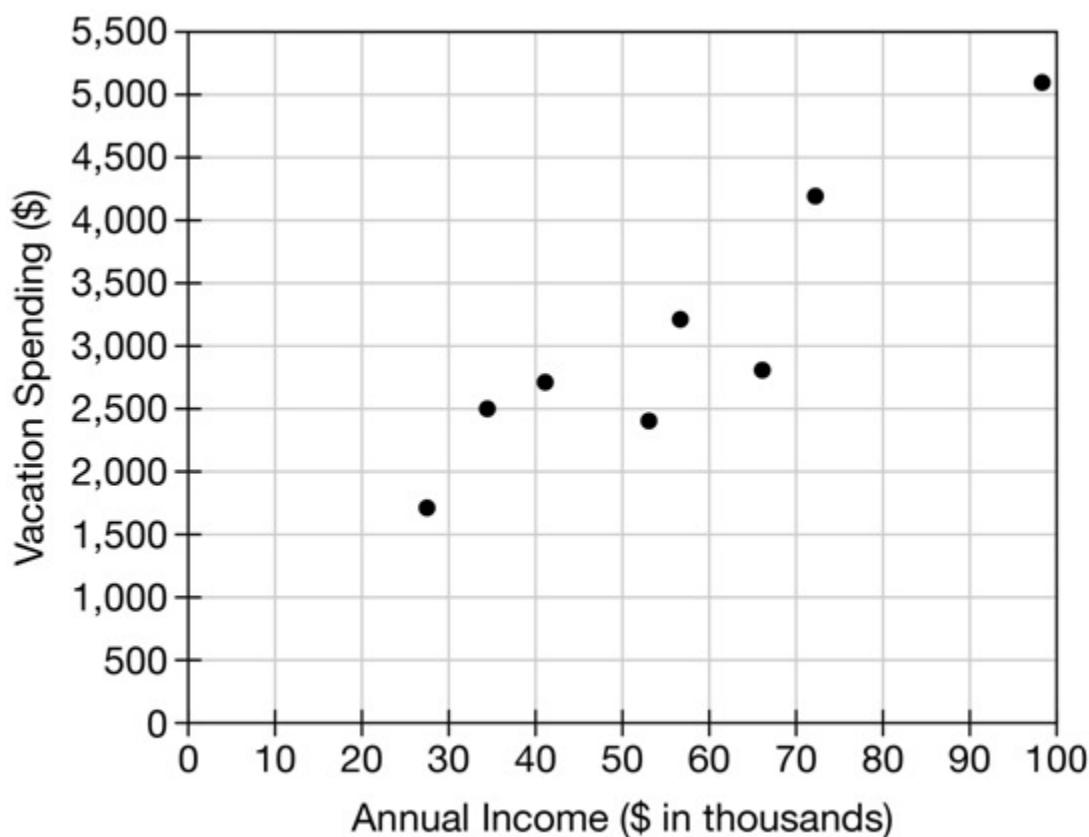
Parts (a), (b), and (c) are each scored as essentially correct (E), partially correct (P), or incorrect (I).



0	1	2	3	4
---	---	---	---	---

All three parts essentially correct

- ☐ Part (a) essentially correct
- ☐ Part (a) partially correct
- ☐ Part (a) incorrect
- ☐ Part (b) essentially correct
- ☐ Part (b) partially correct
- ☐ Part (b) incorrect
- ☐ Part (c) essentially correct
- ☐ Part (c) partially correct

**Unit 2 Progress Check: FRQ**☐ Part (c) incorrectSolution**Part (a):**

## Scoring

**Part (a)** is scored as follows.

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

- The scatterplot contains 8 points AND their locations are reasonably accurate.
- The response includes a correct horizontal axis label (e.g., Annual Income).
- The response includes a correct vertical axis label (e.g., Vacation Spending).
- The response includes at least 3 numbers labeled on the horizontal axis.
- The response includes at least 3 numbers labeled on the vertical axis.

Partially correct (P) if the response includes three or four of the five components

## Unit 2 Progress Check: FRQ

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

Note: There are many different appropriate scales that can be used to represent the data in the scatterplot. The model solution only represents one such example.

**Part (b):** The graph indicates that there is a fairly strong, linear, positive relationship between annual income and vacation spending, with no apparent outliers.

**Part (b)** is scored as follows.

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

- a strong or fairly strong relationship,
- linear,
- positive
- and no outliers.

Partially correct (P) if the response contains two or three of the four components

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

**Part (c):**

The coefficient of determination is  $r^2 = 0.85$ .

The coefficient of determination indicates 85% of the variation in amount spent on vacation can be explained by annual income.

**Part (c)** is scored as follows.

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

- a coefficient of determination of 0.85,
- provides a correct (possibly generic) interpretation of  $r^2$ ,
- interpretation includes context.

Partially correct (P) if the response does one of the following:

- includes the coefficient of determination of 0.85 AND only one of the other two components
- satisfied the three components, but reverses the roles of annual income and amount spent on vacation

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

Notes:

· If a response indicates that the relationship is linear because the value of the coefficient of determination is close to 1, the score should be lowered one level (that is, from an E to a P, or a P to an I).

