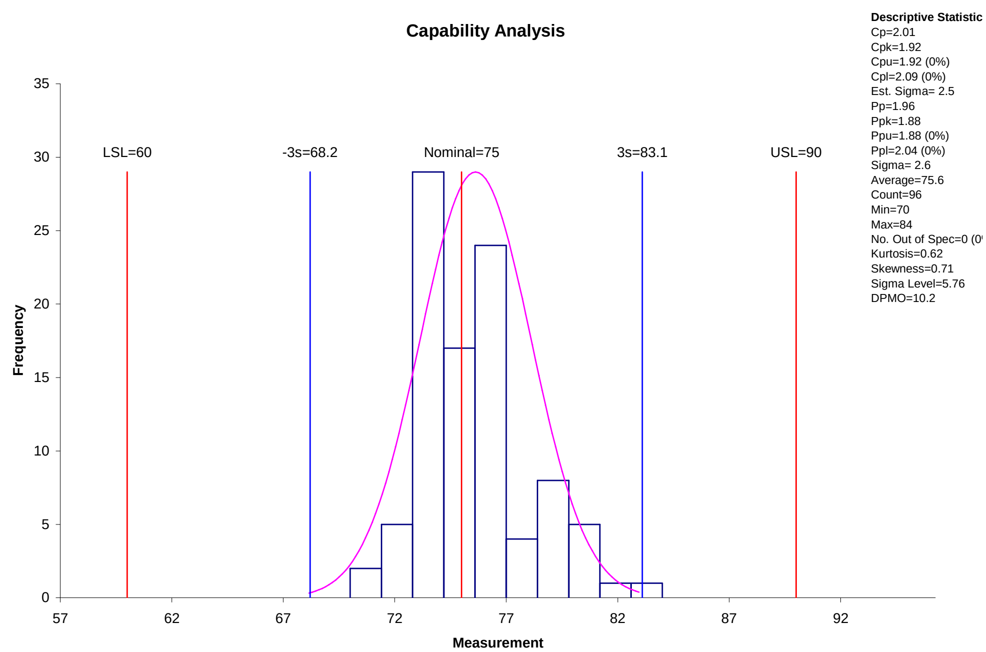
Instruction Manual for SPC for MS Excel V4.0[1]



Thank you for selecting our software package. This manual contains the instructions for using SPC for MS Excel Version 4 ***with the exceptions of the Measurement Systems Analysis, Multiple Linear Regression, and Experimental Design, which are each contained in separate instruction manuals***. This program is owned by BPI Consulting, LLC. This program cannot be copied or used unless under license with BPI Consulting, LLC. BPI Consulting, LLC is not liable for any decisions made based on the use of this software package.

Requirements: This program is a Microsoft Excel® add-in. You must Microsoft Excel® for this program to work. This program supports any version of Excel from 2000 on.

BPI Consulting, LLC
20314 Lakeland Falls
Cypress, TX 77433
800-274-2874
281-304-9504
www.spcforexcel.com
mailto:support@spcforexcel.com

[1] Instructions for Measurement Systems Analysis, Multiple Linear Regression and Design of Experiments are in separate manuals.

