# *Scatter Diagrams*

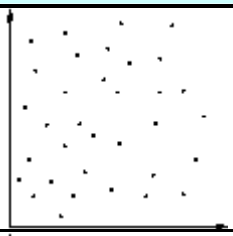
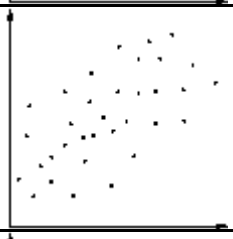
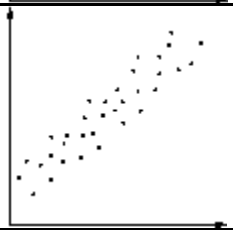
## What is a scatter diagram?

The scatter diagram is one of the tools of quality. A scatter diagram is a graphical technique used to analyze the relationship between two variables. It shows whether or not there is correlation between two variables. Correlation refers to the measure of the relationship between two sets of numbers or variables. Two sets of data are plotted on a graph, with the y-axis being used for the variable to be predicted and the x-axis being used for the variable to make the prediction. The graph will show possible relationships (although two variables might appear to be related, they might not be: Those who know most about the variables must make that evaluation). However, correlation does not necessarily mean a direct cause and effect relationship. If it appears that values for one of the variables can be predicted based on the value of the other variable, then there is correlation.

## Interpreting a scatter diagram:

Scatter diagrams can show different types of correlation. The tables below, taken from the following website, http://syque.com/quality_tools/toolbook/Scatter/how.htm, show some of the different types of correlation and how to interpret them.

**Table 1. Degrees of Correlation**

| Scatter Diagram | Degree of Correlation | Interpretation |
|---|---|---|
|  | None | No relationship can be seen. The 'effect' is not related to the 'cause' in any way. |
|  | Low | A vague relationship is seen. The 'cause' may affect the 'effect', but only distantly. There are either more immediate causes to be found or there is significant variation in the 'effect'. |
|  | High | The points are grouped into a clear linear shape. It is probable that the 'cause' is directly related to the 'effect'. Hence, any change in 'cause' will result in a reasonably predictable change in 'effect'. |

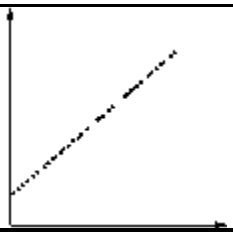
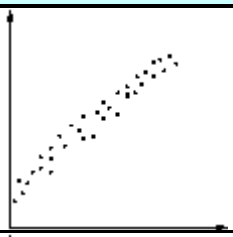
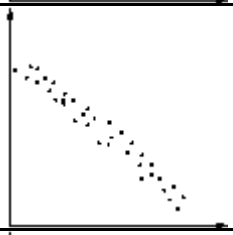
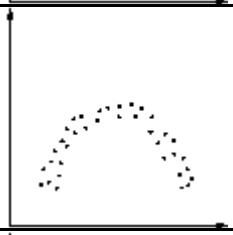
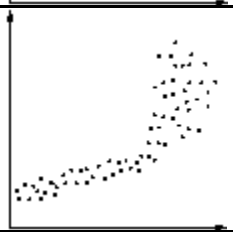
| | | |
|---|---|---|
|  | Perfect | All points lie on a line (which is usually straight). Given any 'cause' value, the corresponding 'effect' value can be predicted with complete certainty. |

**Table 2. Types of Correlation**

| Scatter Diagram | Types of Correlation | Interpretation |
|---|---|---|
|  | Positive | Straight line, sloping up from left to right. Increasing the value of the 'cause' results in a proportionate increase in the value of the 'effect'. |
|  | Negative | Straight line, sloping down from left to right. Increasing the value of the 'cause' results in a proportionate decrease in the value of the 'effect'. |
|  | Curved | Various curves, typically U- or S-shaped. Changing the value of the 'cause' results in the 'effect' changing differently, depending on the position on the curve. |
|  | Part linear | Part of the diagram is a straight line (sloping up or down). May be due to breakdown or overload of 'effect', or is a curve with a part that approximates to a straight line (which may be treated as such). |

## Things to consider when using scatter diagrams:

Some additional factors should be considered when using the scatter diagram. First, not all relationships between variables are linear. The second consideration is that the visible slope of the line does not provide any information about the strength of correlation, since the scales of the graph can be expanded or compressed on either

axis. The third consideration, and probably the most important, is that a direct or strong correlation does not necessarily imply a cause-and-effect relationship. If a scatter diagram shows signs of correlation, investigate further for confirmation. Correspondingly, a diagram that indicates no correlation, should not lead you to dismiss your suspicions. An example is that the volume of ice cream sold per day is strongly correlated to the daily number of fatalities by drowning. Obviously, neither of these variables is the cause of the other, and the most likely explanation is that each of these variables is a result of and therefore strongly correlated to a third variable, such as the outside temperature.

**Steps in using a scatter diagrams:**

1. Select and define the two variables to be analyzed in the individual chart.
2. Measure the two variables, if data has not already been collected.
3. The set of data pairs should consist of the at least thirty, but preferably more like one hundred data pairs.
4. Design the chart by placing the independent variable on the horizontal axis. The independent variable is the factor believed to be governing relationship between the two variables.
5. On the vertical axis, insert the dependent variable—that is, the factor believed to change in proportion to the independent variable.
6. Plot the data pairs themselves in the chart area.
7. Examine the completed chart, looking for patterns that indicate a connection between the two variables.
8. If correlation patterns are identified, investigate any third variable involvement before drawing definite conclusions.

