Name ______ Date ___

4.5 Notetaking with Vocabulary

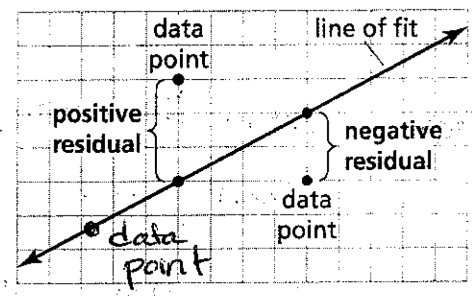
I can use residuals to determine how well lines of fit model data.
I can use technology to find lines of best fit.
I can distinguish between extrapolation and interpolation.
I can distinguish between correlation and causation.

Core Concepts

Residuals

line y
data y − model y

A residual is the difference of the y-value of a data point and the corresponding y-value found using the line of fit. A residual can be + / − / 0.



2nd

A scatter plot of the residuals shows how well a model fits the data. If the model is a good fit, then the absolute values of the residuals are small, and the residual points will be more or less evenly dispersed about the horizontal axis. If the model is not a good fit then the residual points will form some type of pattern that suggests the data are not linear. Widely scattered points suggest that the data might have no correlation.

Graphing calculators and computers use a method called linear regression to find a precise line of fit called a line of best fit (only 1). The line best models a set of data. A calculator often gives a value r, called the correlation coefficient. This value tells whether the correlation is + or — and how closely the equation models the data. Values of r range from −1 to 1. When r is close to −1 or 1, there is a strong correlation between the variables. As r, gets closer to 0, the correlation becomes weaker.

Practice

In Exercises 1 and ~~2~~, use residuals to determine whether the model is a good fit for the data in the table. Explain.

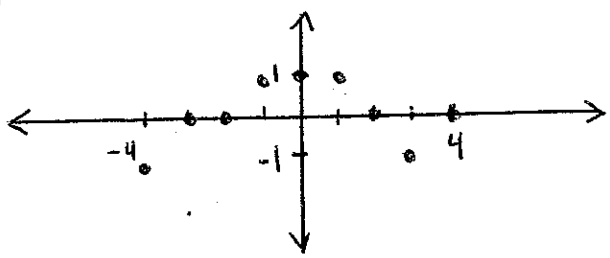
res = data y − model y

1. $y = -3x + 2$ → model

x	−4	−3	−2	−1	0	1	2	3	4
y	13	11	8	6	3	0	−4	−8	−10
model y	14	11	8	5	2	−1	−4	−7	−10
residual	−1	0	0	1	1	1	0	−1	0

} data

(−4, −1)



The model is a good fit to the data because the residuals are small and evenly distributed about the x-axis.

2. $y = -0.5x + 1$

x	0	1	2	3	4	5	6	7	8
y	2	0	−3	−5	−7	−6	−4	−3	−1

3. The table shows the number of visitors y to a particular beach for average daily temperatures x.

Average Daily Temperature (°F)	Number of Beach Visitors
80	100
82	150
83	145
85	190
86	215
88	263
89	300
90	350

a. Use a graphing calculator to find an equation of the line of best fit. Then plot the data and graph the equation in the same viewing window.

DESMOS ⇒ Statistics: Linear Regression

put points in table

$m = 23.57$ $b = -1798.56$

$r = .98$ $y = 23.57x - 1798.56$

b. Identify and interpret the correlation coefficient.

$r = .98$ There is a strong positive correlation between temperature and number of beach vistors.

c. Interpret the slope and y-intercept of the line of best fit.

$m = 23.57 = \frac{\Delta y}{\Delta x}$ $\frac{\text{\# of visitors}}{\text{degrees}}$

For every $1°$ increase in temperature, the model predicts there will 23 to 24 more beach visitors.

$b = -1{,}798.56$ The y-intercept has no meaning in the context of this situation.

predict $\boxed{\text{in}}$side range

Using a graph or its equation to approximate (predict) a value between two known values is called $\boxed{\text{in}}$terpolation. Using a graph or its equations to predict a value outside the range of known values is called extrapolation. In general, the farther removed a value is from the known values, the less confidence you can have in the accuracy of the prediction.

Refer to the example 3. Use the equation of the line of best fit. $y = 23.57x - 1{,}798.56$

a. Approximate the number of beach visitors if the outside temperature is 70° F. Is this extrapolation or interpolation?

Extrapolation

70° outside 80°–90°

$y = 23.57(70) - 1{,}798.56$

$y \approx -149$!

Our answer doesn't make sense.

b. Approximate the number of beach visitors if the outside temperature is 87° F. Is this extrapolation or interpolation?

Interpolation

$y = 23.57(87) - 1{,}798.56$

$y \approx 252$

Our model predicts 252 beach visitors if the temperature is 87°F.

c. What temperature does the model predict if 120 people visited the beach? (Round to the nearest degree)

$$120 = 23.57x - \cancel{1{,}798.56}$$
$$+1{,}798.56 \qquad +\cancel{1{,}798.56}$$
$$\frac{1{,}918.56}{23.57} = \frac{\cancel{23.57}x}{\cancel{23.57}}$$
$$81° \approx x$$

The model predicts a temperature of 81°F if there were 120 beach visitors.

When a change in one variable causes a change in another variable it is called causation.

Causation produces a strong correlation between the two variables. The converse is NOT true. In other words, correlation does not imply causation.

It is very important to identify the independent and dependent variable in the sentence.
x y

Tell whether a correlation is likely in the situation. If so, tell whether there is a causal relationship.

Explain your reasoning.

ind. dep.

a. time spent exercising and the number of calories burned

There would be a positive correlation between time spent exercising and number of calories burned.

There is a causal relationship because as you increase exercise time, you burn more calories.

b. eating organic food and getting a good score on a test

ind. dep var.

There could be a weak positive correlation between eating organic food and getting a good score on the test but there is no causal relationship. Eating organic food does not raise your test scores.

c. the number of banks and the population of a city

ind. dep.

There would be a positive correlation between number of banks and the population of the city. But there is no causal relationship. Building more banks will not increase the population of a city.