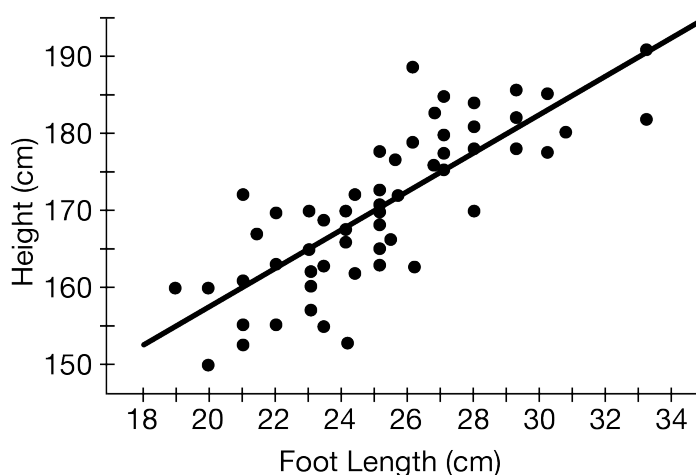
## Unit 2 Progress Check: FRQ

- In component 2, common *incorrect* interpretations of the coefficient of determination include:
  - o The percent variability in the *predicted*  $y$  values that is explained by the linear relationship between  $y$  and  $x$ .
  - o The percent variability in the *data* that is explained by the linear relationship between  $y$  and  $x$ .
  - o The percent variability that is explained by the linear relationship between  $y$  and  $x$ .
  - o The percent variability in  $y$  that is *on average* explained by the linear relationship between  $y$  and  $x$ .

## Unit 2 Progress Check: FRQ

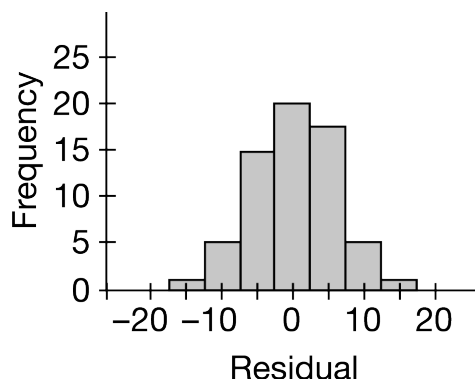
2. Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

A random sample of 65 high school seniors was selected from all high school seniors at a certain high school. The following scatterplot shows the height, in centimeters (cm), and the foot length, in cm, for each high school senior from the sample. The least-squares regression line is shown. The computer output from the least-squares regression analysis is also shown.



Term	Coef	(SE) Coef	T-Value	P-Value
Constant	105.08	6.00	17.51	0.000
Foot length	2.599	0.238	10.92	0.000
S = 5.90181      R-sq = 65.42%				

- (a) Calculate and interpret the residual for the high school senior with a foot length of 20 cm and a height of 160 cm.
- (b) The standard deviation of the residuals is  $s = 5.9$ . Interpret the value in context.
- (c) The following histogram summarizes the 65 residuals.

**Unit 2 Progress Check: FRQ**

Assume that the distribution of residuals is approximately normal with mean 0 cm and standard deviation 5.9 cm. What percent of the residuals are greater than 8 cm? Justify your answer.

(d) Based on your answer to part (c), would it be surprising to randomly select a high school senior from the high school with a foot length of 20 cm and a height greater than 165 cm? Justify your answer.

**Part A, B, C, and D****Intent of Question**

The primary goals of this question are to assess a student's ability to (1) calculate and interpret a residual; (2) interpret the standard deviation of the residuals; (3) calculate a percentage using a normal distribution; and (4) combine information from previous parts to make a conclusion about the likelihood of an event.

Each essentially correct (E) part counts as 1 point.

Each partially correct (P) part counts as  $\frac{1}{2}$  point.

**Scoring**

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

If a response is between two scores (for example,  $2\frac{1}{2}$  points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.

*Reasons to score up:*

- All notation is correct and clearly marked
- All explanations are clear
- No wrong information is included that was not part of the scoring (for example, saying sample size must be greater than 30 when that has nothing to do with the problem)
- No minor calculation errors are made, if they are not part of the scoring
- Interpretation parts are especially strong

## Unit 2 Progress Check: FRQ

Reasons to score down:

- Notation is not wrong, but is spotty and not clearly marked
- Explanations are not wrong, but are hard to follow
- Wrong or extraneous information is included but not part of scoring
- Minor calculation errors that are not part of the scoring are made

Interpretation parts are scored an E but are considered a weak E



0	1	2	3	4
---	---	---	---	---

Parts (a) through (d) sum to 4 points.

OR

Parts (a) through (d) sum to  $3\frac{1}{2}$  points AND a holistic approach is used to decide to score up.

