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 more formal methods for determining a best-fitting line are presented.

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 $y = -4875.021 + 2.578x$. 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 soda 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 (centimeters)	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 whether 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 the number of mint candies and the height of the spurt appears close enough to being linear, and he proceeds to draw a line. His eyeballed line goes through the mean of the three heights for three mint candies and the mean of the three heights for 10 candies. Bill calculates the equation of his eyeballed line 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. In the context of this problem, interpret in words the slope and intercept for Bill's line. Does interpreting the intercept make sense in this context? Explain.

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