AP® STATISTICS
2014 SCORING GUIDELINES

Question 6

Intent of Question

The primary goals of this question were to assess a student's ability to (1) calculate and interpret a residual value; (2) answer questions about residual plots; (3) compare associations between two scatterplots; and (4) identify an appropriate explanatory variable to include in a regression model based on residuals from simpler regression models.

Solution

Part (a):

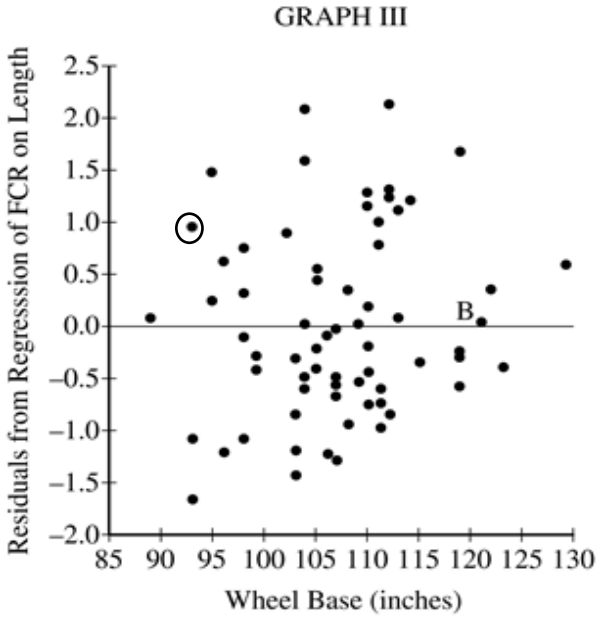
For a car with length 175 inches, the predicted value for the car's FCR, based on the least squares regression line, is

$$\text{predicted FCR} = -1.595789 + 0.0372614(175) \approx 4.92 \text{ gallons per 100 miles.}$$

The actual FCR for the car is 5.88, so the residual is $5.88 - 4.92 = 0.96$. The residual value means that the car's FCR is 0.96 gallons per 100 miles greater than would be predicted for a car of its length.

Part (b):

(i) The point with a wheel base of 93 inches and a residual of 0.96 gallons per 100 miles is circled in graph III below.



(ii) Point B corresponds to a car with an actual FCR that is very close to the FCR that would be predicted for a car with its length by the regression model which predicts FCR using the explanatory variable length.

AP® STATISTICS
2014 SCORING GUIDELINES

Question 6 (continued)

Part (c):

Graph II reveals a moderate association that is positive and linear. In contrast, there is a weak association that is positive and linear in graph III. The association between engine size and residual (from predicting FCR based on length) is stronger than the association between wheel base and residual (from predicting FCR based on length).

Part (d):

Engine size is a better choice than wheel base for including with length in a regression model for predicting FCR. The stronger association between engine size and residual (from predicting FCR based on length) indicates that engine size is more useful than wheel base for reducing the variability in FCR values that remains unexplained (as indicated by residuals) after predicting FCR based on length.

Scoring

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

Part (a) is scored as follows:

Essentially correct (E) if the response provides the following two components:
1. A correct residual value with supporting calculation.
2. A correct interpretation of the residual value, in context.

Partially correct (P) if the response includes only one of the two components listed above.

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

Notes:
- If the residual value is incorrect, the interpretation should be considered correct if it follows from the incorrect residual value.
- Correct interpretation of the residual must include the correct direction and magnitude of the FCR value away from the predicted FCR value.
- A calculated residual value which is slightly different from 0.96 due to the number of significant digits is acceptable.

Part (b) is scored as follows:

Essentially correct (E) if the response provides the following two components:
1. Circles the correct point in graph III.
2. Provides a reasonable interpretation of the car associated with point B having a residual near 0 that refers to predicting FCR based on length.

Partially correct (P) if the response correctly provides only one of the two components listed above.

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

Question 6 (continued)

Note: A correct response for the second component must include reference to the observed FCR value of the car represented by point B, not the point B itself.

Part (c) is scored as follows:

Essentially correct (E) if the response correctly provides the following three components:
1. A description of form *AND* direction for both graphs.
2. A description of the strength of association for both graphs.
3. A comparison between the two graphs.

