Unit 2 Progress Check: MCQ Part B

1. A small business owner has created a linear regression model to predict the number of new customers who will visit a shop based on the number of times the owner has an advertisement played on the radio. What is the explanatory variable and what is the response variable?

(A) Explanatory: number of new customers; response: number of times the advertisement is played

(B) Explanatory: number of times the advertisement is played; response: number of new customers ✓

(C) Explanatory: number of times the advertisement is played; response: number of purchases made by customers

(D) Explanatory: number of purchases made by customers; response: number of times the advertisement is played

(E) Explanatory: number of previous customers; response: number of new customers

Answer B

Correct. The explanatory variable is what is being used to predict, which is the number of times the advertisement is played. The response variable is what is to be predicted. In this case, it is the number of new customers who will visit the shop.

2. Bankers at a large financial institution created the linear regression model $\hat{d} = 0.37 - 0.0004s$ to predict the proportion of customers who would default on their loans, d, based on the customer's credit score, s.

For a customer with a credit score of 700, which of the following is true?

(A) The default proportion is predicted to be 0.09. ✓

(B) The default proportion will be 0.09.

(C) The default proportion is predicted to be approximately 1.75 million.

(D) The default proportion will be approximately 1.75 million.

(E) The default proportion is predicted to be 0.28.

Answer A

Correct. The predicted proportion of customers defaulting on their loans is found by substituting 700 for s in the regression equation, getting $\hat{d} = 0.37 - 0.0004(700) = 0.09$. The default proportion is predicted to be 0.09.

Unit 2 Progress Check: MCQ Part B

3. A marketing consultant, Sofia, has been studying the effect of increasing advertising spending on product sales. Sofia conducts several experiments, each time spending less than $1,000 in advertising. When she analyzed the relationship between $x =$ advertising spending and $y=$ product sales, the relationship was linear with $r = 0.90$. Her boss is thrilled and asks her to estimate product sales for $100,000 in advertising spending. Is it appropriate for her to calculate a predicted amount of product sales with advertising spending of $100,000 ?

(A) Yes, because the association is linear.

(B) Yes, because the association is positive.

(C) Yes, because the association is strong.

(D) No, because the value of the correlation is not equal to 1.

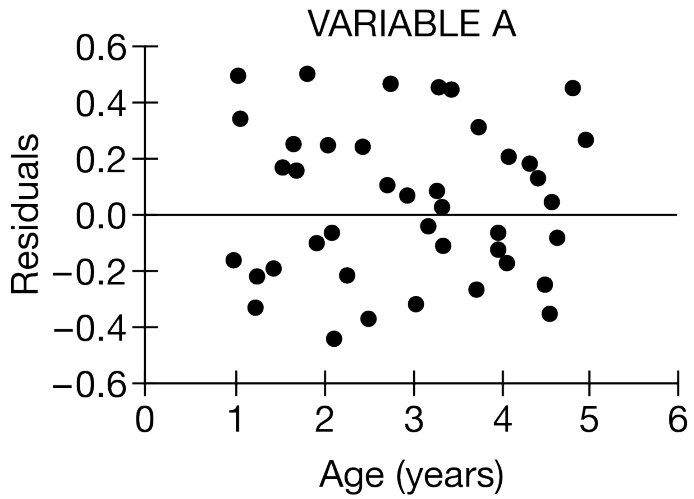
(E) No, because $100,000 is much greater than the values used in the experiment. ✓

