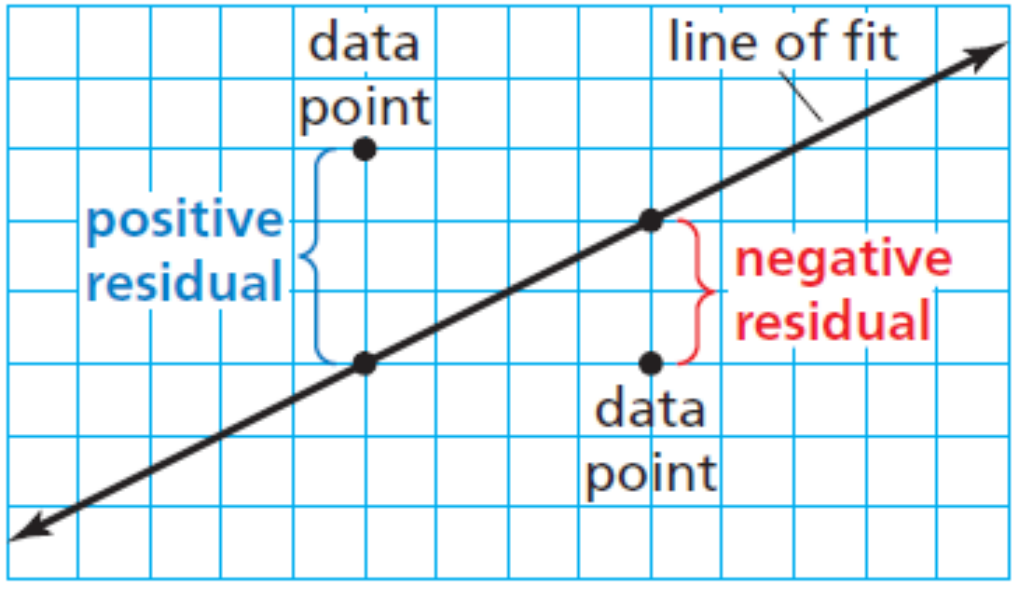
Analyzing Lines of Best Fit

12.6

Key Idea

Residuals

A **residual** is the difference between the y-value of a data point and the corresponding y-value found using the line of fit. A residual can be positive, negative, or zero.



If the model (line of fit) is good, then the residual points will be randomly dispersed about the horizontal axis. If the model is not a good fit, then the residual points will form some type of pattern.

EXAMPLE 1 Using Residuals

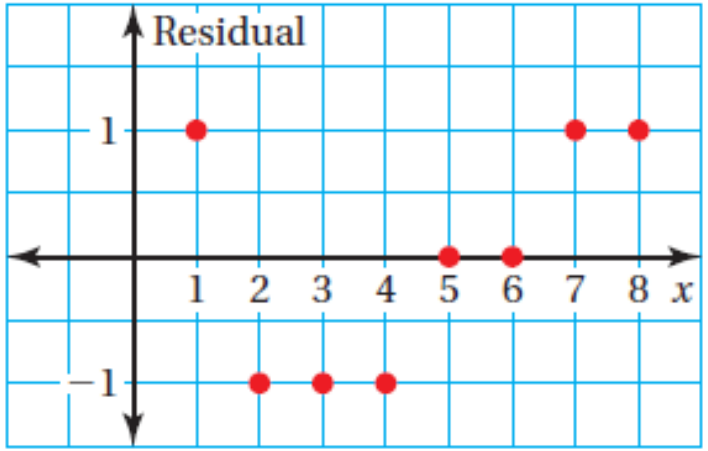
Week, x	Sales (millions), y
1	\$19
2	\$15
3	\$13
4	\$11
5	\$10
6	\$8
7	\$7
8	\$5

In Example 3 in Section 12.5, the equation $y = -2x + 20$ models the data in the table at the left. Is the model a good fit?

