Module 19 Analyzing Data and Statistics
**Lesson 4** Finding a Line of Best Fit

independent practice

For Questions 1 and 2, use the data to make a scatter plot. Then, indicate whether the graph shows a positive correlation, a negative correlation, or no correlation between the two sets of data.

1. Rita runs her school's football concession stand. For each game last season, she recorded the number of spirit banners sold and the team's final score in the game.

Final Score	Banners	Final Score	Banners
14	28	20	40
7	15	3	6
10	20	13	22
21	32	17	22
10	18	3	9

The scatter plot indicates a positive correlation.

**As the team's final score increases, the number of**

banners increases.
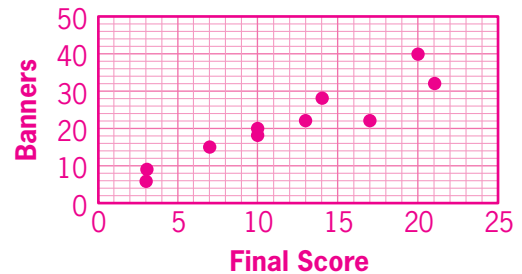
2. Rita also recorded the number of fans present for the marching band's post-game show.

Final Score	Fans	Final Score	Fans
14	31	20	34
7	12	3	7
10	50	13	42
21	37	17	18
10	40	3	6

The scatter plot does not indicate a relationship.

There is no correlation between the home team's

final score and the number of fans present for the
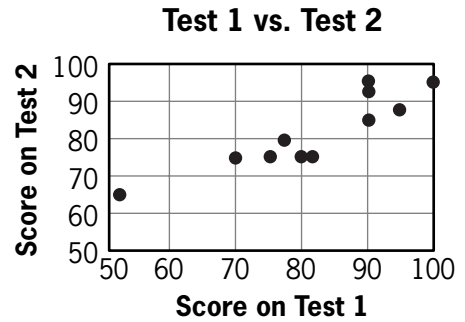
post-game show.
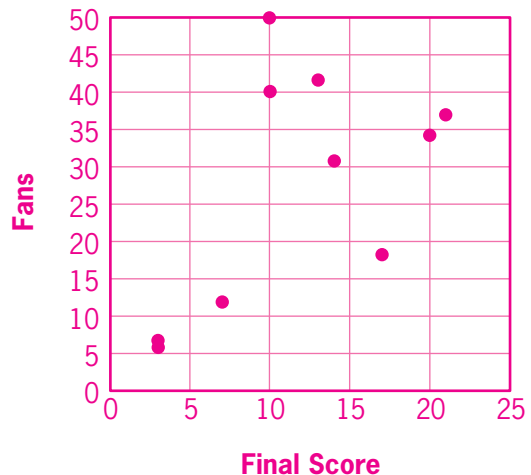
3. Use the scatter plot on the right to write the equation of the line of fit. Use the points (75, 75) and (100, 95) to find the equation.

Slope: 0.8; y = 0.8x + 15

Test 1 vs. Test 2

4. Use the line of fit found in Question 3 to predict the Test 2 score of a student who scored 60 on Test 1.

y = 63; According to the model, a student who

scores 60 on Test 1 will score 63 on Test 2.

5. Use the scatter plot on the right to write the equation of the line of fit. Use the points on the line (5, 4) and (8, 3) to find the equation.
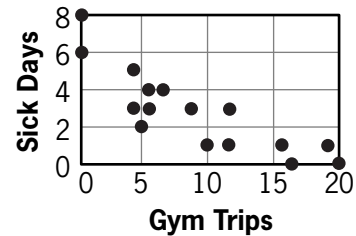
Slope: $-\frac{1}{3}$; y = $-\frac{1}{3}$x + $\frac{17}{3}$

Average Monthly Trips to the Gym vs. Yearly Sick Days Used

6. Use the line of fit found in Question 5 to predict the number of sick days that will be used by a person who goes to the gym about two times a month.