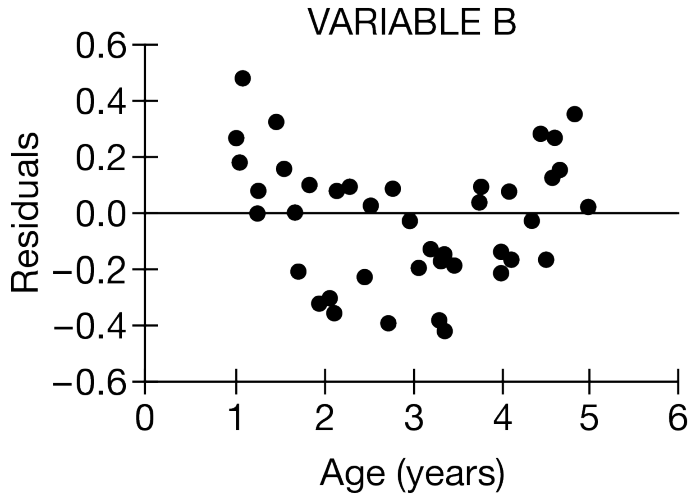
Answer E

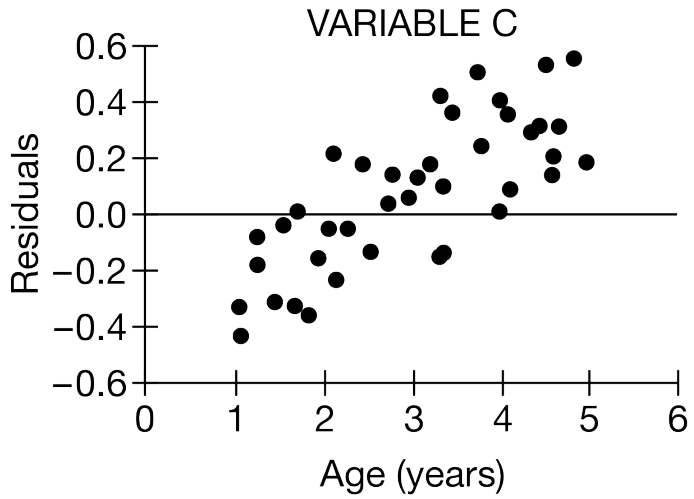
Correct. Predicting sales based on a value well beyond the interval that was used to create the model results in less reliable estimates and would not be appropriate. This is called extrapolation.

Unit 2 Progress Check: MCQ Part B

4. A researcher studying koi fish collected data on three variables, A, B, and C. The following residual plots show the residual for a model for predicting each variable from the age of the fish.







Unit 2 Progress Check: MCQ Part B

A conclusion that a linear model between the variable and age is appropriate is supported by which plot or plots?

(A) The plot for variable A only ✓

(B) The plot for variable B only

(C) The plot for variable C only

(D) The plots for variables A and C

(E) The plots for variables B and C

Answer A

Correct. The residual plot for variable A indicates that a linear model is appropriate for relating variable A to age, since this is the only residual plot in which there is no discernible pattern in the scatter of points.

5. A marketing consultant created a linear regression model to predict the number of units sold by a client based on the amount of money spent on marketing by the client. Which of the following is the best graphic to use to evaluate the appropriateness of the model?