Partially correct (P) if the response correctly provides only two of the three components listed above.

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

Notes:
- Part (c) is focused on the comparison of graph II and graph III. Inferences drawn from patterns in these graphs are considered in part (d).
- Linear is needed for form in graph II.
- Graph III may be described as having no association between wheel base and the residuals of FCR based on length, which is sufficient for describing the form, direction and strength of association of graph III.

Part (d) is scored as follows:

Essentially correct (E) if the response indicates the correct choice with a sound justification based on the following two components:
1. The strong(er) association.
2. Reducing the variability that remains unexplained in the model which predicts FCR based on length.

Partially correct (P) if the response indicates the correct choice and provides a justification based on only one of the two components which are listed above.

Incorrect (I) if the response indicates the incorrect choice;
OR
if the response indicates the correct choice but does not mention either of the two components which are listed above.

Note: Describing the variables in graph II and graph III as residuals is not required but can be used positively in holistic scoring. Incorrect descriptions of graph II or graph III or the variables in graphs are not acceptable.

AP® STATISTICS
2014 SCORING GUIDELINES

Question 6 (continued)

Each essentially correct (E) part counts as 1 point. Each partially correct (P) part counts as ½ point.

4 Complete Response

3 Substantial Response

2 Developing Response

1 Minimal Response

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

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.

Linear Fit
FCR = −1.595789 + 0.0372614 * Length

Summary of Fit	
RSquare	0.250401
Root Mean Square Error	0.902382
Observations	66

(a) The point on 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.

$\hat{FCR} = -1.596 + .0373(\text{length (in.)})$

$= -1.596 + .0373(175)$

$= 4.925$ predicted FCR

residual = actual − predicted

$= 5.88 - 4.925$

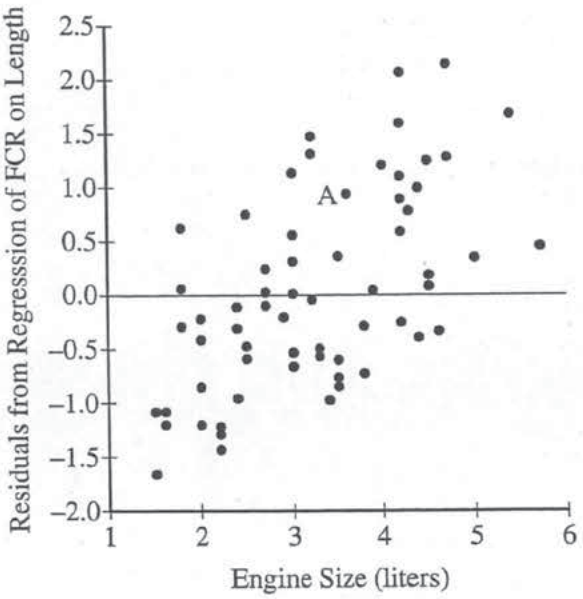
$= .955$

A residual of .955 shows an underestimate by the least squares regression line. This residual shows that the predicted FCR is .955 gallons/100 miles lower than car A's actual consumption rate.