Table of Contents

Pareto Diagrams 4
Histograms....7
Attribute Control Charts 10
p Charts....10
np Control Charts....13
c Control Charts 16
u Control Charts....19
Variable Control Charts 22
Xbar-R Control Chart, Xbar-s Control Charts, and Moving Average/Moving Range Charts 22
X-mR (Individuals) Control Charts....28
Table X-MR (Individuals) Charts....33
Run Charts 35
Updating Charts....39
Changing Chart Options....39
Single Point Actions 40
All Points Action....42
Scatter Diagrams....43
Cause and Effect Diagrams 46
Process Capability....48
Multiple Process Capability Analysis 52
Six Sigma Tools 54
Stacked and UnStacked Data Entry 54
Single/Two Samples....55
One Sample z and t Tests for a Mean 55
One Sample Variance....59
Two Sample z and t Tests for Difference in Two Means....62
Paired Sample Comparison t Test....66
One Proportion Test 69
Two Proportions Test 72
Sample Size Calculator....75
Multiple Means and Variances 76
Fisher's LSD Method for Means....76
Tukey's Method for Means....79
Bonferroni's Method for Means....82
Analysis of Means 85
Box and Whisker Plots 88
Bartlett's Test for Equality of Variances 90
Modified Levene's Test for Equality of Variances....92
Cause and Effect 94
Correlation Coefficients....94
Failure Mode and Effects Analysis Template....96
Plot Multiple Y Variables against One X Variable 97
Scatter Plot Matrix....98
Miscellaneous....100
Normal Probability Plot with Anderson-Darling Test 100
Distributions 102
Side by Side Histogram 104
Chi-Square Goodness of Fit Test....106
Chi-Square Test for Association....108

Nonparametric..110
One Sample Sign Test 110
One Sample Wilcoxon Signed Rank Test 112
Two Sample Mann-Whitney Test 114
Kruskal-Wallis Test for Multiple Samples 116
Select Cells...118
Transfer Charts to PowerPoint or Word..119

Pareto Diagrams

A Pareto diagram is a special type of bar chart that is used to separate the "vital few" from the "trivial many." It is based on the 80/20 rule; e.g., 20% of our customers buy 80% of our products. The horizontal (x) axis most often represents problems or causes of problems (the "categories"). The vertical (y) axis most often represents frequency or cost (the "frequencies". The problem or cause that occurs most frequently (or costs the most) is listed first on the x axis. The second most frequently occurring problem or cause is listed second and so on. A bar is generated for each cause or problem. The height of the bar is the frequency with which that problem or cause occurred. A cumulative percentage line is sometimes added to the Pareto diagram.

For more information on Pareto diagrams, please see our June 2004 newsletter on Pareto diagrams on our website (http://www.spcforexcel.com/articleslist.htm). The example below shows how to construct a Pareto diagram using the SPC for MS Excel software. There are four different options for data entry.

Example

A company is monitoring return goods and reasons for those returns. Data has been collected and we want to use a Pareto diagram to highlight returns goods by product and reason for return.

1. Enter the data into an Excel spreadsheet. There are four different options that can be used for entering data.
    a. Option 1: Frequencies Already Totaled
        i. The categories are in one column or row and the frequencies are in another column or row with the frequencies already totaled. The data entry requirements are given below. The columns and the rows do not have to be adjacent to each other. The table below shows the data entered as columns.

Products	Number of Returns
A	15
B	25
C	8
D	2

    b. Option 2: Frequencies Need to be Totaled
        i. *The categories are listed in a range on the spreadsheet. There is no list of unique categories and the frequency needs to be determined.* The data entry requirements are given below. The data can be in multiple rows and columns. Empty cells can be included. The table below shows the data entered in a number of columns and rows. The program will total each unique entry and develop a Pareto diagram based on those totals.

Customer Did Not Need	Customer Did Not Need	Customer Did Not Need
Broken		Broken
Wrong Quantity	Wrong Quantity	
Wrong Quantity	Wrong Quantity	Wrong Quantity
Salesman Ordered Wrong	Salesman Ordered Wrong	Salesman Ordered Wrong
		Wrong Quantity
Wrong Quantity	Wrong Quantity	Wrong Quantity
Broken	Broken	Broken
Customer Did Not Need	Customer Did Not Need	Customer Did Not Need
Salesman Ordered Wrong		Salesman Ordered Wrong
Wrong Quantity	Wrong Quantity	

c. Option 3: Frequencies per Period in Columns

   i. *The categories are in one column and the frequency is in columns. The frequencies are already summed by column (e.g., week).* The data entry requirements are shown below. The data must be in columns. The program will total the results and develop the Pareto.

