Six Sigma Tools – Scatter Diagrams

Gerald M. Taylor
Managing Consultant
The Performance Management Group LLC

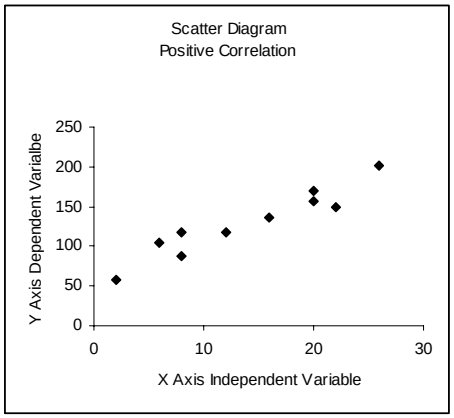
The Scatter Diagram (Scatter Plot) is a tool used in correlation analysis to graphically illustrate possible relationships between 2 quantitative variables.

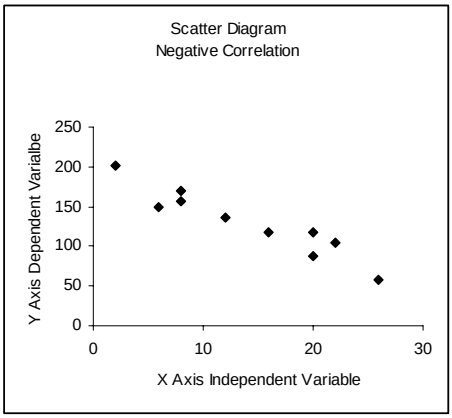
There are three main components of a scatter diagram; the Y axis, the X axis and the Plane.

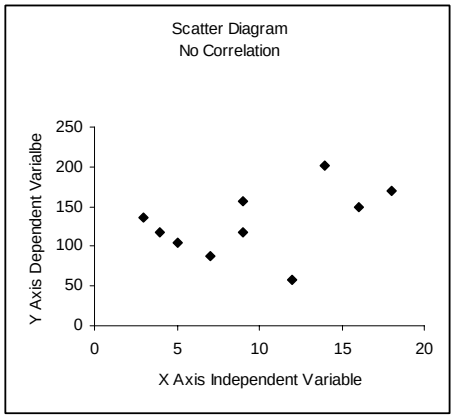
- On the "*Y axis"* is plotted the response variable. The response variable, also known as the dependent variable, is the variable that is acted upon and is the object of correlation analysis.
- On the "*X axis"* is plotted the explanatory variable. The explanatory variable, also known as the independent variable, is the variable that affects the dependent variable and helps explain the response that is plotted.
- The "*plane"* is where the plots of all of the observations of the analysis are made.

Six Sigma professionals often use scatter diagrams to investigate and examine a situation for possible *cause* and *effect* relationships. They also use scatter diagrams test if there is a correlation between one cause and another, or to evaluate whether correlations exist between one cause and multiple other causes.

Below are examples of the types of correlations usually associated with scatter diagrams.







Positive Correlation:

Is determined when an increase in Y depends on an increase in X. If X is controlled, Y will be naturally.

Some examples include:
- Height Vs Weight
- Training Vs Performance
- Speed Vs Distance

Negative Correlation:

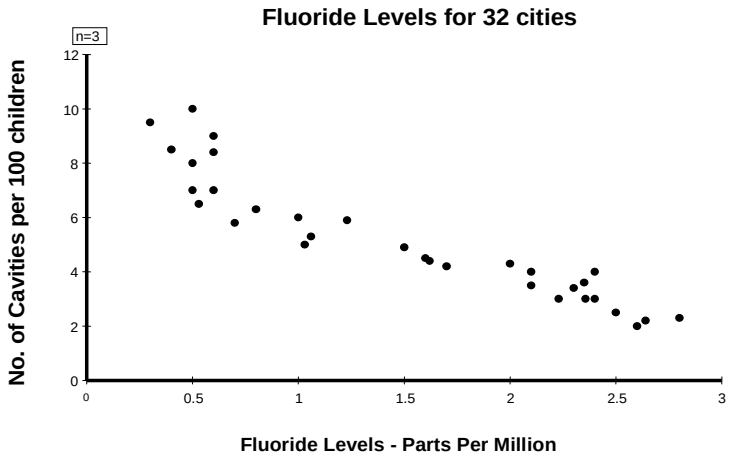
Is determined when an increase in X will cause a decrease in Y.