**y = 5; According to the model, a person who goes to the gym about two**

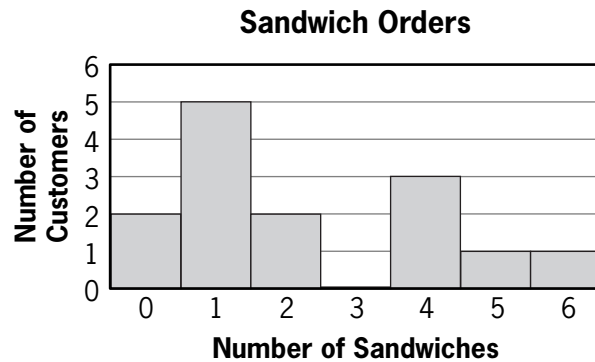**times a month will use about five sick days annually.**

Journal

1. Explain how to use a scatter plot to determine whether there is a correlation between two variables.
2. Explain what a line of fit is. How is a line of fit formed?
3. Arnold argued that the line drawn on the graph is a good line of fit since half the points lie above the line and half lie below it. Do you agree with Arnold? Explain.
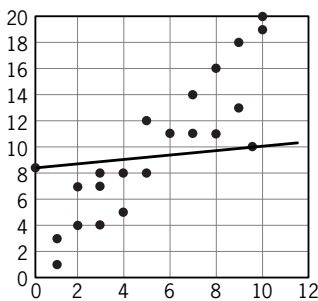
4. In a management training course, Sandra learned that she could predict the number of cups of coffee her convenience store would sell based on the day's predicted high temperature. Her trainer said that the line equation $y = -2x + 500$ is a good line of fit to describe the relationship between the number of cups of coffee sold, y, and the day's predicted high temperature, x. If the line of fit was developed using a scatter plot, what did the scatter plot look like? What type of correlation exists between x and y?

Cumulative Review

The bar graph shows the number of sandwiches ordered by 14 customers at a fast-food restaurant.

Sandwich Orders

A bar graph titled "Sandwich Orders" with the vertical axis labeled "Number of Customers" (0 to 6) and the horizontal axis labeled "Number of Sandwiches" (0 to 6). Bars: 0 sandwiches = 2 customers, 1 sandwich = 5 customers, 2 sandwiches = 2 customers, 3 sandwiches = 0, 4 sandwiches = 3 customers, 5 sandwiches = 1 customer, 6 sandwiches = 1 customer.

1. What is the mean number of sandwiches ordered by a customer?

about 2.29

2. What is the median number of sandwiches ordered by a customer?

1.5

3. What is the mode number of sandwiches ordered by a customer?

One

4. Which statistical measure would be most useful to the restaurant manager when he readies his cooks to prepare sandwiches for a long line of customers?

Because he can expect each customer to order one to 1.5 sandwiches

on average, he should tell the cooks to prepare sandwiches numbering

about one and a half times the number of customers in line.

5. Which statistical measure would be most useful to the restaurant manager in predicting the day's total revenue from sandwiches if each sandwich sells for about $2.50 and about 200 customers visited the restaurant?

Multiplying the mean number of sandwiches sold (2.29) by the selling

price ($2.50) and then, by the number of customers (200) will predict

the day's total revenue from sandwiches.

Calculator Problem

You can use a graphing calculator to find the *line of best fit*. The calculator uses a statistical process to determine the equation of the line that best represents the data.

The data in the table shows the number of bags of gummi-slugs consumed by Ferd during a movie of the given length.

Time	Bags
1	2
1.5	2
1.5	4
1.75	4
2	5.5
2.25	5
2.5	7.5
3.5	7

Begin by entering the data into the calculator. In the (STAT) menu, select **1:Edit…**. Enter the Time as List 1 (L_1) and the Bags as List 2 (L_2). See Figure 1.

Next, select (2nd) (Y=) and then **1:Plot1…Off** and turn it on. Scroll through the icons to select the scatter plot. Be sure the **Xlist** is L_1, and the **Ylist** is L_2. Before graphing, adjust the viewing (WINDOW) so that the domain is $0 \le x \le 8$, and the range is $0 \le y \le 8$. Press (GRAPH) to show the scatter plot. See Figure 2.

