Approximate Linear Relationships

In the real world, rarely do things follow trends perfectly. When the trend is expected to behave linearly, or when inspection suggests the trend is behaving linearly, it is often desirable to find an equation to approximate the data. This equation will help us understand the behavior of the data and allow us to use the linear model to make predictions about the data.

To find the equation, we start with a scatter plot of all the ordered pairs of the data points that were collected from measurements or experiments. This means that a scatter plot is a relation, and not necessarily a function. We should not expect that all the points will fit perfectly on a straight line. Rather the points will be scattered about a straight line. There are many reasons why the data does not fall perfectly on a line such as measurement error and outliers.

Measurement error is always present as no measurement device is perfectly accurate.

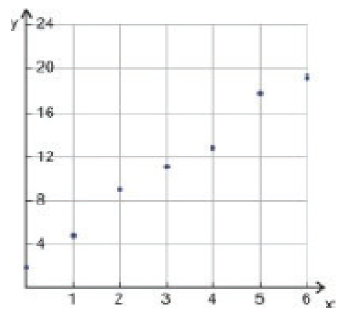
An outlier is a data point that does not fit with the general pattern of the data. It is a statistical fluctuation.

To make a scatter plot, simply plot the ordered pairs from the list of given ordered pairs or from a given table.

Example 1:

Make a scatter plot of the following order pairs:

(0, 2), (1, 4.5), (2, 9), (3, 11), (4, 13), (5, 18), (6, 19.5)

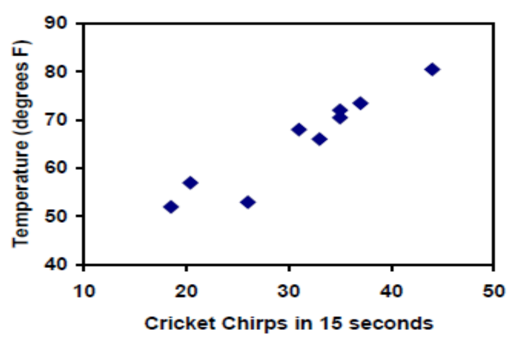


Notice that the relationship looks linear, even though the points don't fit perfectly on a straight line.

Example 2:

The table below shows the number of cricket chirps in 15 seconds, and the air temperature, in degrees Fahrenheit. Plot this data, and determine whether the data appears to be linearly related.

Chirps	44	35	20.4	33	31	35	18.5	37	26
Temp.	80.5	70.5	57	66	68	72	52	73.5	53



The trend appears roughly linear.

In both examples above, the points look like they might be part of a straight line, although they would not fit perfectly on a straight line. To draw the line that best fits the data, we use a line that is closest to most of the points on the graph. Remember that there may be outliers.

Once the line that best fits the data is drawn, the equation that approximates the linear relationship can easily be found. This process is called linear regression.

The steps used to perform linear regression:

1. Select two data points that are very close to the line of best fit.
2. Find the slope using those two points.
3. Use the point-slope form to write the equation.

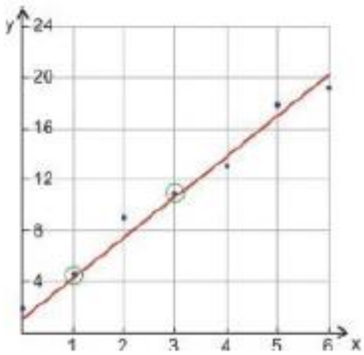
This linear equation can then be used to approximate behaviors and predictions about the trend. The equation is the best guess as to how the relationship will behave outside of the values of given data. There is a difference, though, between making predictions inside the data and outside the data.

Interpolation is predicting a value inside the data.

Extrapolation is predicting a value outside the data (future value).

Example 3:
Using the ordered pairs and graph in Example 1, find a linear function of the line that best fits the data.

Draw a line closest to most of the points



Two points very close to the line of best fit are (1, 4.5) and (3, 11)

Find the slope: $m = \frac{11-4.5}{3-1} = \frac{6.5}{2} = 3.25$

Find the equation: $y - 11 = 3.25(x - 3)$

$$y - 11 = 3.25x - 9.75$$
$$\underline{\quad + 11 \qquad\qquad + 11\quad}$$
$$y = 3.25x + 1.25$$

The function: $f(x) = 3.25x + 1.25$

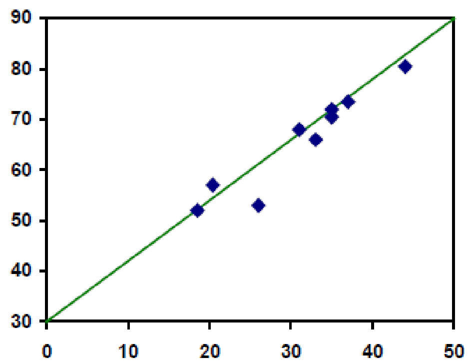
Example 4:
Using the table of values and graph in Example 2, find the following:

a. A linear function of the line that best fits the data

b. Would predicting the temperature when the crickets are chirping 30 times in 15 seconds be interpolation or extrapolation? Make the prediction. Is it reasonable?

c. Would predicting the number of chirps crickets will make at 40 degrees be interpolation or extrapolation? Make the prediction. Is it reasonable?

a. Draw a line closest to most of the points



Two points on the line of best fit are (18.5, 52) and (35, 72)

Find the slope: $m = \frac{72-52}{35-18.5} = \frac{20}{16.5} \approx 1.2$

Find the equation: $y - 72 = 1.2(x - 35)$

$y - 72 = 1.2x - 42$

$+72 \qquad +72$

$y = 1.2x + 30$

The function: $f(x) = 1.2x + 30$

b. A prediction of 30 chirps per 15 seconds is inside the data, so it would be interpolation. Using the model:

$$\begin{aligned} f(30) &= 1.2(30) + 30 \\ &= 36 + 30 \\ &= 66 \text{ degrees} \end{aligned}$$

Based on the data, this value seems reasonable

c. A prediction at 40 degrees is outside the data, so it would be extrapolation. Using the model:

$$40 = 1.2x + 30$$

$$\underline{-30 \qquad\qquad -30}$$

$$\frac{10}{1.2} = \frac{1.2x}{1.2}$$

$x \approx 8.33$ chirps in 15 seconds

While this might be possible, we have no reason to believe the model is valid outside the given data.