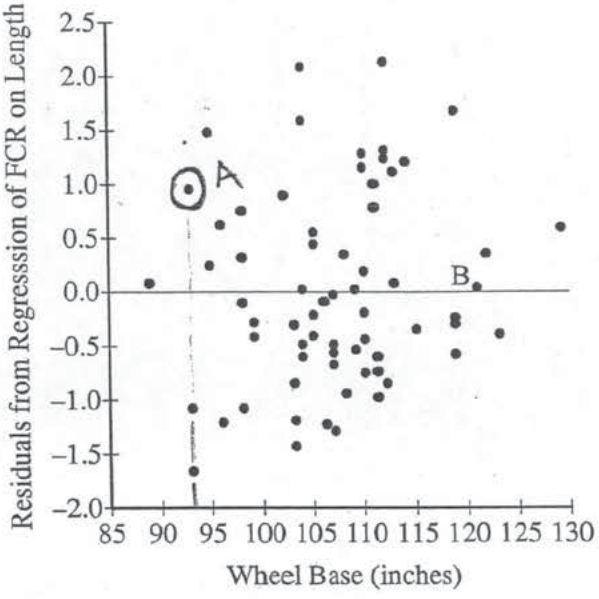
6A2

Jamal knows that it is possible to predict a response variable using more than one explanatory variable. He wants to see if he can improve the original model of predicting FCR from length by including a second explanatory variable in addition to length. He is considering including engine size, in liters, or wheel base (the length between axles), in inches. Graph II is a scatterplot showing the engine size of the 66 cars plotted with the corresponding residuals from the regression of FCR on length. Graph III is a scatterplot showing the wheel base of the 66 cars plotted with the corresponding residuals from the regression of FCR on length.

GRAPH II



GRAPH III



(b) In graph II, the point labeled A corresponds to the same car whose point was labeled A in graph I. The measurements for the car represented by point A are given below.

FCR	Length (inches)	Engine Size (liters)	Wheel Base (inches)
5.88	175	3.6	93

(i) Circle the point on graph III that corresponds to the car represented by point A on graphs I and II.

(93, .955) resid from part A

(ii) There is a point on graph III labeled B. It is very close to the horizontal line at 0. What does that indicate about the FCR of the car represented by point B?

This indicates that the predicted FCR for that particular car when using the least squares regression based on the car's length was very accurate, as it resulted in a very small residual. In other words, the prediction made using Jamal's initial least squares regression line was very close to the car's true FCR.

GO ON TO THE NEXT PAGE.

(c) Write a few sentences to compare the association between the variables in graph II with the association between the variables in graph III.

The association between engine size and residuals in Graph II shows a positive, roughly linear association, with weak to moderate strength. Graph III, between wheel base and residual shows no apparent pattern. Both graphs contain no obvious outliars. Graph III appears to have a larger scatter that Graph II.

(d) Jamal wants to predict FCR using length and one of the other variables, engine size or wheel base. Based on your response to part (c), which variable, engine size or wheel base, should Jamal use in addition to length if he wants to improve the prediction? Explain why you chose that variable.

Jamal should choose to use engine size in addition to length to improve his prediction. The scatter in Graph II, between engine size and residuals from FCR on length, appears to have a stronger association than Graph III. This means that more of the variation in the residuals will be able to be accounted for if engine size is added than if wheel base were to be added. The more variation that can be accounted for, the better Jamal will be able to make predictions.

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.

Linear Fit
FCR = −1.595789 + 0.0372614 * Length

Summary of Fit	
RSquare	0.250401
Root Mean Square Error	0.902382
Observations	66

(a) The point on 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.

FCR(175 in) = −1.595789 + .0372614(175 in)

FCR = 4.925

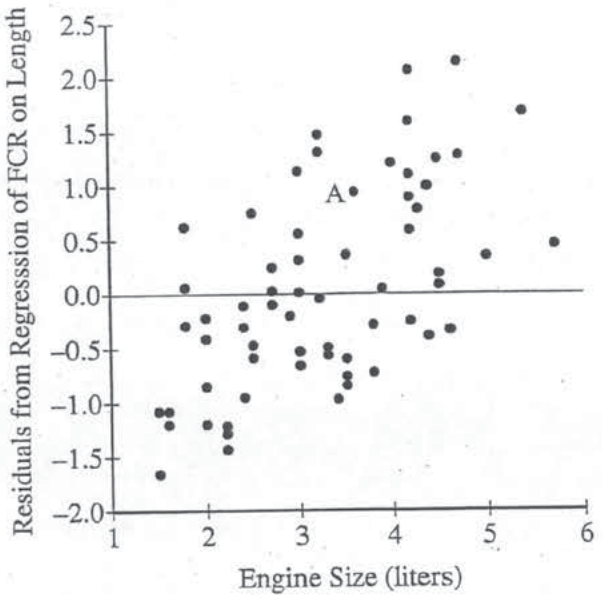
Residual = 5.88 − 4.925 = .955

A residual of .955 means that, for a car length of 175 inches, the observed FCR is .955 greater than the FCR predicted by the linear model.