To find the line of best fit, press the (STAT) button and scroll to the **CALC** tab. Select **4:LinReg(ax+b)**. Press (ENTER) twice. The graphing calculator produces a line of best fit with equation $y = 2.303030303x + 0.189393939$. See Figure 3.
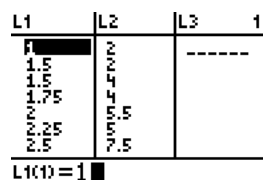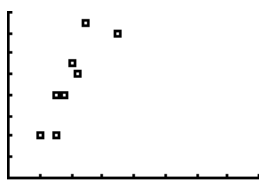
Figure 1

Figure 2

Figure 3

Computer/Spreadsheet Problem

An automated spreadsheet program, like Microsoft® Excel, can also be used to find the line of best fit.

Begin by entering ordered pairs so that *x*-values appear in column A and *y*-values in column B. Then, on the Insert menu, select Chart, scroll down to the scatter plot image, and select it. Click Finish. See Figure 4. From the Chart menu, select Add Trendline. Select the Linear Regression type and select "Display Equation on Chart" from the Options menu. Click OK. The graph now shows the scatter plot, including the line of best fit and its equation. (Figure 5)

Note that the TI-83 and Microsoft® Excel produced the same line of best fit for this scatter plot.
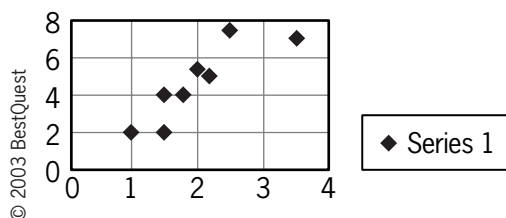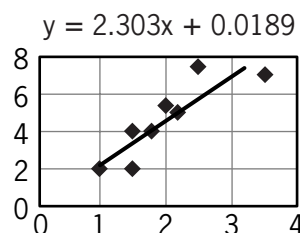
Figure 4

Figure 5

$y = 2.303x + 0.0189$

Use a graphing calculator or spreadsheet to find the equation of the line of best fit for each set of data.

1.   **Bug Juice Consumption**

Time	Ounces
11	20
1.5	22
1.5	18
1.75	20
2	22
2.25	25
2.5	30
3.5	32

y ≈ 5.88x + 11.87

2.   **Number of Previews Shown**

Time	Previews
1	5
1.5	4
1.5	5
1.75	6
2	4
2.25	3
2.5	2
3.5	2

y ≈ –1.52x + 6.91

3.   **Longest Run vs. Maximum Number of Push-Ups**

Run (mi)	Push-Ups
0.75	8
1.25	10
1.5	10
2.25	18
2.5	20
2.75	22
3.25	25
3.5	28

y ≈ 7.93x + 0.28

4.   **Roller Coaster Line Length vs. Rider Satisfaction**

People in line	Rider Satisfaction
12	90
15	95
24	84
36	82
55	80
78	78
100	75
115	73
118	70
120	73

y ≈ –0.17x + 91.35

Possible Journal Answers

1. Look at the scatter plot to determine whether or not the points generally lie in a line. If so, determine whether the line rises or falls from left to right. If the line has positive slope, the variables have a positive correlation; if the line has negative slope, the variables have a negative correlation. If the points lie in a random pattern, there is no correlation.

2. A line of fit is a line that describes the trend of the data. It represents the data without describing each point exactly. To find a line of fit, one method is to draw a line that shows the general trend of the data, with about half the data points above the line and half below it.

3. Arnold's line of fit has half the data points above and half below; however, the line does not represent the trend of the data well. It is not steep enough. Arnold's is not a good line of fit.

4. The scatter plot would show points indicating a trend that falls from left to right. As the temperature increases, the number of cups of coffee sold decreases. There is a negative correlation between temperature and number of cups of coffee sold.