# Chapter 3

# Visual Representation of Data: Charts and Graphs for Six Sigma

# Chapter Outline

## *Describing Data using Charts and Graphs: Introduction*

The graphical techniques described in this chapter are useful in the following ways:

- The techniques will help you gain insight into the way the variable or variables seem to behave.
- The graphical techniques enable one to understand how the values of a random variable under study are distributed.
- The shapes produced using the graphical techniques help select an appropriate theoretical distribution for the random variable in question.
- The charts and graphs help us visualize the important characteristics of data which are usually not apparent from the raw data.
- Some of the visual representation of the data provide excellent means of comparing data from processes, checking the variation, and taking corrective actions when the deviation from stable conditions occur.

***In this chapter, we have presented numerous graphical techniques. We assume that you are familiar with many of them; therefore, we will not discuss the theory behind them in detail. Instead, we focus on applications. We explain how to construct these graphs and charts using the computer and explain their important characteristics.***

## *Charts and Graphs in this Chapter*

- Histograms with outline and group
- Graphical summary of data
- Histogram with normal curve
- Stem-and-leaf, box plots, and dot plots
- Character graphs (to construct stem-and-leaf, box plot and dot plot)
- Widely used plots including bar charts, pie charts, scatter plots, interval plots, and time series plots
- Empirical cumulative density function (CDF), probability plots, matrix plots, marginal plots, 3D scatter plots, 3D surface plots, contour plots, and others.

*Some Examples of Graphs & Charts useful in Six Sigma Analysis (The data files to construct the charts below are available with the book)*

## *Constructing a Default Histogram*

To construct a default histogram, follow the instructions in Table 3.1 below.

Table 3.1

| | |
|---|---|
| **CONSTRUCTING A DEFAULT HISTOGRAM** | Open the worksheet **Demand.MTW** then select<br>**Graph ➢Histogram**<br>Click on **Simple** then click **OK**.<br>Complete the dialog box by selecting or typing the response shown:<br>Graph variables: **C1** or **Demand**<br>Click on **Scale** Box<br>Click on **Y-scale type** and select **Frequency**<br>Click **OK**<br>Click **Labels**<br>In the Title box, type: **Histogram of Demand Data**<br>Click on **Data Labels** and select **Use Y-value label**<br>Click **OK**<br>Click on **Data View** and check **Bars**, then click **OK** |

You will be back to the **Histogram-simple** dialog box. Click **OK**. The histogram shown in Figure 3.1 will be displayed on the graphics window.

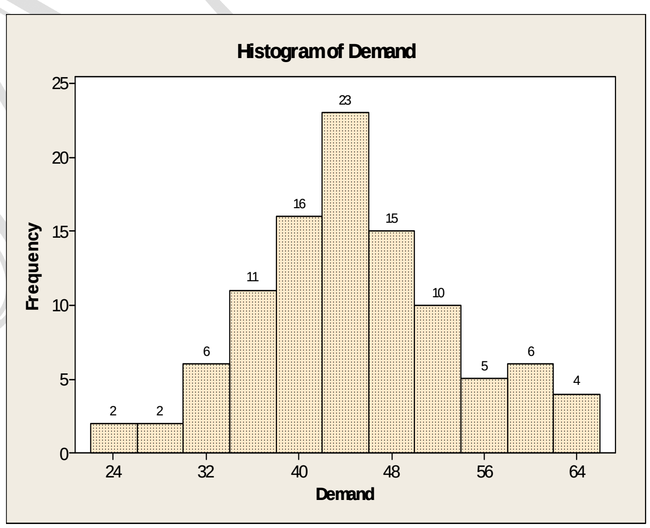



Figure 3.1: A Default Histogram

:
:
:

## *Other examples:*

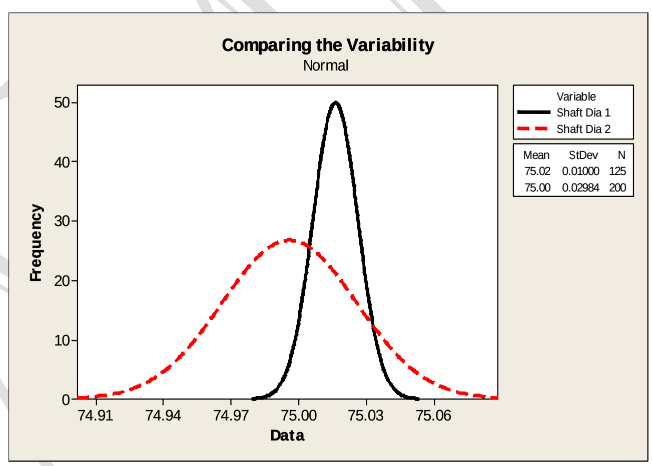
### *Histogram with Fit and Groups*

This option can be used to compare the mean and variability of two sets of data. Suppose you want to compare the variability in diameter of the shafts produced by two manufacturers. A sample of 124 shafts from manufacturer 1 and a sample of 200 shafts from manufacturer 2 were measured. The data are in the file **ShaftDia.MTW**. Table 3.6 shows the instructions.

Table 3.6

| |
|---|
| Open the worksheet **ShaftDia. MTW** |
| Select **Graph** ➢ **Histogram** |
| **:** |
| **:** |
| Click **Labels** and type **Comparing the Variability** in the **Title** box |
| Click **OK**. |

The graph in below will be displayed.