(A) A dotplot

(B) A histogram

(C) A residual plot ✓

(D) A boxplot

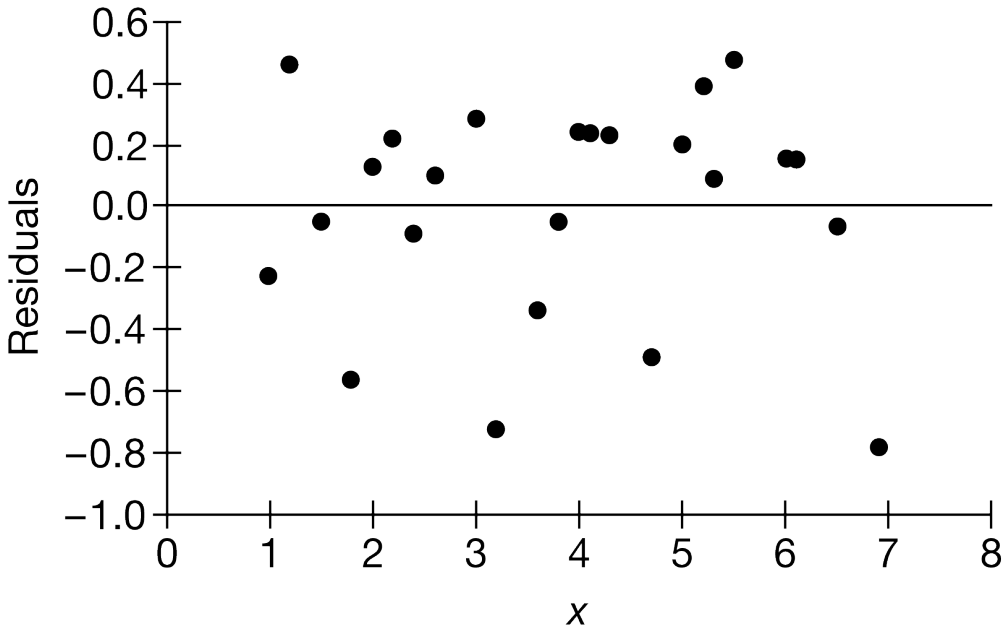
(E) A bar chart

Answer C

Correct. A residual plot can be used to evaluate the appropriateness of the regression model.

Unit 2 Progress Check: MCQ Part B

6. The following is a residual plot from a regression of a variable with the independent variable x.



Based on the plot, is it reasonable to conclude that a linear model is appropriate?

(A) Yes, because the plot shows no apparent pattern. ✓

(B) Yes, because the points in the plot display less variation as x increases.

(C) Yes, because the sum of the residuals is close to zero.

(D) No, because the plot shows no apparent pattern.

(E) No, because the points in the plot display more variation as x increases.

Answer A

Correct. No pattern in a residual plot is an indication that a linear model is appropriate.

7. A grocery store wants to examine the relationship between the sales amounts each day at two different locations, store A and store B. The sales amount each day, in dollars, was recorded for 10 days at each store. The least-squares regression line is $\hat{y} = -3{,}000 + 1.2x$, where x represents the sales amounts each day at store A and y represents the sales amounts each day at store B. If the mean of the 10 sales amounts for store B is \$45,000, what is the mean of the 10 sales amounts for store A?

Unit 2 Progress Check: MCQ Part B

(A) $35,000

(B) $40,000 ✓

(C) $42,000

(D) $45,000

(E) $51,000

Answer B

Correct. A property of the least-squares regression line is that the line always contains the point $(\bar{x}, \bar{y})$, so the equation for the least-squares regression line is true when $\bar{x}$ is substituted for x and $\bar{y}$ is substituted for $\hat{y}$. This yields $\bar{y} = -3{,}000 + 1.2\bar{x}$. Substituting $\bar{y} = 45{,}000$ gives $45{,}000 = -3{,}000 + 1.2\bar{x}$, or $1.2\bar{x} = 48{,}000$, so $\bar{x} = \frac{48{,}000}{1.2} = 40{,}000$. The mean of the 10 sales amounts at store A is $40,000.

8. The least-squares regression model $\hat{y} = -3.4 + 5.2x$ and correlation coefficient $r = 0.66$ were calculated for a set of bivariate data with variables x and y. Which of the following is closest to the proportion of the variation in y that cannot be explained by the explanatory variable?

(A) 81%

(B) 66%

(C) 56% ✓

(D) 44%

(E) 34%

Answer C

Correct. The coefficient of determination, r^2, is the proportion of the response variable variation that can be explained by the explanatory variable. Thus $1 - r^2 = 1 - (-0.66)^2 = 1 - 0.4356 = 0.5644$ is the proportion of the variation in the response variable that cannot be explained by the explanatory variable.

9. A botanist created a linear model to predict plant height from soil acidity (pH level) for a certain type of plant. The slope of the model was 2.5 centimeters per pH level, the standard deviation of the sample of plant heights was 4 centimeters, and the standard deviation of the soil acidities was 1 pH level. What is the value of the correlation coefficient?