No Correlation:

No apparent correlation exists between X or Y.

Scatter Diagram – An example (2)

Below is an example of a scatter diagram of fluoride levels in 32 cities. The response variable being studied is the number of cavities per 100 children. That variable is being correlated against fluoride levels in the water of different cities. The results of the analysis seem to indicate that children with fewer cavities reside in cities with greater concentrations fluoride in their water. This type of relationship that can be defined as a negative correlation, where an increase in fluoride concentration in water (X) seems to cause a decrease in the number of cavities children experience (Y).



How to construct a Scatter Diagram

1. Collect data in paired samples. Create a data sheet as shown below.

Number of Observations	Fluoride Levels (PPM)	Number of Cavities per 100 Children.
1	0.5	10
2	0.3	9.5
3	0.6	9
4	0.5	8
5	0.4	8.5
32	2.8	2.3

2. Draw both vertical and horizontal axes of the Scatter Diagram and plot the explanatory (independent) variable on the horizontal (X) axis and the response (dependent variable) on the vertical (Y) axis. Scale your chart accordingly.

3. Plot the data on the plane of Scatter Diagram.

To build a scatter diagram in Microsoft Excel, the following steps are helpful:

1. Create the columns of your data sheet, with appropriate headings, as shown above.
- Place column 1 as your subject category.
- Place column 2 as your explanatory (X) variable.
- Place column 3 as your response (Y) variable.

2. Highlight the data (numbers) in columns two and three.

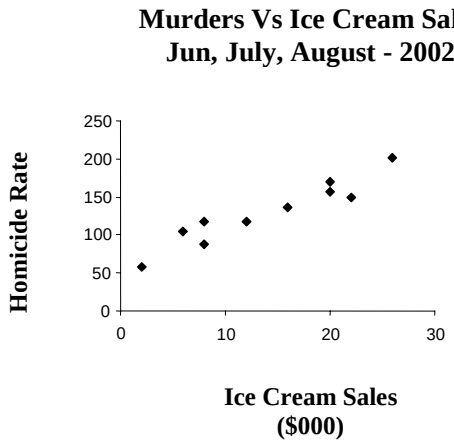
Special note: Do not sort your data – leave your data and values in its original form.

3. Select "*Insert*" from your menu then select "*Chart*". Your chart wizard will appear.

4. Under the *Standard Types* tab, Select the "*XY (Scatter)*" icon and press next.

5. Follow the appropriate sequence to properly title and label your Scatter Diagram.

Effective use of scatter diagrams can make it clear whether cause and effect relationships exist and what their relative strengths are. It is important to understand, however, that a graphical representation of a positive correlation does not necessarily mean that the correlation exists. Case in point: The scatter diagram below examines the possible correlation between ice cream sales and murders in a major US city. It is unlikely that increased ice cream sales explain the up tick in murders over the 3 months studied. As a caveat, this diagram underscores the concept that scatter plots test for "possible" correlations between two variables and that the use of the scatter plot is merely a beginning to understanding the nature of a relationship between two variables.



Scatter Diagramming – An Exercise

The Travel Time Scenario[1]

Suppose your company has just implemented a flextime work policy. This policy would allow you to come to work anytime between 0730 hrs and 0930 hrs and leave eight and one half hours later. You want to pick your work hours to minimize your drive time. Over the next month, you leave your home at various times between 0700 hrs and 0900 hrs.

You record your data as follows:

Trip	Departure Time	Travel Time (Min)
1	7:05	16
2	7:10	18
3	7:15	20
4	7:25	17
5	7:35	21
6	7:45	24
7	7:50	22
8	7:55	25
9	8:01	25
10	8:05	28
11	8:15	35
12	8:20	30
13	8:24	46
14	8:28	24
15	8:35	26
16	8:37	28
17	8:40	24
18	8:46	21
19	8:50	18
20	08:58	16

Construct your scatter diagram and analyze it. What conclusions can you draw from it. For the correct Scatter Diagram to compare your results to, email TPMG Educational Services at performance1@helpingmakeithappen.com