Histogram with Fit and Groups

### *Other Graph Options*

Using the command sequence, **Stat** ➢**Basic Statistics** ➢ **Display Descriptive Statistics** and selecting the **Graphs** in the **Display Descriptive Statistics** dialog box, provides options for graphs, such as, histogram, histogram of data with normal curve, individual value plot, and box plot.

### *Graphical Summary of Data*

This option provides useful statistics of the data along with graphs. To produce a graphical summary of the data, follow the steps in Table 3.8.

Table 3.8

| |
|---|
| Open the worksheet **ShaftDia.MTW**<br>Select **Stat ➢Basic Statistics ➢Graphical Summary**<br>: |

Figure 3.10 shows the graphical summary of the data.

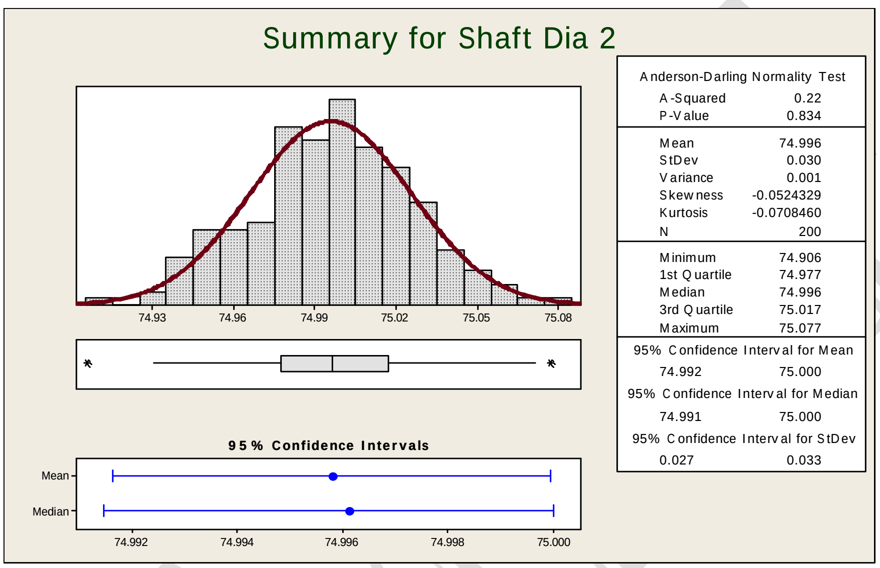



Figure 3.10: Graphical Summary of the Shaft Diameter Data

(Note: when the graph is displayed, you can double click any where on the bars and edit to change color and fill of the bars).

## *Stem-and-leaf Plots*

Stem-and-leaf plots are very efficient way of displaying data, checking the variation and shape of the distribution. Stem-and-leaf plots are obtained by dividing each data value into two parts, stem and leaf. For example, if the data are two-digit numbers, e.g., 34, 56, 67, etc., then the first number (the tens digit) is considered the stem value, and the second number (the ones digit) is considered the leaf value. Thus, in data value 56, 5 is the stem and 6 is the leaf. In a three digit data value, the first two digits are considered as the stem and the last digit as the leaf. To construct a stem-and-leaf plot, follow the steps in Table 3.9.

Table 3.9

| |
|---|
| Open worksheet **DEFECTS.MTW**<br>From the main menu, select **Graph ➢ Stem-and-Leaf**<br>For **Graph variables**, select **No. of Defects**<br>Click **OK** |

The stem-and-leaf plot of number of defects will be displayed on the session window.

The plot is shown in Figure 3.11.

```
Stem-and-Leaf Display: No. of Defects (out of 1000)

Stem-and-leaf of No. of Defects(out of 1000)  N  = 200
Leaf Unit = 1.0

 5    5  23333
 6    5  4
 12   5  666777
 26   5  88888889999999
 48   6  0000000000011111111111
 76   6  2222222333333333333333333333
(30)  6  444444444444445555555555555555
 94   6  6666666666677777777777777777
 66   6  888888888888889999999
 45   7  00000000000001111111
 25   7  2222222222333
 12   7  44444555
 4    7  667
 1    7  8
```

Figure 3.11: Stem-and-Leaf Plot of Number of Defects

## ***Another Example on Stem-and-leaf plot***

```
Stem-and-Leaf Display: Moisture Content

Stem-and-leaf of Moisture Content  N  = 55
Leaf Unit = 0.10

 1    8   0
 2    8   8
 3    9   4
 4    9   7
 9    10  02334
 16   10  5667888
 26   11  1223333444
(5)   11  56677
 24   12  24444
 19   12  5567
 15   13  2333
 11   13  578
 8    14  01
 6    14  6
 5    15  23
 3    15  56
 1    16  1
```

Stem-and-leaf Plot of Moisture Content in Samples of Clay (in percent)

## *Box-Plots*

The box-plot displays the smallest and the largest values in the data along with the three quartiles: $Q_1$, $Q_2$, and $Q_3$. The display of these five numbers (known as five measure summary) is used to study the shape of the distribution. There are different types of box plots you can do including:

(a) One Y - Simple
(b) One Y - With groups
(c) Multiple Y's - Simple
(d) Multiple Y's - With Groups

### *(b) One Y - With groups*

This plot is useful when there is one y-variable (diameter in this case) that you want to monitor by assessing several days of production. You would like to plot the box plot for each day of production. To do this plot, follow the steps in Table 3.12.

