

10.E: Prediction (Optional Exercises)

Use the following information to answer the next two exercises. An electronics retailer used regression to find a simple model to predict sales growth in the first quarter of the new year (January through March). The model is good for 90 days, where x is the day. The model can be written as follows:

$$\hat{y} = 101.32 + 2.48x \quad (10.E.1)$$

where \hat{y} is in thousands of dollars.

? Exercise 12.6.2

What would you predict the sales to be on day 60?

Answer

\$250,120

? Exercise 12.6.3

What would you predict the sales to be on day 90?

Use the following information to answer the next three exercises. A landscaping company is hired to mow the grass for several large properties. The total area of the properties combined is 1,345 acres. The rate at which one person can mow is as follows:

$$\hat{y} = 1350 - 1.2x \quad (10.E.2)$$

where x is the number of hours and \hat{y} represents the number of acres left to mow.

? Exercise 12.6.4

How many acres will be left to mow after 20 hours of work?

Answer

1,326 acres

? Exercise 12.6.5

How many acres will be left to mow after 100 hours of work?

? Exercise 12.6.7

How many hours will it take to mow all of the lawns? (When is $\hat{y} = 0$?)

Answer

1,125 hours, or when $x = 1,125$

Table contains real data for the first two decades of AIDS reporting.

Adults and Adolescents only, United States

Year	# AIDS cases diagnosed	# AIDS deaths
Pre-1981	91	29
1981	319	121
1982	1,170	453
1983	3,076	1,482

1984	6,240	3,466
1985	11,776	6,878
1986	19,032	11,987
1987	28,564	16,162
1988	35,447	20,868
1989	42,674	27,591
1990	48,634	31,335
1991	59,660	36,560
1992	78,530	41,055
1993	78,834	44,730
1994	71,874	49,095
1995	68,505	49,456
1996	59,347	38,510
1997	47,149	20,736
1998	38,393	19,005
1999	25,174	18,454
2000	25,522	17,347
2001	25,643	17,402
2002	26,464	16,371
Total	802,118	489,093

? Exercise 12.6.8

Graph “year” versus “# AIDS cases diagnosed” (plot the scatter plot). Do not include pre-1981 data.

? Exercise 12.6.9

Perform linear regression. What is the linear equation? Round to the nearest whole number.

Answer

Check student’s solution.

? Exercise 12.6.10

Write the equations:

- Linear equation: _____
- $a =$ _____
- $b =$ _____
- $r =$ _____
- $n =$ _____

? Exercise 12.6.11

Solve.

- When $x = 1985$, $\hat{y} = \underline{\hspace{2cm}}$
- When $x = 1990$, $\hat{y} = \underline{\hspace{2cm}}$
- When $x = 1970$, $\hat{y} = \underline{\hspace{2cm}}$ Why doesn't this answer make sense?

Answer

- When $x = 1985$, $\hat{y} = 25, 52$
- When $x = 1990$, $\hat{y} = 34, 275$
- When $x = 1970$, $\hat{y} = -725$ Why doesn't this answer make sense? The range of x values was 1981 to 2002; the year 1970 is not in this range. The regression equation does not apply, because predicting for the year 1970 is extrapolation, which requires a different process. Also, a negative number does not make sense in this context, where we are predicting AIDS cases diagnosed.

? Exercise 12.6.11

Does the line seem to fit the data? Why or why not?

? Exercise 12.6.12

What does the correlation imply about the relationship between time (years) and the number of diagnosed AIDS cases reported in the U.S.?

Answer

Also, the correlation $r = 0.4526$. If r is compared to the value in the 95% Critical Values of the Sample Correlation Coefficient Table, because $r > 0.423$, r is significant, and you would think that the line could be used for prediction. But the scatter plot indicates otherwise.

? Exercise 12.6.13

Plot the two given points on the following graph. Then, connect the two points to form the regression line.


 Blank graph with horizontal and vertical axes.

Figure 10.E. 1.

Obtain the graph on your calculator or computer.

? Exercise 12.6.14

Write the equation: $\hat{y} = \underline{\hspace{2cm}}$

Answer

$$\hat{y} = 3,448,225 + 1750x$$

? Exercise 12.6.15

Hand draw a smooth curve on the graph that shows the flow of the data.

? Exercise 12.6.16

Does the line seem to fit the data? Why or why not?

Answer

There was an increase in AIDS cases diagnosed until 1993. From 1993 through 2002, the number of AIDS cases diagnosed declined each year. It is not appropriate to use a linear regression line to fit to the data.

? Exercise 12.6.17

Do you think a linear fit is best? Why or why not?

? Exercise 12.6.18

What does the correlation imply about the relationship between time (years) and the number of diagnosed AIDS cases reported in the U.S.?

Answer

Since there is no linear association between year and # of AIDS cases diagnosed, it is not appropriate to calculate a linear correlation coefficient. When there is a linear association and it is appropriate to calculate a correlation, we cannot say that one variable “causes” the other variable.

? Exercise 12.6.19

Graph “year” vs. “# AIDS cases diagnosed.” Do not include pre-1981. Label both axes with words. Scale both axes.

? Exercise 12.6.20

Enter your data into your calculator or computer. The pre-1981 data should not be included. Why is that so?

Write the linear equation, rounding to four decimal places:

Answer

We don’t know if the pre-1981 data was collected from a single year. So we don’t have an accurate x value for this figure.

Regression equation: $\hat{y}(\text{\#AIDS Cases}) = -3,448,225 + 1749.777(\text{year})$

	Coefficients
Intercept	-3,448,225
X Variable 1	1,749.777

? Exercise 12.6.21

Calculate the following:

- a. $a =$ _____
- b. $b =$ _____
- c. correlation = _____
- d. $n =$ _____