- ☐ Part (a) essentially correct
- ☐ Part (a) partially correct
- ☐ Part (a) incorrect
- ☐ Part (b) essentially correct
- ☐ Part (b) partially correct
- ☐ Part (b) incorrect
- ☐ Part (c) essentially correct
- ☐ Part (c) partially correct
- ☐ Part (c) incorrect
- ☐ Part (d) essentially correct
- ☐ Part (d) partially correct
- ☐ Part (d) incorrect

## Solution

**Part (a):**

From the computer output, the regression equation is  $\hat{y} = 105.08 + 2.599x$  where  $x$  represents the foot length, in cm and  $y$  represents the height, in cm. For  $x = 20$ ,  $\hat{y} = 105.08 + 2.599(20) = 157.06$ . Therefore, the residual is equal to  $160 - 157.06 = 2.94$  cm (that is, the observed value minus the predicted value). The high school senior is 2.94 cm taller than expected based on his or her foot length.

## Unit 2 Progress Check: FRQ

### Scoring

**Part (a)** is scored as follows.

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

The response states that the predicted value is 157.06 cm.

- There is evidence that the response subtracts the predicted value from the actual value to calculate the residual.
- The interpretation contrasts the actual height with the predicted (expected) height, in context.
- The interpretation includes correct direction (e.g., taller, greater) OR includes a direction consistent with an incorrect calculation.

Partially correct (P) if the response includes two or three of the four components needed for an E

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

Note: Estimating the value of the residual from the graph does not satisfy components 1 or 2.

## Solution

**Part (b):** Actual heights for the high school seniors typically vary by about 5.9 cm from the heights predicted by the least-squares regression line using  $x$  = foot length.

### Scoring

**Part (b)** is scored as follows.

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

- The interpretation contrasts the actual  $y$ -values with the predicted  $y$ -values OR contrasts the points in the scatterplot with the least-squares regression line.
- The interpretation includes that idea that the standard deviation measures a typical or average distance.
- The interpretation uses the variable name (height).

Partially correct (P) if the response includes two of the three components needed for an E

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

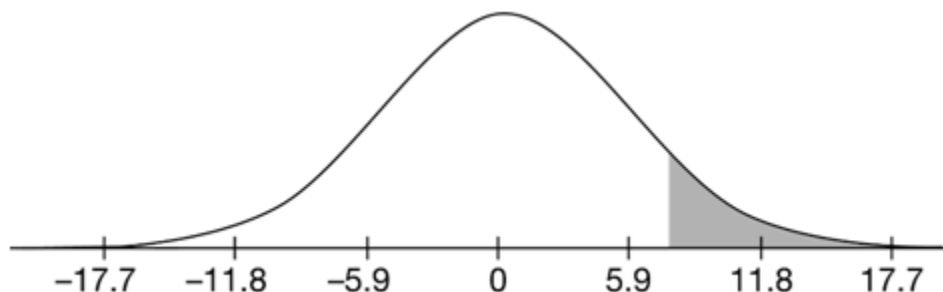
Note: Including the name of the  $x$  variable (foot length) is not required, but it is considered a positive in terms of holistic scoring.

### Solution

**Part (c):**



## Unit 2 Progress Check: FRQ



The  $z$ -score associated with a residual of 8 cm is  $z = \frac{8-0}{5.9} = 1.36$ .

Using technology, or the standard normal table, the area under the normal curve to the right of 8 cm is 0.0869. About 9% of residuals are greater than 8 cm.

Scoring

**Part (c)** is scored as follows.

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

- The response indicates use of a normal distribution with mean equal to 0 and standard deviation equal to 5.9.
- The response indicates that the values of interest are greater than 8.
- The response provides an answer consistent with components 1 and 2.

Partially correct (P) if the response includes two of the three components needed for an E.

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

Notes:

- To satisfy component 1, a response should include a well-labeled sketch (with mean and  $\pm 1$  standard deviation labeled) OR a calculator command that includes the word normal and has the mean and standard deviation labeled OR a  $z$ -score calculation along with an indication of normality (by using the word normal or drawing a sketch of a normal curve).
- To satisfy component 2, a response can include a well-labeled sketch with a boundary line at 8 and shading to the right OR a calculator command that specifies a lower bound of 8 and an upper bound at least 5 standard deviations greater than the mean OR an inequality statement such as  $P(X > 8)$  or  $P(z > 1.36)$ .

Solution

**Part (d):** No. A high school senior with a foot length of 20 cm is expected to be 157.06 cm tall (calculated in part a), so we'd expect about 9% (calculated in part c) of all high school seniors from the high school with a foot length of 20 cm to be at least  $157.06 + 8 = 165.06$  cm tall.

Scoring

**Part (d)** is scored as follows.

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

## Unit 2 Progress Check: FRQ

- The response recognizes that a height of 165 cm is about 8 cm greater than the predicted height for a high school senior with a foot length of 20 cm.
- The response uses the value from part (c) to determine what percent of high school seniors with a foot length of 20 cm will have a height greater than 165 cm.
- The response states that selecting such a high school senior at random would not be surprising.

Partially correct (P) if the response includes two of the three components needed for an E.

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

Note: If the response suggests that the percent from part (c) is small (e.g., “only 9%”), then component 3 can be satisfied if the response also states that selecting such a high school senior would be surprising.