Table 3.12

| |
|---|
| Open the worksheet **DIAMETER.MTW** |
| From the main menu, select **Graph ➢Boxplot** (you can also select **Stat ➢EDA ➢Boxplot**) |
| Under **One Y** select **With Groups** |
| Click **OK** |
| Select **Diameter** for Graph Variables |
| For **Categorical variables for grouping (1-4, outermost first),** select **Day for attribute assignment** |
| **:** |
| **:** |
| Click **OK** in all the dialog boxes. |

The box-plot in Figure 3.14 will be displayed.

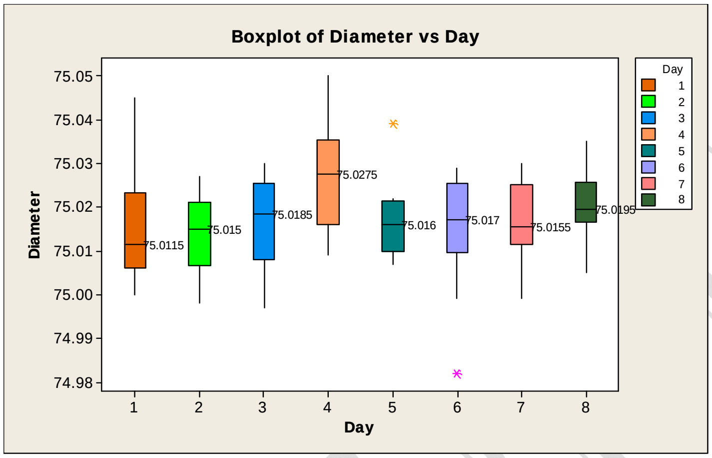



Figure 3.14: Box Plot of Diameter vs. Day

Suppose you want to check the consistency of the diameters of 5 samples with respect to three machine operators. You can construct a box plot to check the consistency. The steps are same as described for the box plot in Figure 3.15. Open worksheet **DIAMETER3.MTW**, follow the steps above but select **Operator** for **categorical variables grouping** and select **Operator** in the **Data View** box. The box plot is shown in Figure 3.17. Figures 3.16 and 3.17 are useful in checking the distribution and consistency of critical production parameter with respect to categorical variables such as, machine and operator.

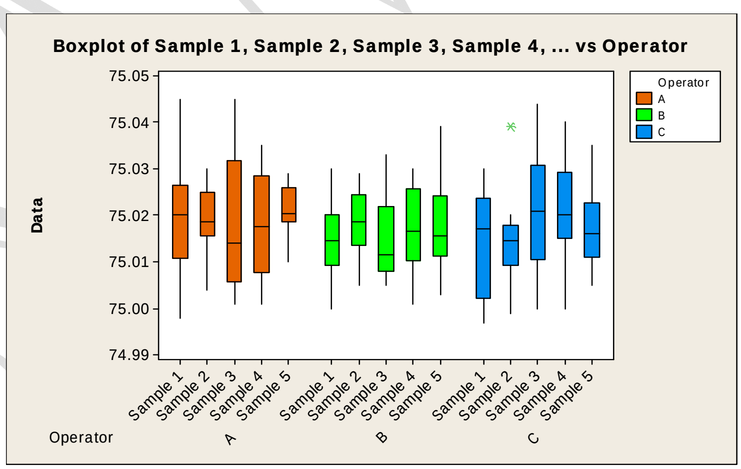



Figure 3.17: Box plot of Samples vs. Operators

## *Bar Charts*

MINITAB provides several options for bar charts including simple, clustered, and stacked bar charts. There are several variations of these charts.

**A Simple bar chart:** Suppose you want to plot monthly sales for your company. You have two columns of data; the first column is the categorical variable (month) and the second column contains the sales values. To do a bar chart, follow the steps in Table 3.26.

Table 3.26

| |
|---|
| Open the worksheet **BAR1.MTW**<br>From the main menu, select **Graph** ➢**Bar Chart**<br>Below **Bars represent**, click the down arrow and select **Values from a table**<br>Make sure **Simple** is highlighted under **One column of values**<br>Click **OK**<br>For **Graph variables**, select **Sales ($)**<br>In **Categorical variable**, select **Month**<br>Click **Labels** box then click **Titles/Footnotes** and enter a title for your graph<br>Click **Data Labels** tab<br>Click the circle next to **Use y-values labels**<br>Click **Data View**, in the **Categorical variables for attribute assignment**, type or select **Month**<br>Click **OK** in all the dialog boxes. |

The bar chart in Figure 3.29 will be displayed.

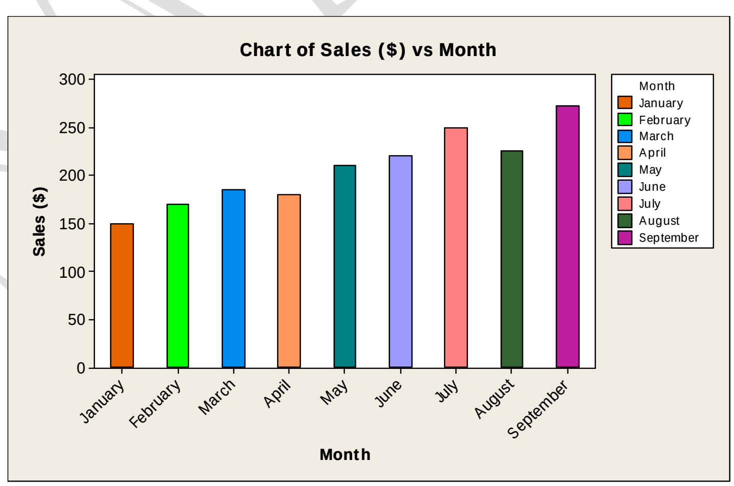



Figure 3.29: A Bar chart of Monthly Sales

## Bar Chart of Categorical Variables

The worksheet **BAR1.MTW** contains a categorical variable: Causes of Failure. This column contains various causes of failure for machined parts. We want to know the frequency of each type of failure.