Unit 2 Progress Check: MCQ Part B

(A) 0.015
(B) 0.10
(C) 0.25
(D) 0.625 ✓
(E) 1.60

Answer D

Correct. If the equation of the regression line is $\hat{y} = a + bx$, where b is the slope, then $b = r\frac{s_y}{s_x}$. Substituting the appropriate values in the formula yields $2.5 = r\left(\frac{4}{1}\right)$, which results in a correlation of 0.625.

10. In baseball, two statistics, the ERA (Earned Run Average) and the WHIP (Walks and Hits per Inning Pitched), are used to measure the quality of pitchers. For both measures, smaller values indicate higher quality. The following computer output gives the results from predicting ERA by using WHIP in a least-squares regression for the 2017 baseball season.

Variable	DF	Estimate	SE	T
Intercept	1	-5.0	0.26	-19.3
WHIP	1	6.8	0.14	47.4

Which of the following statements is the best interpretation of the value 6.8 shown in the output?

(A) ERA is predicted to increase by 6.8 units for each 1 unit increase of WHIP. ✓
(B) WHIP is predicted to increase by 6.8 units for each 1 unit increase of ERA.
(C) For a pitcher with 0 units of WHIP, the ERA is predicted to be approximately 6.8 units.
(D) For a pitcher with 0 units of ERA, the WHIP is predicted to be approximately 6.8 units.
(E) Approximately 6.8% of the variability in ERA is due to its linear relationship with WHIP.

Answer A

Correct. The estimated slope is 6.8, which is the change in response (ERA) as the explanatory variable (WHIP) increases by a single unit.

Unit 2 Progress Check: MCQ Part B

11. Jordan is working on a business model for a sandwich shop. Based on past data, he developed the model $\hat{n} = 150 - 3p$, where $\hat{n}$ represents the predicted number of turkey sandwiches sold in one day for a price of p dollars per sandwich.

Which of the following is the best description of the slope of the model?

(A) For each increase of \$3 in the price of the sandwich, the number sold is predicted to decrease, on average, by 150.

(B) For each increase of \$3 in the price of the sandwich, the number sold is predicted to increase, on average, by 150.

(C) For each increase of \$1 in the price of the sandwich, the number sold is predicted to decrease, on average, by 3. ✓

(D) For each increase of \$1 in the price of the sandwich, the number sold is predicted to increase, on average, by 3.

(E) For each increase of \$1 in the price of the sandwich, the number sold is predicted to decrease, on average, by 150.

Answer C

Correct. The slope of the line is -3, which indicates a one unit increase in the independent variable, price, is associated with a 3 unit decrease in the dependent variable, number sold. Because the number sold is not deterministic, the slope represents an average change in the long run.

12. Researchers are investigating how the amount of monthly rainfall, measured in centimeters (cm), affects the monthly growth, in cm, of a certain plant. From a sample of data, the researchers created a least-squares regression line. Computer output is shown in the following table.

Variable	DF	Estimate	SE	T
Intercept	1	0.75	0.350	2.14
Rainfall	1	0.15	0.025	6.00

Which of the following statements is an interpretation of the value 0.75 shown in the table?

(A) Monthly growth is expected to increase by 0.75 cm when rainfall increases by 1 cm.

(B) Rainfall is expected to increase by 0.75 cm when monthly growth increases by 1 cm.

(C) For a month with 0 cm of rainfall, the monthly growth is expected to be approximately 0.75 cm. ✓

(D) For a plant with 0 cm of monthly growth, the month had an expected rainfall of approximately 0.75 cm.

