Lesson 11: Using Linear Models in a Data Context

## Classwork

Exercise 1

Old Faithful is a geyser in Yellowstone National Park. The following table offers some rough estimates of the length of an eruption (in minutes) and the amount of water (in gallons) in that eruption.

Length (min.)	1.5	2	3	4.5
Amount of water (gal.)	3,700	4,100	6,450	8,400

a. Chang wants to predict the amount of water in an eruption based on the length of the eruption. What should he use as the dependent variable? Why?

b. Which of the following two scatter plots should Chang use to build his prediction model? Explain.
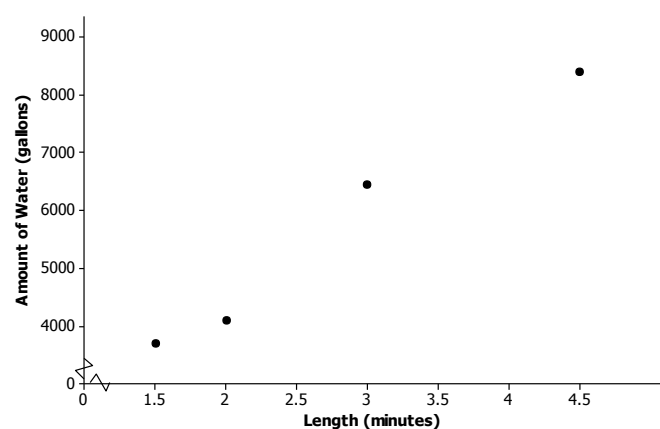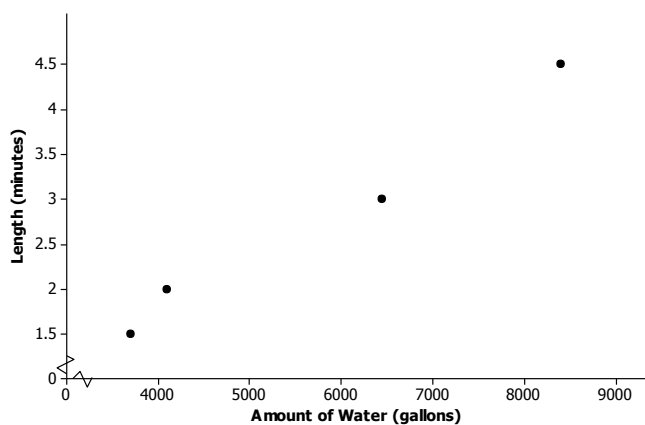
c. Suppose that Chang believes the variables to be linearly related. Use the *first* and *last* data points in the table to create a linear prediction model.

d. A friend of Chang's told him that Old Faithful produces about 3,000 gallons of water for every minute that it erupts. Does the linear model from part (c) support what Chang's friend said? Explain.

e. Using the linear model from part (c), does it make sense to interpret the y-intercept in the context of this problem? Explain.

Exercise 2

The following table gives the times of the gold, silver, and bronze medal winners for the men's 100 meter race (in seconds) for the past 10 Olympic Games.

Year	2012	2008	2004	2000	1996	1992	1988	1984	1980	1976
Gold	9.63	9.69	9.85	9.87	9.84	9.96	9.92	9.99	10.25	10.06
Silver	9.75	9.89	9.86	9.99	9.89	10.02	9.97	10.19	10.25	10.07
Bronze	9.79	9.91	9.87	10.04	9.90	10.04	9.99	10.22	10.39	10.14
Mean time	9.72	9.83	9.86	9.97	9.88	10.01	9.96	10.13	10.30	10.09

a. If you wanted to describe how mean times change over the years, which variable would you use as the independent variable, and which would you use as the dependent variable?

b. Draw a scatter plot to determine if the relationship between mean time and year appears to be linear. Comment on any trend or pattern that you see in the scatter plot.

c. One reasonable line goes through the 1992 and 2004 data. Find the equation of that line.

d. Before he saw these data, Chang guessed that the mean time of the three Olympic medal winners decreased by about 0.05 seconds from one Olympic Games to the next. Does the prediction model you found in part (c) support his guess? Explain.

e. If the trend continues, what mean race time would you predict for the gold, silver, and bronze medal winners in the 2016 Olympic Games? Explain how you got this prediction.

f. The data point $(1980, 10.3)$ appears to have an unusually high value for the mean time (10.3). Using your library or the Internet, see if you can find a possible explanation for why that might have happened.

> **Lesson Summary**
>
> In the real world, it is rare that two numerical variables are exactly linearly related. If the data are roughly linearly related, then a line can be drawn through the data. This line can then be used to make predictions and to answer questions. For now, the line is informally drawn, but in later grades you will see more formal methods for determining a best-fitting line.

Problem Set

1. From the United States Bureau of Census website, the population sizes (in millions of people) in the United States for census years 1790–2010 are as follows.

Year	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	
Population size	3.9	5.3	7.2	9.6	12.9	17.1	23.2	31.4	38.6	50.2	63.0	
Year	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
Population size	76.2	92.2	106.0	123.2	132.2	151.3	179.3	203.3	226.5	248.7	281.4	308.7

a. If you wanted to be able to predict population size in a given year, which variable would be the independent variable and which would be the dependent variable?

b. Draw a scatter plot. Does the relationship between year and population size appear to be linear?

c. Consider the data only from 1950 to 2010. Does the relationship between year and population size for these years appear to be linear?

d. One line that could be used to model the relationship between year and population size for the data from 1950 to 2010 is $y = -4,875.021 + 2.578x$. Suppose that a sociologist believes that there will be negative consequences if population size in the United States increases by more than $2\frac{3}{4}$ million people annually. Should she be concerned? Explain your reasoning.

e. Assuming that the linear pattern continues, use the line given in part (d) to predict the size of the population in the United States in the next census.

2. In search of a topic for his science class project, Bill saw an interesting YouTube video in which dropping mint candies into bottles of a soda pop caused the pop to spurt immediately from the bottle. He wondered if the height of the spurt was linearly related to the number of mint candies that were used. He collected data using 1, 3, 5, and 10 mint candies. Then he used two-liter bottles of a diet soda and measured the height of the spurt in centimeters. He tried each quantity of mint candies three times. His data are in the following table.

Number of mint candies	1	1	1	3	3	3	5	5	5	10	10	10
Height of spurt (cm)	40	35	30	110	105	90	170	160	180	400	390	420

a. Identify which variable is the independent variable and which is the dependent variable.

b. Draw a scatter plot that could be used to determine if the relationship between height of spurt and number of mint candies appears to be linear.

c. Bill sees a slight curvature in the scatter plot, but he thinks that the relationship between number of mint candies and the height of the spurt appears close enough to being linear and proceeds to draw a line. His eye-ball line goes through the mean of the three heights for 3 mint candies and the mean of the three heights for 10 candies. Bill calculates the equation of his eye-ball to be

$$y = -27.617 + (43.095)x,$$

where the height of the spurt (y) in centimeters is based on the number of mint candies (x).

Do you agree with this calculation? He rounded all of his calculations to three decimal places. Show your work.

d. Interpret in words the slope and intercept for Bill's line in the context of this problem. Does interpreting the intercept make sense in this context? Explain.

e. If the linear trend continues for greater numbers of mint candies, what would you predict the height of the spurt will be if 15 mint candies are used?