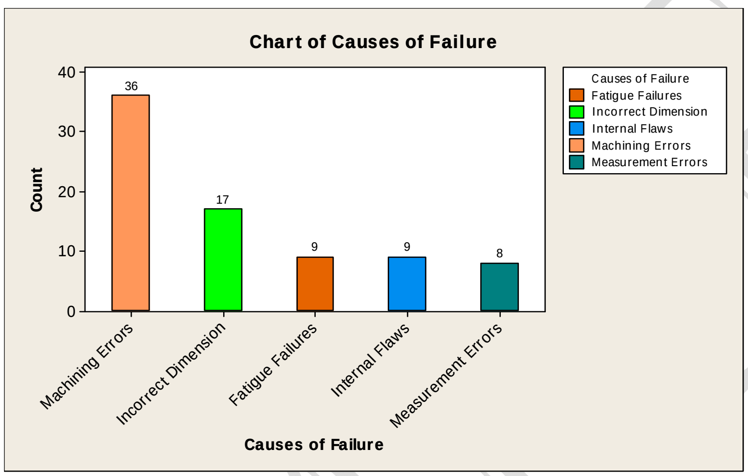



Figure 3.30: A Bar chart of Categorical Data

## A Cluster Bar Chart

Suppose you want to compare the quarterly sales for the past four years. A good way is to plot the sales of each quarter using a bar chart. We can use the cluster option to group the four quarters of each of the four years.

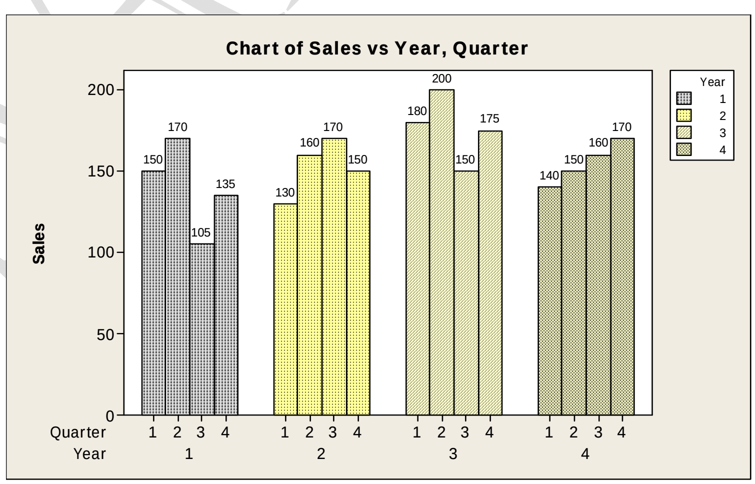



Figure 3.32: A Cluster Bar Chart to Compare Quarterly Sales

## *Pie Charts*

A pie chart is used to show the relative magnitudes of parts to a whole. In this chart relative frequencies of each group of data are plotted. A circle is constructed and is divided into distinct sections. Each section represents one group of data. The area of each section is determined by multiplying the relative frequency of each section by the angle of a circle. Since there are $360^0$ in a circle, each section is multiplied by $360^0$ to obtain the correct number of degrees for each section.

The data file **PIE.MTW** contains the causes of failures in machined parts. To construct a pie chart of this data, follow the steps in Table 3.30.

Table 3.30

| |
|---|
| Open the worksheet **PIE.MTW** |
| From the main menu, select **Graph ➢Pie Chart** |
| In the Pie Chart dialog box, click on the circle next to **Chart values from a table** |
| In the **Categorical variable** box, type or select **Failures** |
| In the **Summary variables** box, type or select **Count** |
| Click on **Pie Chart Options** tab |
| Click on the circle next to **Decreasing volume** |
| Click **OK** |
| Click on **Labels** then click on **Titles/Footnotes** and type a title for your plot |
| Click on **Slice Labels** and check **Frequency and Percentage** |
| Click **OK** in all the boxes |

The pie chart in Figure 3.33 will be displayed. You may double click on the pie chart click custom and select a pattern from the Type box.