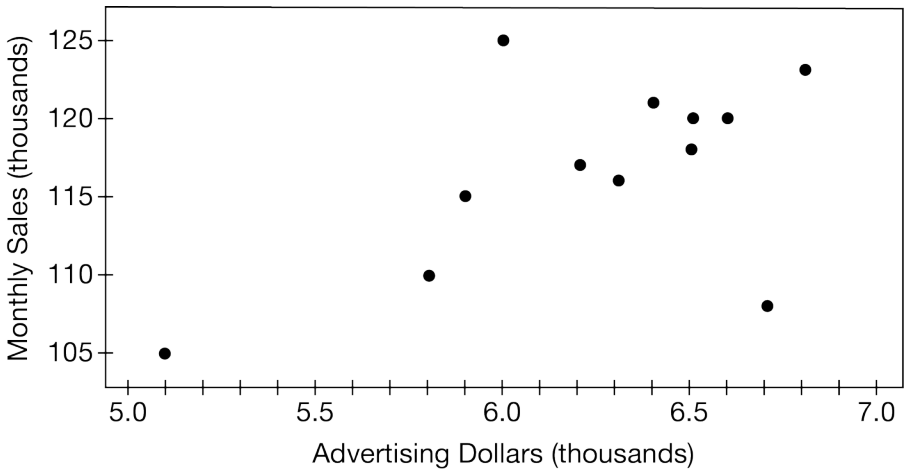
(E) Approximately 75% of the variability in monthly growth is due to its linear relationship with rainfall.

Unit 2 Progress Check: MCQ Part B

Answer C

Correct. The intercept is the estimated value of the response variable (monthly growth) when the explanatory variable (rainfall) is zero.

13. The following scatterplot shows a company's monthly sales, in thousands of dollars, versus monthly advertising dollars spent, in thousands of dollars.



Which of the following points is most likely a high-leverage point with respect to a regression of monthly sales versus advertising dollars?