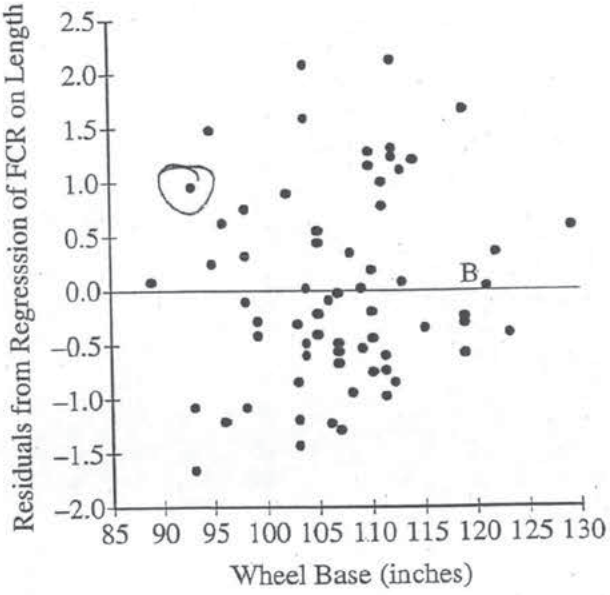
6B2

Jamal knows that it is possible to predict a response variable using more than one explanatory variable. He wants to see if he can improve the original model of predicting FCR from length by including a second explanatory variable in addition to length. He is considering including engine size, in liters, or wheel base (the length between axles), in inches. Graph II is a scatterplot showing the engine size of the 66 cars plotted with the corresponding residuals from the regression of FCR on length. Graph III is a scatterplot showing the wheel base of the 66 cars plotted with the corresponding residuals from the regression of FCR on length.

GRAPH II



GRAPH III



(b) In graph II, the point labeled A corresponds to the same car whose point was labeled A in graph I. The measurements for the car represented by point A are given below.

FCR	Length (inches)	Engine Size (liters)	Wheel Base (inches)
5.88	175	3.6	93

(i) Circle the point on graph III that corresponds to the car represented by point A on graphs I and II.

(ii) There is a point on graph III labeled B. It is very close to the horizontal line at 0. What does that indicate about the FCR of the car represented by point B?

Since the point is very close to the horizontal line at 0, it must have a very small residual, meaning that the observed FCR is very close to the FCR predicted by the linear model of FCR vs. length.

GO ON TO THE NEXT PAGE.

(c) Write a few sentences to compare the association between the variables in graph II with the association between the variables in graph III.

Graph II has a linear shape whereas Graph III has no clear shape.

Graph II has a positive direction compared to Graph III which has no clear direction

Graph II has a greater strength (it's positive linear association is more evident) than Graph III. Thus, there appears to be a relatively strong positive linear association between the variables in Graph II while there is no clear association between the variables of graph III.

(d) Jamal wants to predict FCR using length and one of the other variables, engine size or wheel base. Based on your response to part (c), which variable, engine size or wheel base, should Jamal use in addition to length if he wants to improve the prediction? Explain why you chose that variable.

Jamal should choose wheel base in addition to length to improve his prediction because there is no clear relationship between the residuals of the regression of FCR on length and wheel base (in). The graph is uniformly scattered, meaning that the linear model is a better representation of the data.

In Graph II, the fairly strong positive linear relationship between residuals and engine size indicates that for large and small engine sizes, the linear model becomes less accurate, so Jamal should not use engine size.

6C1

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.

Linear Fit
FCR = −1.595789 + 0.0372614 * Length

Summary of Fit	
RSquare	0.250401
Root Mean Square Error	0.902382
Observations	66

(a) The point on 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.