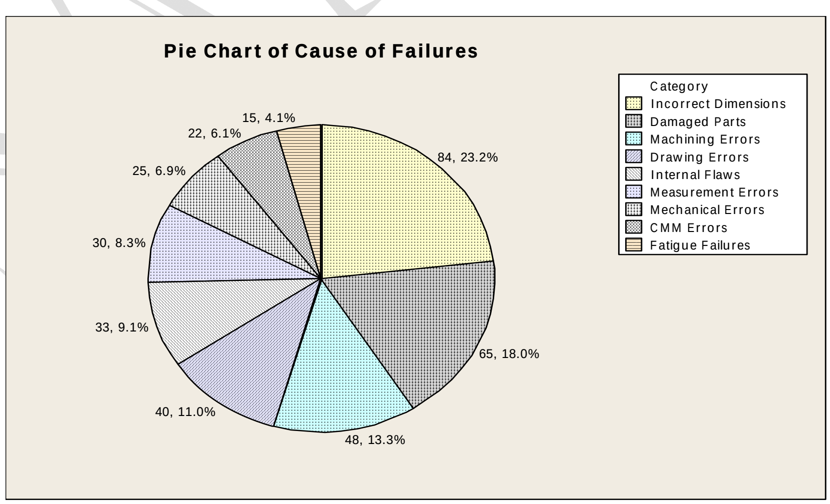
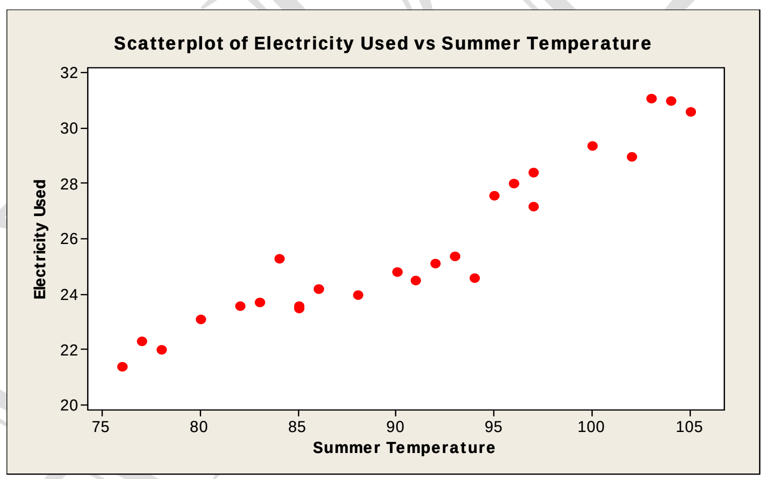



Figure 3.33: Pie Chart of Failure Data

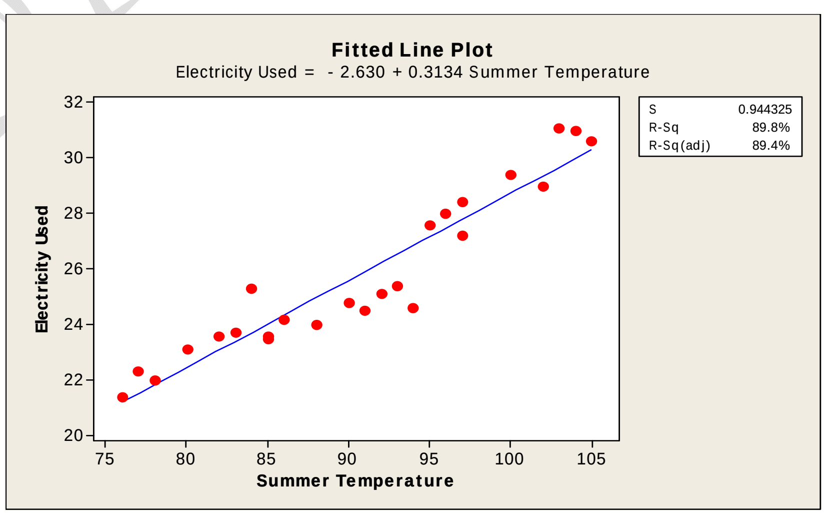
# Scatter Plots

Scatter plots are helpful in investigating the relationship between two variables. One of these variables is considered as a dependent variable and the other an independent variable. The data value is thought of as having an x value and a y value. Thus, we have $(x_i,y_i)$, i=1,2,3,.....,n pairs. If we are interested in the relationship between the two variables, one of the easiest way to investigate this relationship is to plot the (x,y) pairs. This type of plot is known as a scatter plot. Several options are available for scatter plot. We have demonstrated some below. In the example given below, we have plotted the summer temperature and the amount of electricity used by customers (in millions of kilowatts). The data is shown in **SCATTER1.MTW**. Follow the steps in Table 3.32 to do a scatter plot of the data.