(A) $(5.1, 105)$ ✓

(B) $(5.8, 110)$

(C) $(6.0, 125)$

(D) $(6.7, 108)$

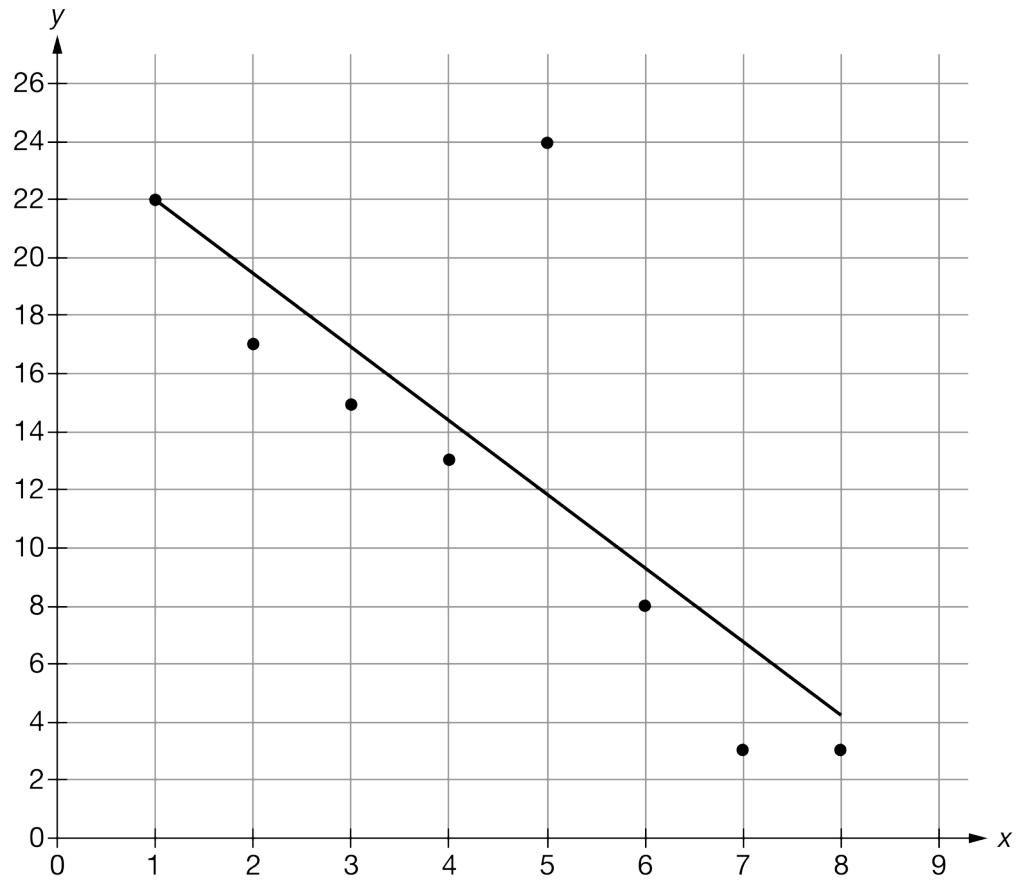
(E) $(6.8, 123)$

Answer A

Correct. A high-leverage point in regression has a much larger or much smaller x-value than the other observations. The value of 5.1 along the horizontal axis is substantially farther away from the other values along the axis, which tend to fall between 5.8 and 6.8.

Unit 2 Progress Check: MCQ Part B

14. The following scatterplot shows two variables along with a least-squares regression line.



Which of the following points is an outlier for the data?

(A) $(1, 22)$

(B) $(2, 17)$

(C) $(4, 13)$

(D) $(5, 24)$ ✓

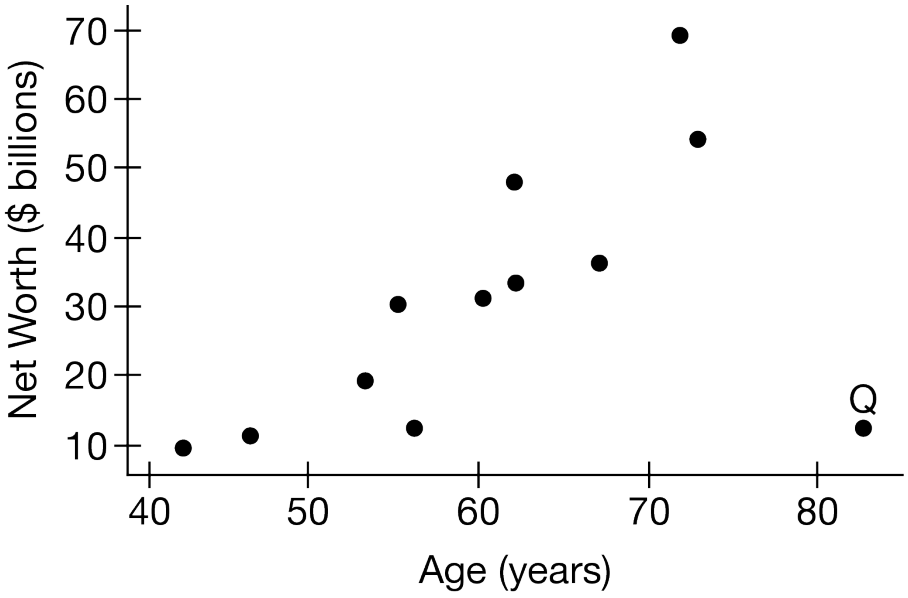
(E) $(8, 3)$

Answer D

Correct. The point $(5, 24)$ is an outlier because it does not follow the general pattern of the rest of the data. Also, since the observed y-value is much higher than predicted (the least-squares regression line), this point will have a very large residual.

Unit 2 Progress Check: MCQ Part B

15. The following scatterplot shows the ages, in years, of 12 of the wealthiest people in the world along with their net worth, in billions of dollars. The data point at age 83 is labeled Q.



Suppose point Q is removed from the data set. Which of the following is likely not affected by the removal?

(A) The correlation coefficient

(B) The sign of the slope coefficient ✓

(C) The value of the slope coefficient

(D) The sum of the squared residuals

(E) The net worth intercept

Answer B

Correct. Although removing point Q will change the least squares regression line significantly, there is a positive relationship between the variables whether or not point Q is considered in the data set.

