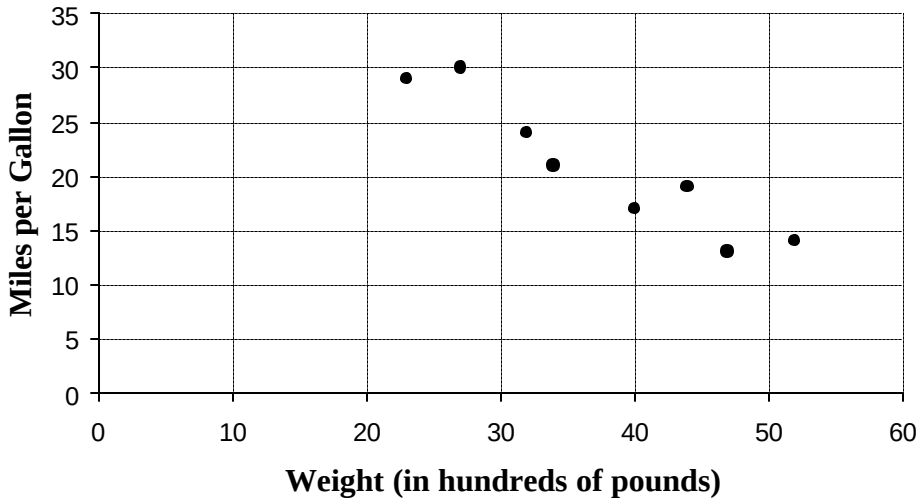
TOPIC 13: Linear Regression

Do heavier cars use more gasoline? If you know the weight of a car, how can you predict its miles per gallon? To answer these questions we need to collect some data. The data in the table below are based on information taken from *Consumer Reports.* Eight cars were randomly selected and we let x be the weight of the car (in hundreds of pounds), and let y be the miles per gallon (mpg).

x	27	44	32	47	23	40	34	52
y	30	19	24	13	29	17	21	14

1. What factors, other than the weight of a car, can affect gasoline usage?

To help us analyze the data in the table, we create a ***scatter plot*** by plotting the ordered pairs (x, y) using our graphing calculator.



Notice the points in the scatter plot, although not on a straight line, show a linear pattern. We see that the points slope *downward to the right*, which indicates that as the weight of the car *x increases*, the miles per gallon *y tends to decrease*. This is an example of a ***negative association*** between x and y. Similarly, there is a ***positive association*** between x and y if, when *x increases*, *y tends to increase*. In this case we would see a linear pattern in the scatter plot with the points sloping *upward to the right*.

For exercises 2 and 3, use the data in the table below.

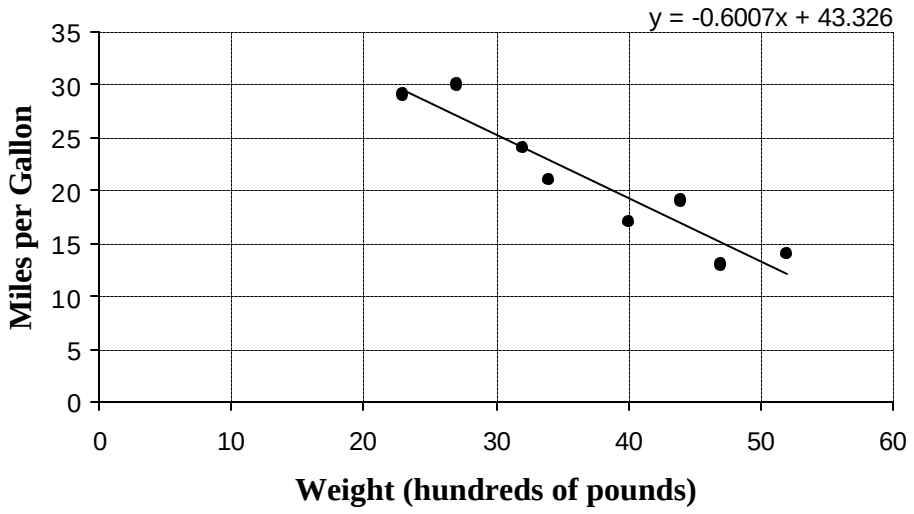
A car dealership has been using 1-minute spot ads on a local TV station. During a 10-week period, the dealer kept a weekly record of the number of TV ads versus the number of cars sold. Let x be the number of TV ads in a week, and let y be the number of cars sold that week.

x	6	20	0	14	25	16	28	18	10	8
y	15	31	10	16	28	20	40	25	12	15

2. Make a scatter plot of the data.

3. Does the scatter plot show a positive association, a negative association, or neither? Explain.

Now look again at the weight/miles per gallon data above. The points in the scatter plot show a linear pattern, so we will try to find a straight line that "fits" the data. Of course there are many possible lines we can draw on the scatter plot, but in some sense the "best" line should be the one that comes closest to each of the points. The line we will use is called the ***least-squares regression line***. We can use our graphing calculator to find the equation of the regression line and graph it on the scatter plot.



Notice that some points in the scatter plot are above the regression line, and some are below it. However, all the points are reasonably "close" to the line. The equation of the regressions line is $y = -0.6007x + 43.326$ (shown in the upper-right corner of the scatter plot).

The equation $y = -0.6007x + 43.326$ indicates that the slope of the line is –0.6007. What does this mean? Recall that the slope is the change in y when x increases by 1. Therefore, there is a 0.6007 *decrease* in miles per gallon when the weight of the care increases by 100 pounds.

In the following example we see how to use the equation of the regression line to make predictions about gasoline usage.

Example 1: Suppose you buy a new car that weighs 3000 pounds. How many miles per gallon would you expect to get?

If we look at the regression line on the scatter plot, we see that a 3000-pound car ($x = 30$) would get slightly more than 25 mpg. To get a more accurate estimate of gas usage, we can substitute $x = 30$ into the equation of the regression line:

$$y = -0.6007x + 43.326$$
$$y = -0.6007(30) + 43.326$$
$$y = -18.021 + 43.326$$
$$y = 25.305$$

You would expect a 3000-pound car to get about 25.3 miles per gallon.

Note that the y-intercept of the regression line $y = -0.6007x + 43.326$ is (0, 43.326). This means we would expect 43.326 miles per gallon for a car weighing 0 pounds! Clearly, this makes no sense in this context.

For exercises 4-7, refer to the data for exercises 2 and 3.

4. Find the least-squares regression line and graph it on your scatter plot.

5. What is the slope of the regression line? What does it mean?

6. What is the y-intercept of the regression line? What does it mean? Does it make sense?

7. The car dealer can afford 12 ads per week. At that level of advertisement, how many cars can he expect to sell each week?