Step 1: Calculate the residuals and organize your results in a table.

x	y	y-Value from Model	Residual
1	19	$-2(1) + 20 = 18$	$19 - 18 = 1$
2	15	$-2(2) + 20 = 16$	$15 - 16 = -1$
3	13	$-2(3) + 20 = 14$	$13 - 14 = -1$
4	11	$-2(4) + 20 = 12$	$11 - 12 = -1$
5	10	$-2(5) + 20 = 10$	$10 - 10 = 0$
6	8	$-2(6) + 20 = 8$	$8 - 8 = 0$
7	7	$-2(7) + 20 = 6$	$7 - 6 = 1$
8	5	$-2(8) + 20 = 4$	$5 - 4 = 1$

Step 2: Use the points (x, residual) to make a scatter plot.



The points are randomly dispersed about the x-axis. The equation is a good fit.

EXAMPLE 2 Using Residuals

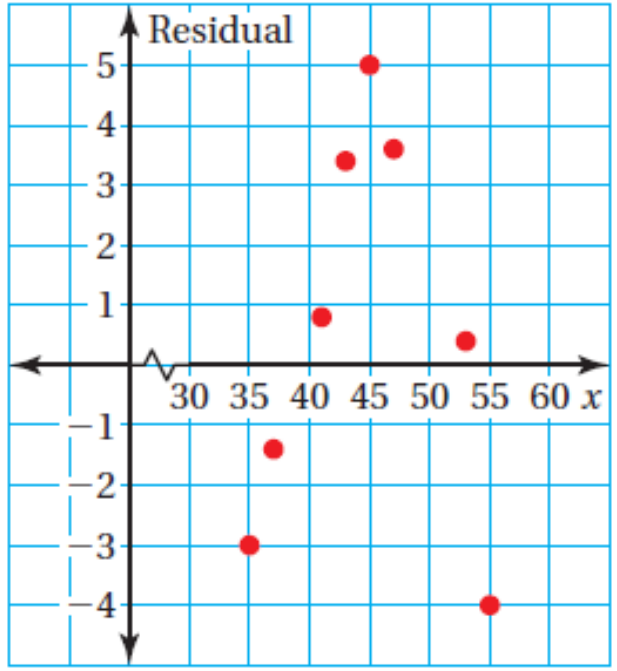
Age, x	Salary, y
35	42
37	44
41	47
43	50
45	52
47	51
53	49
55	45

The table at the left shows the ages x and salaries y (in thousands of dollars) of eight employees at a company. The equation $y = 0.2x + 38$ models the data. Is the model a good fit?

Step 1: Calculate the residuals and organize your results in a table.

x	y	y-Value from Model	Residual
35	42	$.2(35) + 38 = 45$	$42 - 45.0 = -3.0$
37	44	$.2(37) + 38 = 45.4$	$44 - 45.4 = -1.4$
41	47	$.2(41) + 38 = 46.2$	$47 - 46.2 = 0.8$
43	50	$.2(43) + 38 = 46.6$	$50 - 46.6 = 3.4$
45	52	$.2(45) + 38 = 47$	$52 - 47.0 = 5.0$
47	51	$.2(47) + 38 = 47.4$	$51 - 47.4 = 3.6$
53	49	$.2(53) + 38 = 48.6$	$49 - 48.6 = 0.4$
55	45	$.2(55) + 38 = 49$	$45 - 49.0 = -4.0$

Step 2: Use the points (x, residual) to make a scatter plot.



The points form a U-shaped pattern. This is not a good line of fit.

On Your Own

1. The table shows the attendance y (in thousands) at an amusement park from 2000 to 2009, where $x = 0$ represents the year 2000. The equation $y = -9.8x + 850$ models the data. Is the model a good fit?

Year, x	0	1	2	3	4	5	6	7	8	9
Attendance, y	850	845	828	798	800	792	785	781	775	760

The points $(x, \text{residual})$ are randomly dispersed about the horizontal axis. So, the model $y = -9.8x + 850$ is a good fit.