STATISTICS

SECTION II

Part B

Question 6

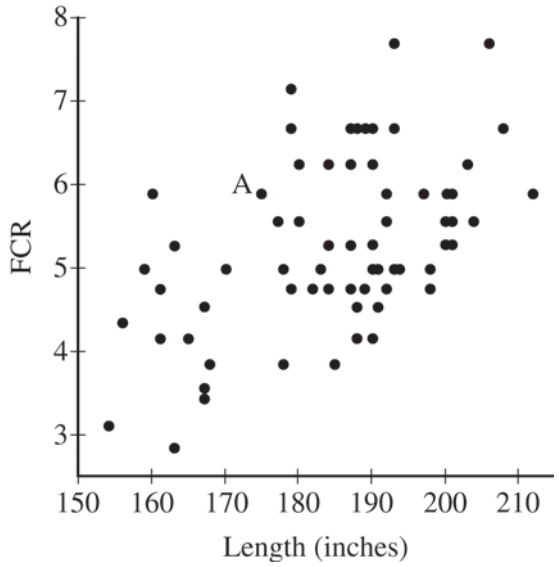
Spend about 25 minutes on this part of the exam.

Percent of Section II score—25

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.

6. Jamal is researching the characteristics of a car that might be useful in predicting the fuel consumption rate (FCR); that is, the number of gallons of gasoline that the car requires to travel 100 miles under conditions of typical city driving. The length of a car is one explanatory variable that can be used to predict FCR. Graph I is a scatterplot showing the lengths of 66 cars plotted with the corresponding FCR. One point on the graph is labeled A.

GRAPH I



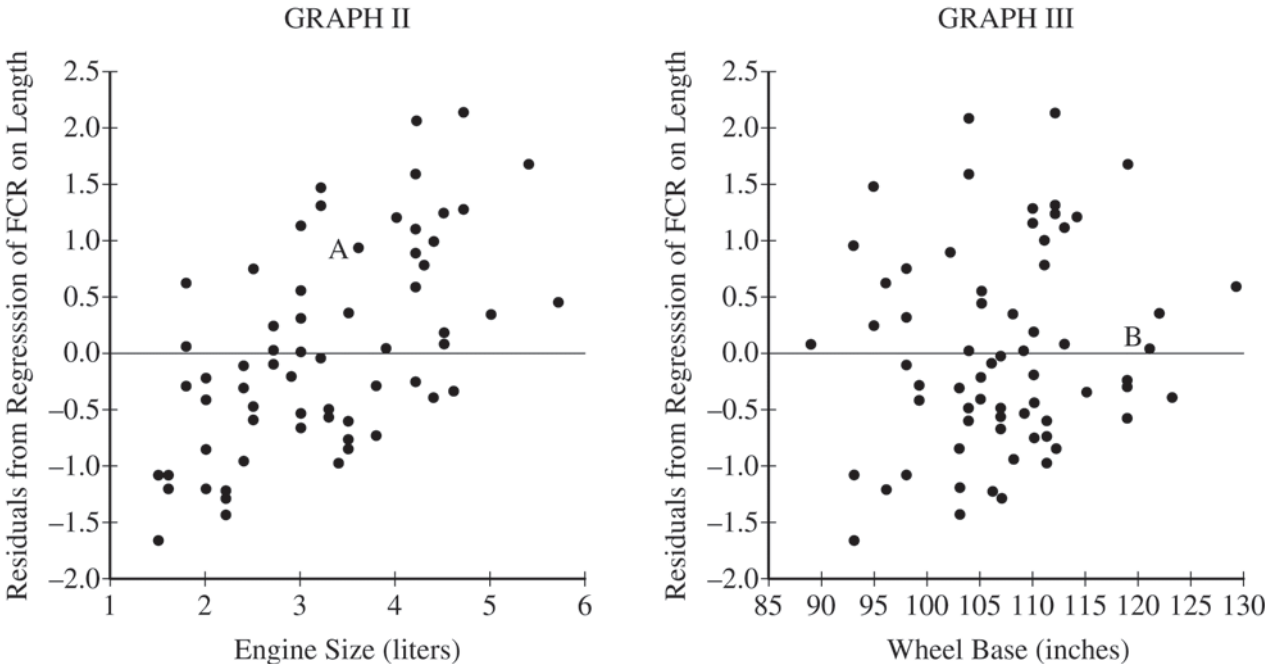
Jamal examined the scatterplot and determined that a linear model would be a reasonable way to express the relationship between FCR and length. A computer output from a linear regression is shown below.

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.

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?

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):

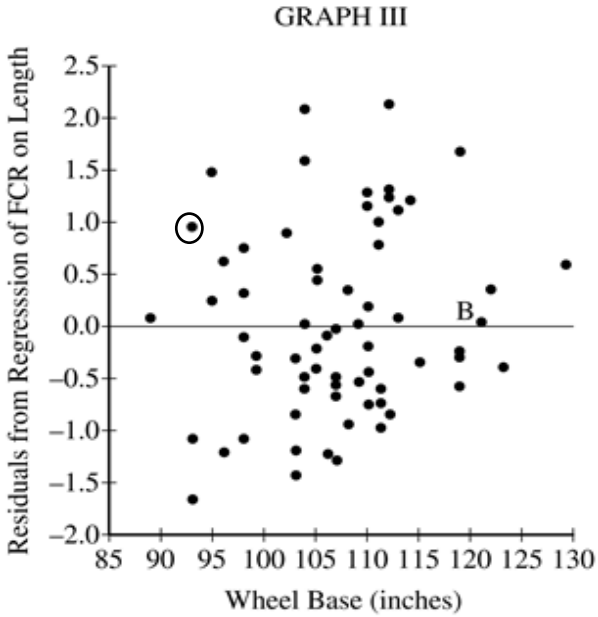
For a car with length 175 inches, the predicted value for the car's FCR, based on the least squares regression line, is

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

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.