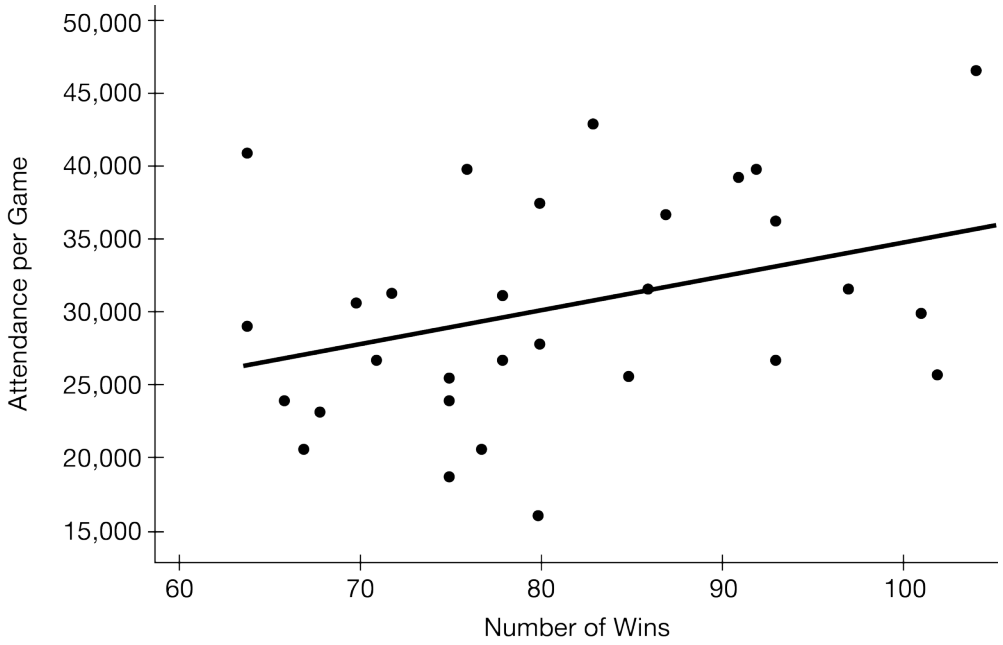
Analyzing Departures from Linearity Quiz

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 scatterplot shows the number of wins and the attendance per game for 30 baseball teams in 2017. Also shown are the least-squares regression line and computer output.



Term	Coef	SE Coef	T-Value	P-Value
Constant	10834	9716	1.12	0.274
Wins	235	119	1.98	0.058
$S = 7{,}377$		$R - \text{sq} = 12.29\%$		$\text{Adj}\, R - \text{sq} = 9.16\%$

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

(b) Explain why it is not reasonable to use the least-squares regression model to predict attendance per game for 0 wins.

(c) What is the value of the correlation coefficient for the sample?

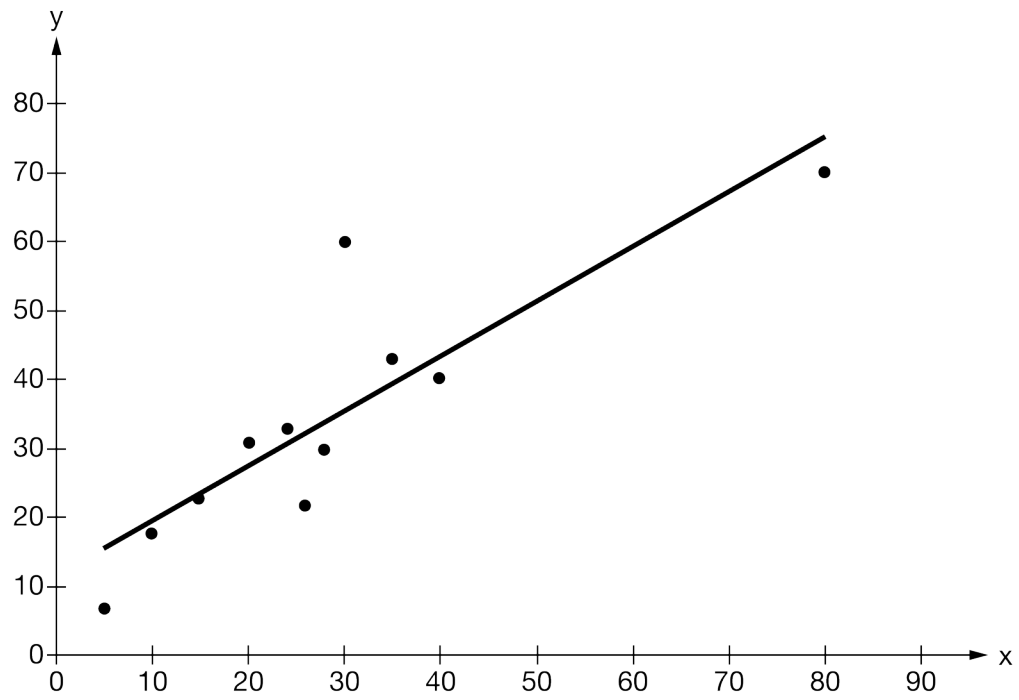
(d) If the point representing 64 wins and attendance of $40{,}786$ people per game is removed from the set of data and a new regression analysis is conducted, how would the following be impacted? Explain your reasoning.