$\hat{p} = -1.595789 + .0372614(175)$

$\hat{p} = 4.924956$

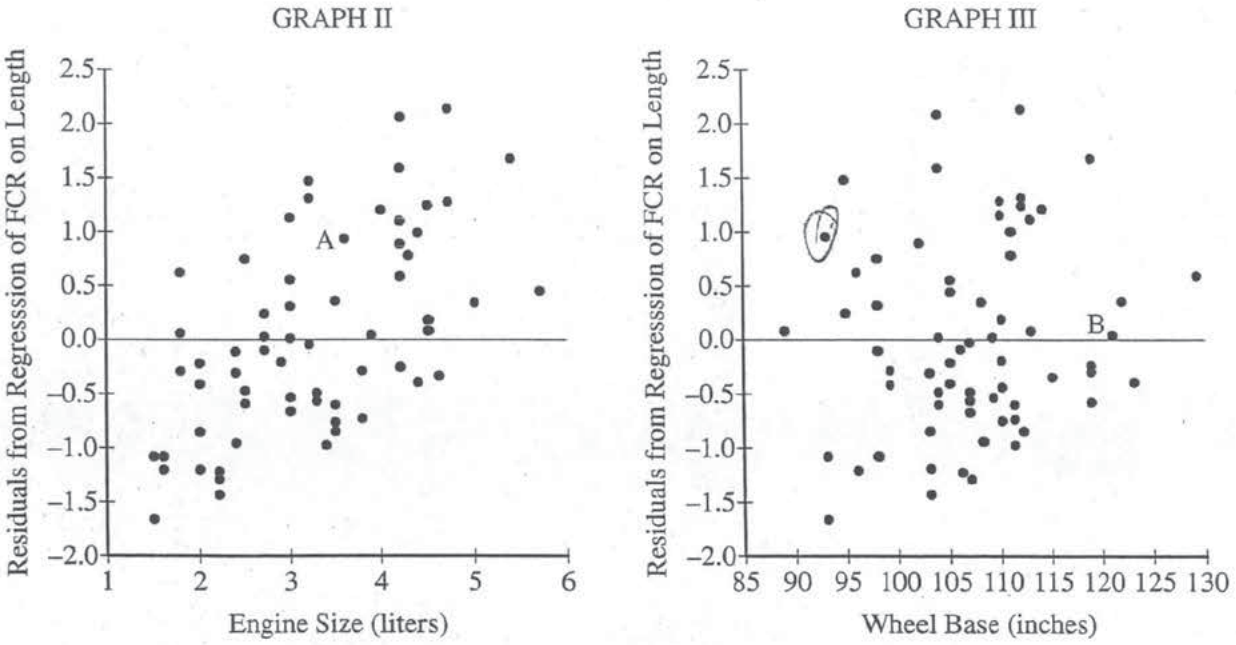
observed − expected = 5.88 − 4.925

= .955

The observed value of 5.88 FCR is .955 away from the expected value of 4.925 FCR that we get using the least squares regression line

GO ON TO THE NEXT PAGE.

Jamal knows that it is possible to predict a response variable using more than one explanatory variable. He wants to see if he can improve the original model of predicting FCR from length by including a second explanatory variable in addition to length. He is considering including engine size, in liters, or wheel base (the length between axles), in inches. Graph II is a scatterplot showing the engine size of the 66 cars plotted with the corresponding residuals from the regression of FCR on length. Graph III is a scatterplot showing the wheel base of the 66 cars plotted with the corresponding residuals from the regression of FCR on length.



(b) In graph II, the point labeled A corresponds to the same car whose point was labeled A in graph I. The measurements for the car represented by point A are given below.

FCR	Length (inches)	Engine Size (liters)	Wheel Base (inches)
5.88	175	3.6	93

(i) Circle the point on graph III that corresponds to the car represented by point A on graphs I and II.

(ii) There is a point on graph III labeled B. It is very close to the horizontal line at 0. What does that indicate about the FCR of the car represented by point B?

It indicates that the observed value of FCR was almost the same as the predicted value of FCR

6C3

(c) Write a few sentences to compare the association between the variables in graph II with the association between the variables in graph III.

In graph II there seems to be a mild, positive association between engine size and residuals of FCR. while in graph III there seems to be no association between wheel base and the residuals of FCR

(d) Jamal wants to predict FCR using length and one of the other variables, engine size or wheel base. Based on your response to part (c), which variable, engine size or wheel base, should Jamal use in addition to length if he wants to improve the prediction? Explain why you chose that variable.

I would use engine size in addition to length to improve the prediction because it has an association to FCR which would make the predicted be closer to the observed, while wheel base has no effect on the FCR.

AP® STATISTICS
2014 SCORING COMMENTARY

Question 6

Overview

The primary goals of this question were to assess a student's ability to (1) calculate and interpret a residual value; (2) answer questions about residual plots; (3) compare associations between two scatterplots; and (4) identify an appropriate explanatory variable to include in a regression model based on residuals from simpler regression models.