16. A real estate agent wants to predict the selling price of single-family homes from the size of each house. A scatterplot created from a sample of houses shows an exponential relationship between price, in thousands of dollars, and size, in 100 square feet. To create a linear model, the natural logarithm of price was taken and the least-squares regression line was given as $\ln\left(\widehat{price}\right) = 2.08 + 0.11\left(size\right)$. Based on the model, which of the following is closest to the predicted selling price for a house with a size of 3,200 square feet?

Unit 2 Progress Check: MCQ Part B

(A) $54,500

(B) $270,000 ✓

(C) $354,000

(D) $398,000

(E) $560,000

Answer B

Correct. Since the variable size is in the units of 100 square feet, 32 needs to be the number used for price. Substituting 32 into the equation for price gives $\ln\left(\widehat{price}\right) = 5.6$. The number e raised to the power of 5.6 is about 270. Multiplying the result by 1,000 (because the units for price are in thousands of dollars) gives $270,000.

17. Workers at a warehouse of consumer goods gather items from the warehouse to fill customer orders. The number of items in a sample of orders and the time, in minutes, it took the workers to gather the items were recorded. A scatterplot of the recorded data showed a curved pattern, and the square root of the number of items was taken to create a linear pattern. The following table shows computer output from the least-squares regression analysis created to predict the time it takes to gather items from the number of items in an order.

Predictor	**Coef**
Constant	3.0979
Square root of items	2.7633
	R-Sq = 96.7%

Based on the regression output, which of the following is the predicted time, in minutes, that it took to gather the items if the order has 22 items?

(A) 7.99

(B) 16.06 ✓

(C) 17.29

(D) 27.49

(E) 63.89

Unit 2 Progress Check: MCQ Part B

Answer B

Correct. The table shows the coefficient of the square root of the items. Substituting 22 into the regression equation $\hat{y} = 3.0979 + 2.7633\sqrt{x}$, where x represents the number of items and y represents the time to gather the items, yields 16.06.

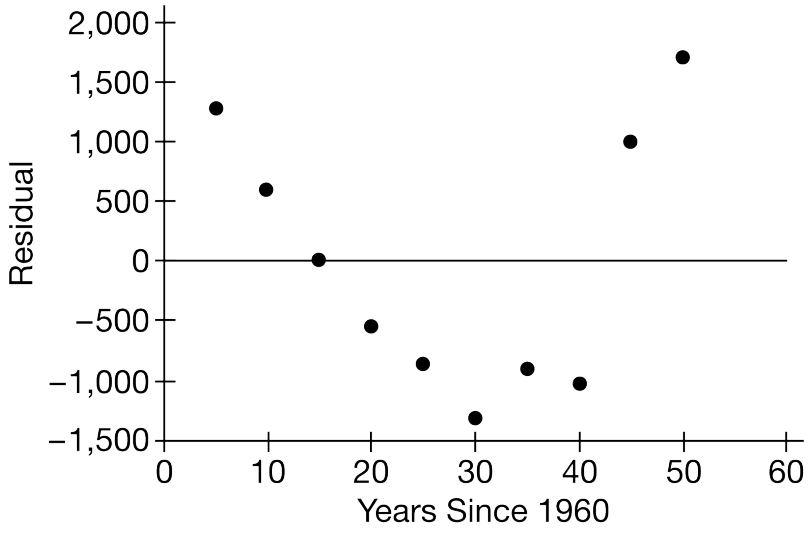
Unit 2 Progress Check: MCQ Part B

18. A new town was incorporated in 1960. The size of the town's population was recorded every 5 years after 1960. Using the variables x, for number of years since 1960, and y, for the size of the population, three models were created to predict the population from the number of years since 1960.