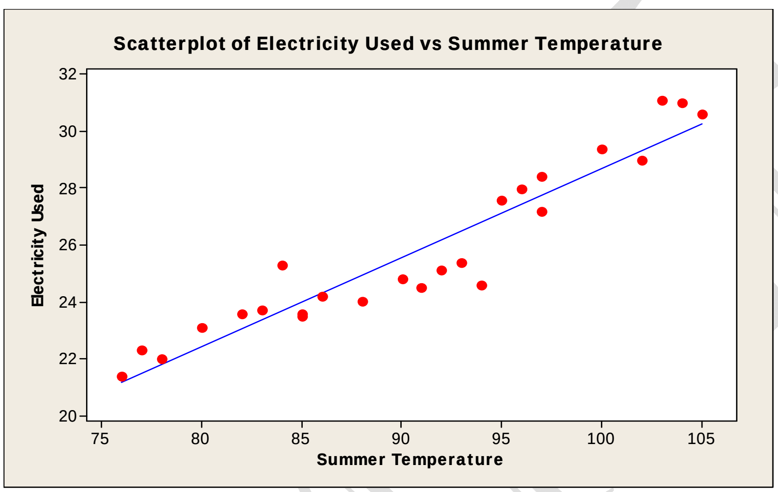
## A simple Scatter Plot




## Scatter Plot with Regression Line

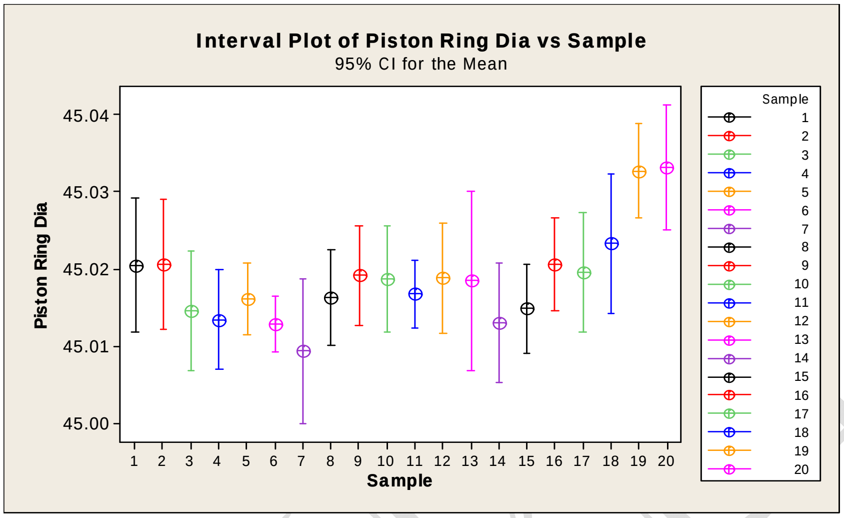


### *Fitted Line Plot with Regression Equation*




## *Interval Plots*

The interval plot displays means and/or confidence intervals for one or more variables. This plot is useful for assessing the measure of central tendency and variability of data. The default confidence interval is 95% however; this can be changed by using the command sequence **Editor ➢Edit Interval Bar ⍰ Options**. We will demonstrate the interval plot using the data in **INTERVAL2.MTW**. This data file contains the amount of beverage in 16 oz. cans from 5 different production lines. The operations manager suspects that the mean content of the cans differs from line to line. He randomly selected 5 cans from each line and measured the contents. The data is shown in **INTERVAL2.MTW**. The interval plot below is created from this data.




## *Time Series Plots*

A time series plots the data over time. The graph plots the ($x_i$,$y_i$) pairs of points and connects these plots through a straight line where the x values are time. The plot is helpful in visualizing a trend or pattern in a data set. In the example below, a time series plot of demand data over time is explained. The data file **TIMESERIES1.MTW** shows weekly demand data for five quarters. Each quarter is divided into 13 weeks.

### *A Simple Time Series Plot*

To do a simple time series plot of the demand data, follow the steps in Table 3.44.

Table 3.44

| |
|---|
| Open the worksheet **TIMESERIES1.MTW** |
| In the **Time Series** dialog box, click on **Simple** then click **OK** |
| Type or select **Demand** for **Series** |
| Click on the **Time/Scale** box |
| Under **Time Scale**, click on the circle next to **Index** |
| Click **OK** in all dialog boxes. |

A simple time series plot shown in Figure 3.48 will be displayed. Note that the **Index** under **Time Scale** is used to label the x-axis with integer values starting from 1. If you want x-axis label as Week, double click on Index when the graph is created, and change the index to week in the dialog box that is displayed.

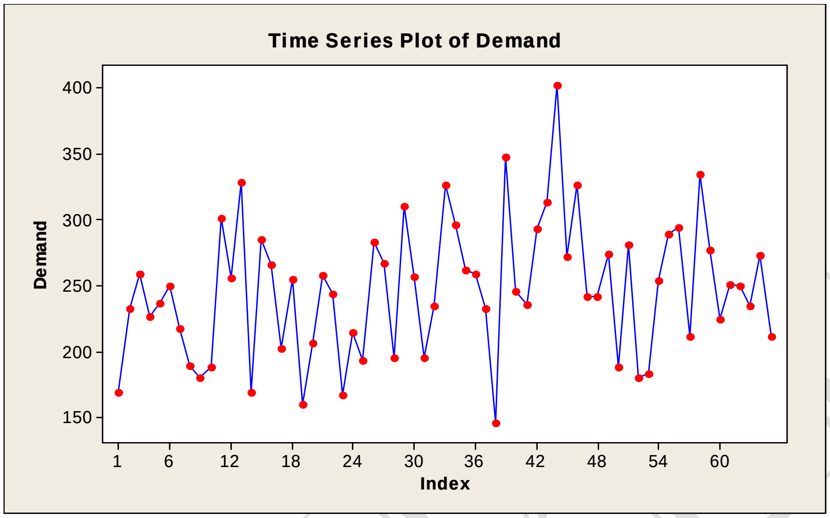



Figure 3.48: A Simple Time Series Plot of Demand Data

## A Simple Time Series Plot using Stamp as Time/Scale

If you want to show quarter and week on the x-axis, use the stamp option under Time/Scale. The Stamp option is used to label x-axis with values from one or more stamp columns. This option is explained below.

A time series plot shown in Figure 3.49 will be displayed.