Sample: 6A
Score: 4

In part (a) the residual is calculated correctly as 0.955, and it is stated that a residual of 0.955 shows that the predicted FCR is 0.955 gallons per 100 miles lower than car A's actual consumption rate. The response includes supporting calculations for the residual, and a correct interpretation of the residual value of 0.955 in context. Part (a) was scored as essentially correct. In part (b) the correct point was circled and labeled "A" on graph III, satisfying the first component. It is reported that the predicted FCR for the car corresponding to point B was very accurately predicted by the least squares regression based on the car's length. The response also states that the prediction made using Jamal's initial least squares regression line was very close to the car's true FCR, and the second component is satisfied. Part (b) was scored as essentially correct. In part (c) the association between engine size and residuals in graph II is described as positive, roughly linear with weak to moderate strength. No apparent pattern is reported for the association between engine size and residuals for graph III. Graph III is indicated to have a larger scatter than graph II. The stronger association of engine size and residuals than wheel base and residuals is specifically stated and used in part (d) in the choice of engine size. Thus, there is a description of form, direction and strength of association for both graphs and a comparison of strength of association. Part (c) was scored as essentially correct. In part (d) the correct choice of engine size is made. The choice is justified by both the stronger association in graph II than in graph III and by a greater reduction in the variation of the residuals when engine size is added to the model. Part (d) was scored as essentially correct. Because all four parts were scored as essentially correct, the response earned a score of 4.

Sample: 6B
Score: 3

In part (a) the residual is correctly calculated as 0.955, and it is stated that a residual of 0.955 means that, for a car of 175 inches, the observed FCR is 0.955 greater than the FCR predicted by the linear model. The response includes supporting calculations for the residual and a correct interpretation of the residual value of 0.955 in context. Part (a) was scored as essentially correct. In part (b) the correct point was circled on graph III, satisfying the first component. It is stated that the very small residual implies that the observed FCR is very close to the FCR predicted by the linear model of FCR versus length, and the second component is satisfied. Part (b) was scored as essentially correct. In part (c) the two graphs are compared on form "Graph II has a linear shape whereas Graph III has no clear shape", direction "Graph II has a positive direction compared to Graph III which has no clear direction", and strength of association with graph II having the stronger association. Thus, part (c) was scored as essentially correct. In part (d) the incorrect choice of wheel base is made. The choice of wheel base resulted from an incorrect interpretation of the residuals from a regression of FCR with the explanatory variable engine size in graph II and an incorrect interpretation of the residuals from a regression of FCR with the explanatory variable wheel base in graph III. Part (d) was scored as incorrect. Because three parts were scored as essentially correct and one part was scored as incorrect, the response earned a score of 3.

Question 6 (continued)

Sample: 6C
Score: 2

In part (a) the residual is correctly calculated as 0.955, and it is stated that the observed value of 5.88 FCR is 0.955 away from the expected value of 4.925 FCR that is obtained from the least squares line for this car. The response includes supporting calculations for the residual and an interpretation of the residual value of 0.955 in context, but is missing direction of the amount away from the least squares line. Part (a) was scored as partially correct. In part (b) the correct point was circled on graph III, satisfying the first component. The response minimally refers to the observed value of FCR as almost the same as the predicted value of FCR, and the second component is satisfied. Thus, part (b) was scored as essentially correct. In part (c) the association between engine size and residuals of FCR is described as mild and positive while no association is reported between wheel base and the residuals of FCR. Describing the residuals as the residuals from regression of FCR on length is more accurate, but the response contains a description of only direction and strength of association for both graphs (form of association is not described in the response). A comparison of stronger association for engine size than no association for wheel base is included. Thus, part (c) was scored as partially correct. In part (d) the correct choice of engine size is made. The choice is justified by an association of engine size to FCR instead of the stronger association of engine size to the residuals from the regression of FCR based on length which are the correct variables in graph II and graph III. Thus, part (d) was scored as incorrect. Because one part was scored as essentially correct, two parts were scored as partially correct, and one part was scored as incorrect, the response earned a score of 2.