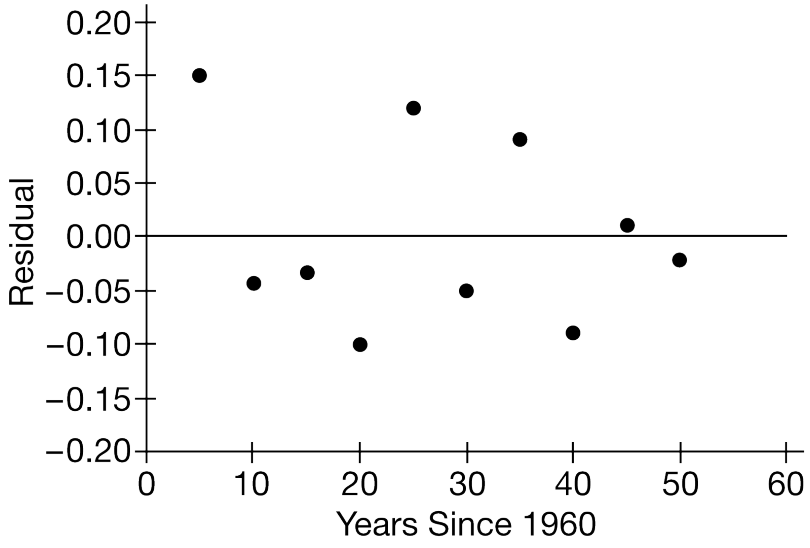
Model I predicts y from x.
Model II predicts $\ln(y)$, the natural logarithm of y, from x.
Model III predicts $\ln(y)$ from $\ln(x)$.

The following graphs show the residual plot for each model.

Residual Plot for Model I

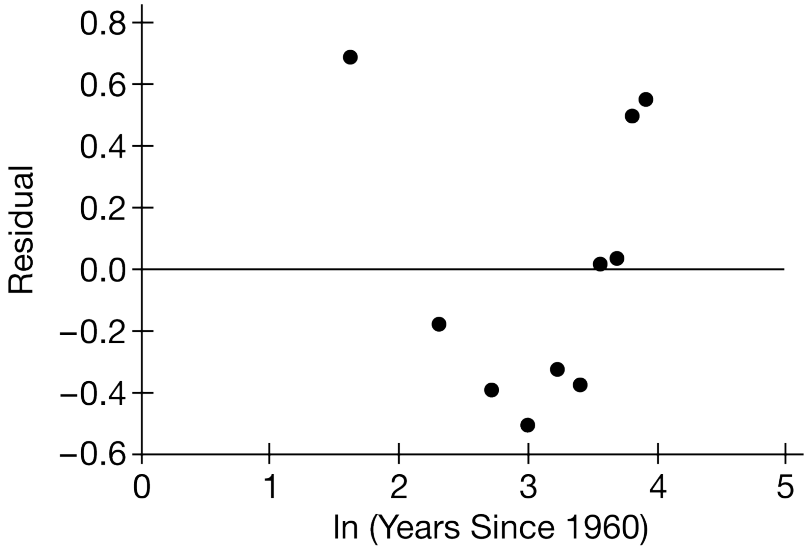


Residual Plot for Model II



Unit 2 Progress Check: MCQ Part B

Residual Plot for Model III



Which of the following statements is the best interpretation of the residual plots?

(A) The residual plot for model I indicates that a quadratic model is the most appropriate among the three models.

(B) The residual plots for models I and III indicate that either model is appropriate and better than model II.

(C) The residual plot for model II indicates that it is the most appropriate among the three models. ✓

(D) All the residual plots indicate that any of the three models are appropriate for the prediction.

(E) All the residual plots indicate that none of the three models is appropriate for the prediction.

Answer C

Correct. The residual plot for model II shows no pattern in the scatter of points, indicating that a linear model is appropriate for relating $\ln(y)$ to x. There are discernible patterns that are U-shaped in the residual plots of both model I and model III, indicating that the response and explanatory variables may not be linearly related and a linear model is not appropriate for those models.