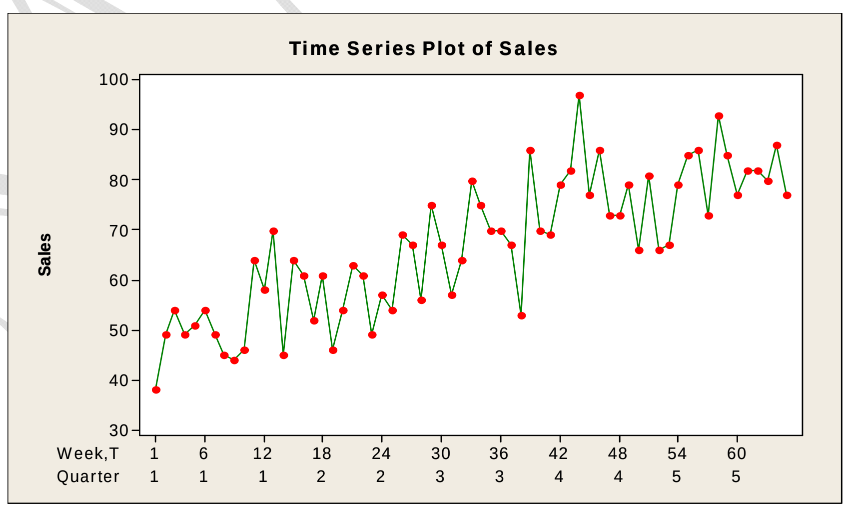



Figure 3.49: A Simple Time Series Plot of Sales Data

### A Multiple Time Series Plot

Use this option to plot two or more series of data (multiple values). For example, suppose you want to plot the actual sales and the forecast on the same plot.

Table 3.46

| |
|---|
| Open the worksheet **TIMESERIES2.MTW** |
| In the Time Series dialog box, click on **Multiple** then click **OK** |
| Type or select **Sales** and **Forecast** for **Series** |
| Click on the **Time/Scale** box |
| Click on the circle next to **Stamp** under **Time Scale** |
| Click on **Stamp columns (1-3, innermost first)** then select **Period,T** |
| Click **OK** in all dialog box. |

A multiple time series plot shown in Figure 3.50 will be displayed.

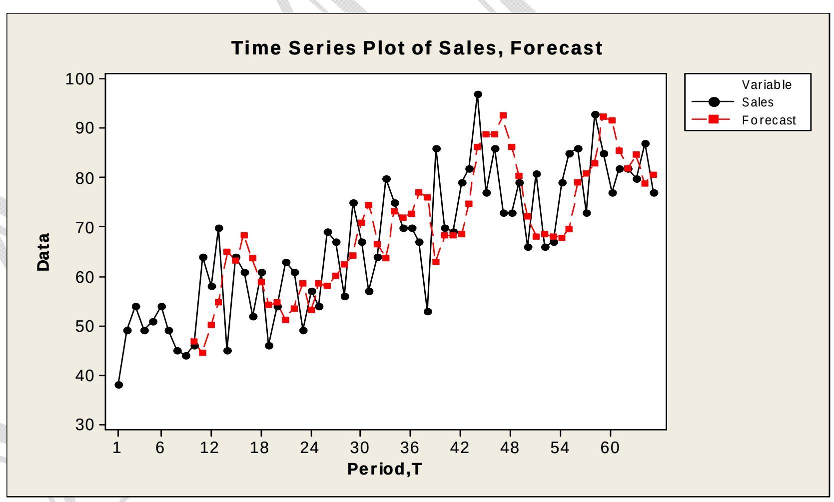



Figure 3.50: A Multiple Time Series Plot Showing Sales and Forecast

## Probability Plots

Probability plots are used to determine if a particular distribution fits sample data. The plot allows us to determine if a distribution is appropriate and also, to estimate the parameters of fitted distribution. We have also seen that an empirical cdf can be used to fit the distribution to the data. The curve in cdf (Figure 3. 51) is not convenient to use and may be misleading sometimes. The probability plots are a better alternative and are achieved by the use of probability paper.

MINITAB provides individual probability plots for the selected distribution for one or more variables. The steps to probability plotting procedure are

1. Hypothesize the distribution: select the assumed distribution that is likely to fit the data
2. Order the observed data from smallest to largest. Call the observed data $x_1, x_2, x_3, \ldots, x_n$
3. Calculate the cumulative percentage points or the plotting position (PP) for the sample of size n (i=1,2,3,....,n) using the following

$$PP = \frac{(i - 0.5)100}{n}$$

Tabulate the $x_i$ values and the cumulative percentage (probability values or PP). Depending on the distribution and the layout of the paper, several variations of cumulative scale are used.
4.Plot the data using the graph paper for the selected distribution. Draw the best fitting line through these points.
5. Draw your conclusion about the distribution.

MINITAB provides the plot based on the above steps. To test the hypothesis, an Anderson-Darling (AD) goodness-of-fit statistic and associated p-value can be used. These values are calculated and displayed on the plot. If the assumed distribution fits the data:

- the plotted points will form a straight line (or approximate a straight line)
- the plotted points will be close to a straight line
- the Anderson-Darling (AD) statistic will be small, and the p-value will be larger than the selected significance level, $\alpha$ ( commonly used values are 0.05 and 0.10).