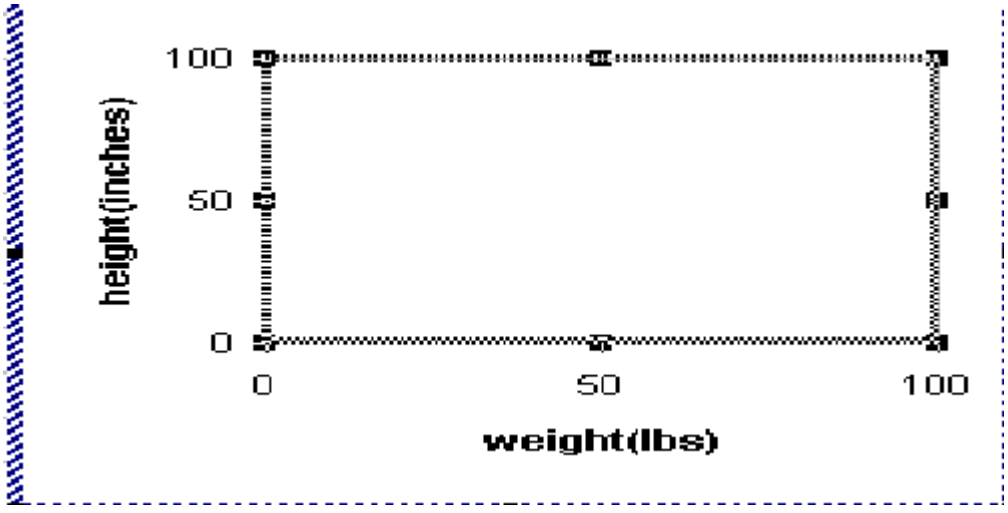
**An example (taken from the following website - http://deming.eng.clemson.edu/pub/tutorials/qctools/scatm.htm) using a scatter diagram:**

Situation: The new commissioner of the American Basketball League wants to construct a scatter diagram to find out if there is any relationship between a player's weight and her height. How should she go about making her scatter diagram?

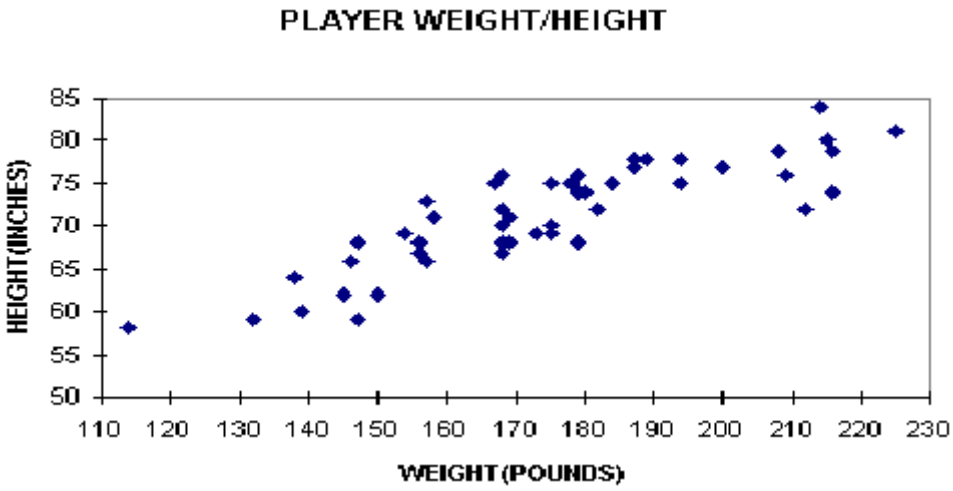
1. Collect the data (Remember to use 30-100 paired samples).

| PLAYER | WEIGHT | HEIGHT | PLAYER | WEIGHT | HEIGHT |
|---|---|---|---|---|---|
| 1 | 132 | 59 | 26 | 154 | 69 |
| 2 | 114 | 58 | 27 | 189 | 78 |
| 3 | 187 | 78 | 28 | 175 | 75 |
| 4 | 214 | 84 | 29 | 146 | 66 |
| 5 | 158 | 71 | 30 | 138 | 64 |
| 6 | 212 | 72 | 31 | 178 | 75 |
| 7 | 175 | 69 | 32 | 208 | 79 |
| 8 | 147 | 59 | 33 | 145 | 62 |
| 9 | 173 | 69 | 34 | 168 | 68 |
| 10 | 182 | 72 | 35 | 225 | 81 |
| 11 | 194 | 78 | 36 | 157 | 66 |
| 12 | 215 | 80 | 37 | 168 | 67 |
| 13 | 180 | 74 | 38 | 179 | 76 |
| 14 | 147 | 68 | 39 | 139 | 60 |
| 15 | 168 | 76 | 40 | 168 | 72 |
| 16 | 157 | 73 | 41 | 179 | 74 |
| 17 | 168 | 70 | 42 | 156 | 68 |
| 18 | 179 | 68 | 43 | 184 | 75 |
| 19 | 216 | 74 | 44 | 169 | 71 |
| 20 | 200 | 77 | 45 | 175 | 70 |
| 21 | 194 | 75 | 46 | 168 | 72 |
| 22 | 147 | 68 | 47 | 167 | 75 |
| 23 | 156 | 67 | 48 | 216 | 79 |

2. Draw and label your x and y axes.




3. Plot the data on the diagram.




4. Interpret your chart.

According to this scatter diagram the new commissioner was right. There does seem to be a positive correlation between a player's weight and her height. In other words, the taller a player is the more she tends to weight.