

Scatter plots and lines of best fit :

A scatter plot displays two sets of data on one graph using points showing the relationship, or correlation, between the two sets of data. A line of best fit shows the trend in the data, if there is a trend.

Example:

The students on the track team recorded the number of practice miles each ran during the week and their time in that week's mile race. The data is shown below.

Number of Practice Miles Ran	12	12	13	14	15	17	18	19
Mile Race Time (minutes)	8.25	7.5	6.6	7	6	5.8	5.5	5

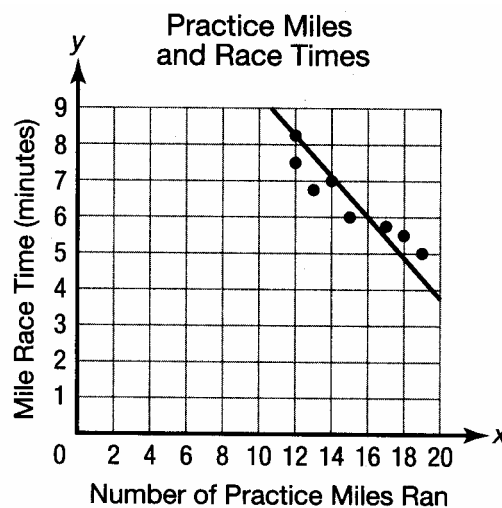
Display the data in a scatter plot and draw the line of best fit, if possible.

First, plot the points on a coordinate grid. Set up the scatter plot on the coordinate grid. Make sure you label the axes and include a title.

Next, List the point in (x, y) format with the number of practice miles ran as the x -value and time as the y -value. Plot the points. The points would look like:

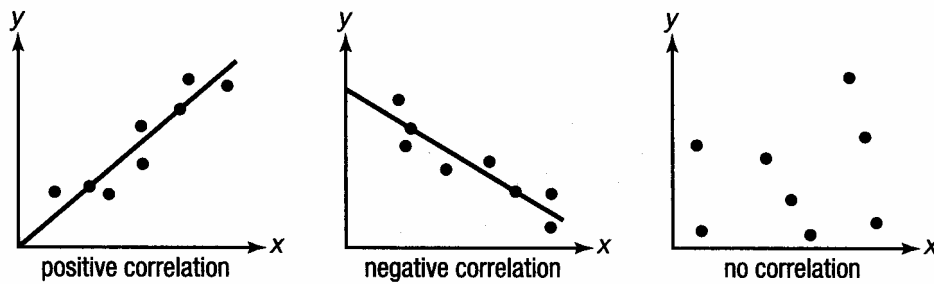
$(12, 8.25)$ $(12, 7.5)$ $(13, 6.6)$ $(14, 7)$ $(15, 6)$ $(17, 5.8)$ $(18, 5.5)$ $(19, 5)$

Then, draw a line of best fit, if possible. To draw the line of best fit, draw a line that has approximately an equal number of points above and below it. Do not "connect the dots." The plot will look like this:



Determine whether data has a positive, negative or no correlation :

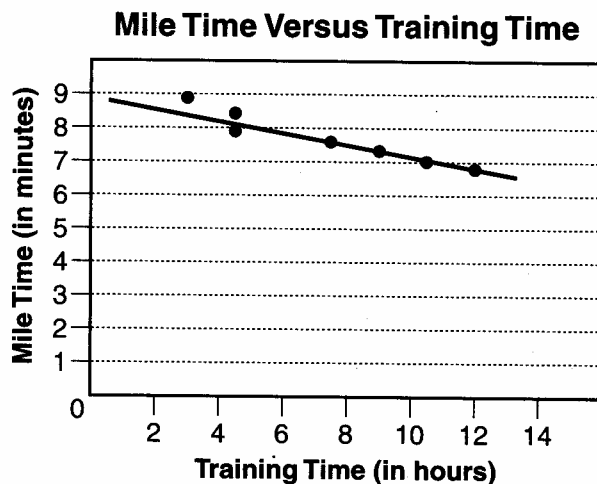
When two data sets are related, we say they are correlated. A direct relationship represents a positive correlation (both sets of data are increasing together). An inverse relationship represents a negative correlation (one set of data is increasing while the other set of data is decreasing). The closer the data points come to forming a straight line, the more strongly they are correlated. When the data points lie in a straight line, the data sets are perfectly correlated. Two data sets that have no relationship have no correlation. Study the following three graphs.



When data sets show a positive or negative correlation, a line of best fit can be drawn to approximate missing data. The line of best fit has close to the same number of data points above it as it does below it.

Example:

The following scatter plot shows the number of hours Stuart trained for each of his last 7 weekly timed miles and the associated mile time for each week. The line of best fit has been drawn in the scatter plot for you.



The scatter plot shows a negative correlation between training time and mile time. The line of best fit has been drawn through the data points in order to make predictions about the data values. Looking at the line of best fit, it appears that if Stuart trains for 14 hours he will run his timed mile in about 6.5 minutes. (This is a prediction, since the table does not include a training time of 14 hours.)