***Probability Plot: Example 1***

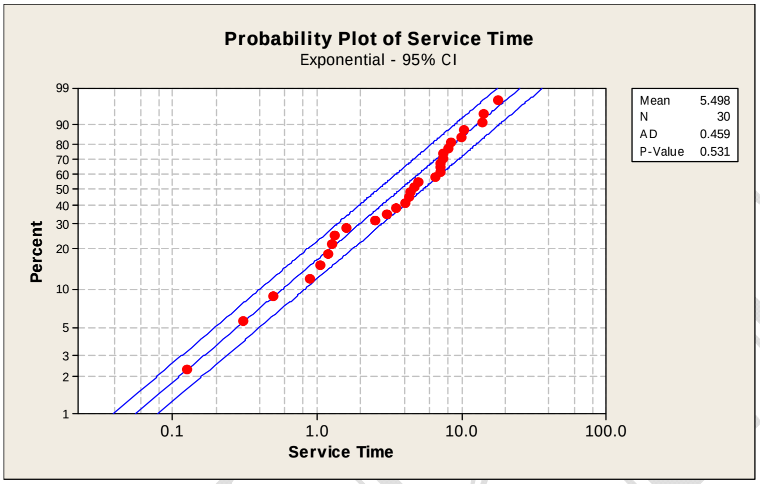
To demonstrate the probability plot, we will use the data in **PROBABILITY Plot. MTW**. The file contains the length of 15 cast iron tubes from a manufacturing process. We want to use the probability plot to check if the data follow a normal distribution. Follow the steps in Table 3.48.

Table 3.48

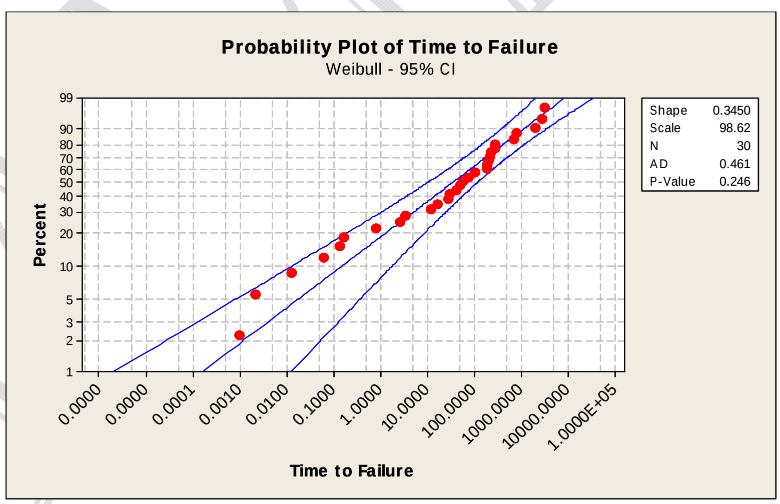
| |
|---|
| Open the worksheet **PROBABILITY Plot. MTW** |
| From the main menu select, **Graph ➢Probability Plot** |
| Click on **Single** then click **OK** |
| For **Graph Variables**, type or select **Length (Cm)** |
| Click on **Distribution** and select **Normal** from drop down menu in the distribution box (Leave **Historical Parameters** box blank) |
| Click **OK** in all boxes. |

The probability plot is shown in Figure 3.52. From the plot we can see that the cumulative percentage points approximately form a straight line and the points are close to the straight line. The calculated p-value is 0.508. At a 5% level of significance ($\alpha$=0.05), p-value is greater than $\alpha$ so we cannot reject the null hypothesis that the data follow a normal distribution. If you place the cursor on the middle line, you can see the calculated cumulative percentage points and also the lower and the upper bounds for the points.

### *Figure 3.52: Example of a Probability Plot of Length Data*




### *Probability Plot: Example 2*

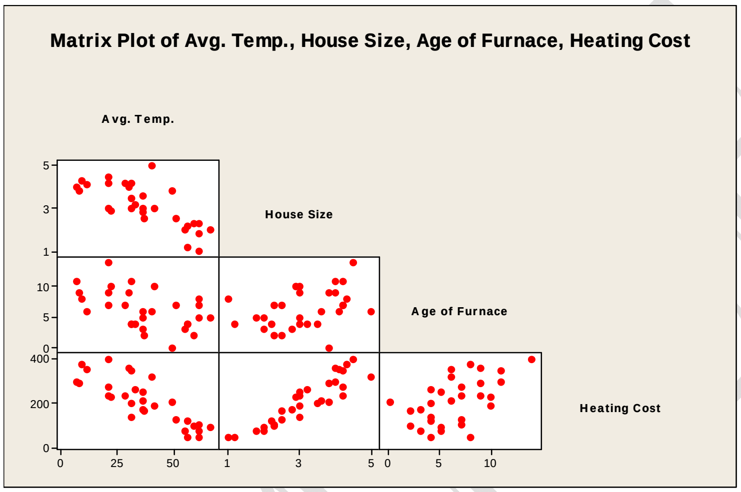



## *Matrix Plot*

A matrix plot can be used to investigate the relationships between pairs of variables by creating an array of scatterplots.

One of the options under matrix plot is **Each Y versus each X**. You can specify y and x variables and create a plot for each possible xy pairs. This is useful when the relationships between certain pairs of variables are of interest. Using **Each Y versus each X**, you can create a simple, with groups, and with smoother plots. Some examples of matrix plot are given below.

***Matrix of Plots: Simple***




***Chapter 3 contains numerous charts/graphs used in Six Sigma, Lean, and Design for Six Sigma. The chapter contains detailed computer instruction with several examples of each plot.***

***To buy chapter 3 or Volume I of Six Sigma Quality Book, please click on our products on the home page.***