(i) The slope of the least-squares line:

Analyzing Departures from Linearity Quiz

(ii) The correlation coefficient:

2. The following scatterplot shows two variables, x and y, along with a least-squares model.



Which of the following is a high leverage point with respect to the regression?

(A) $(5, 8)$

(B) $(20, 31)$

(C) $(27, 22)$

(D) $(30, 60)$

(E) $(80, 70)$

3. An exponential relationship exists between the explanatory variable and the response variable in a set of data. The common logarithm of each value of the response variable is taken, and the least-squares regression line has an equation of $\log(\hat{y}) = 7.3 - 1.5x$. Which of the following is closest to the predicted value of the response variable for $x = 4.8$?

(A) 0.1

(B) 0.68

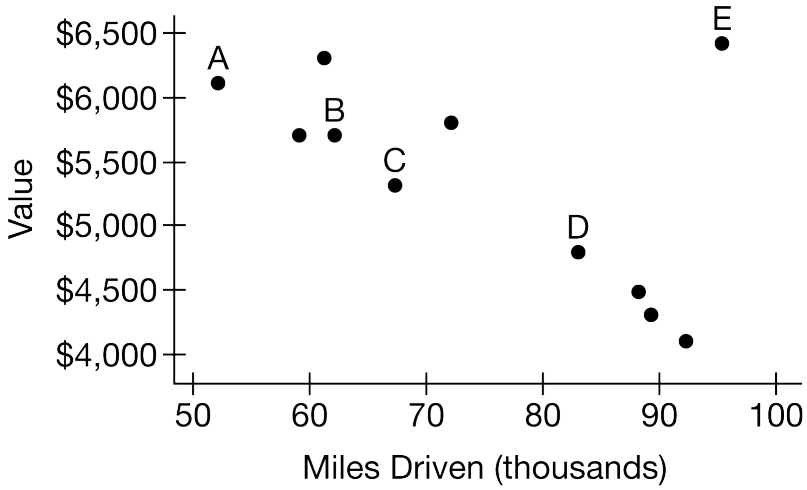
(C) 1.105

(D) 1.26

(E) 14.5

Analyzing Departures from Linearity Quiz

4. In a study to determine whether miles driven is a good predictor of trade-in value, 11 cars of the same age, make, model, and condition were randomly selected. The following scatterplot shows trade-in value and mileage for those cars. Five of the points are labeled A, B, C, D, and E, respectively.



Which of the five labeled points is the most influential with respect to a regression of trade-in value versus miles driven?

(A) A

(B) B

(C) C

(D) D

(E) E

5. Data were collected on two variables, x and y, to create a model to predict y from x. A scatterplot of the collected data revealed a curved pattern with a possible cubic relationship ($y = ax^3$, where a is a constant) between the variables. Which of the following transformations would be most appropriate for creating linearity between the variables?

(A) Taking the cube of y

(B) Taking the cube root of y

(C) Taking the cube root of both y and x

(D) Taking the log of y

(E) Taking the log of both y and x

6. The relationship between carbon dioxide emissions and fuel efficiency of a certain car can be modeled by the least-squares regression equation $\ln(\hat{y}) = 7 - 0.045x$, where x represents the fuel efficiency, in miles per gallon, and $\hat{y}$ represents the predicted carbon dioxide emissions, in grams per mile.

Which of the following is closest to the predicted carbon dioxide emissions, in grams per mile, for a car of this type with a fuel efficiency of 20 miles per gallon?

Analyzing Departures from Linearity Quiz

(A) 1.8

(B) 6.1

(C) 446

(D) 2,697

(E) 1,250,000

7. Which of the following statements about a least-squares regression analysis is true?

I. A point with a large residual is an outlier.
II. A point with high leverage has a y-value that is not consistent with the other y-values in the set.
III. The removal of an influential point from a data set could change the value of the correlation coefficient.

(A) I only

(B) II only

(C) I and III only

(D) III only

(E) I, II, and III