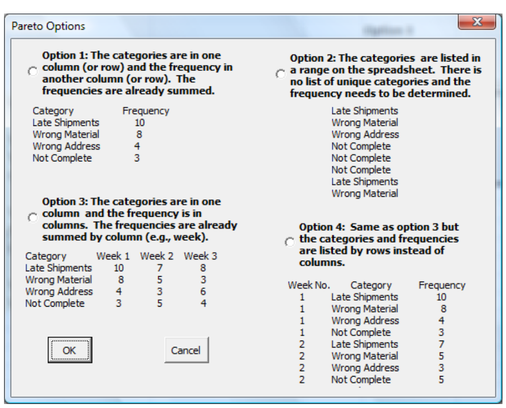
Category	Week 1	Week 2	Week 3
Late Shipments	10	7	8
Wrong Material	8	5	3
Wrong Address	4	3	6
Not Complete	3	5	4
Wrong Price	6	2	1

d. Option 4: Frequencies per Period in One Column

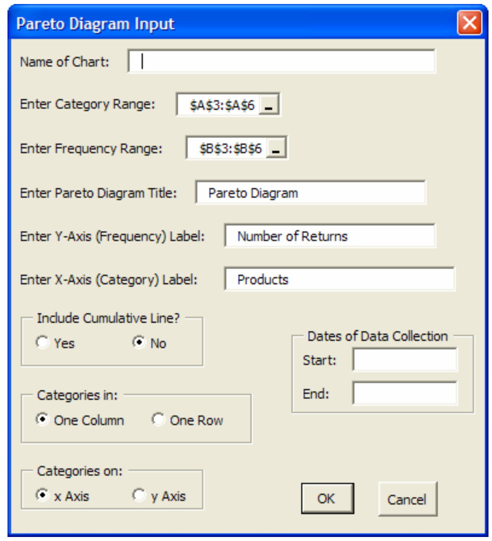
   i. *Same as option 3 but the categories and frequencies are listed by rows instead of columns.* The data entry requirements are shown below. The data must be in columns. The program will total the results and develop the Pareto.

Week	Product	Number of Returns
1	A	4
1	B	2
1	C	12
1	D	3
2	A	0
2	B	3
2	C	1
2	D	4
3	A	4
3	B	2
3	C	5
3	D	1
4	A	8
4	B	1
4	C	10
4	D	1

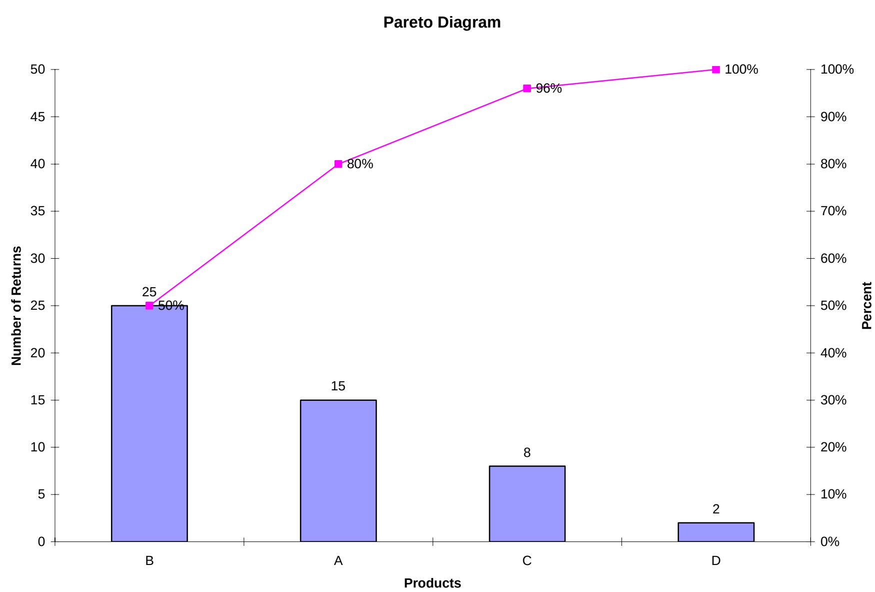
2. Select the range shaded in gray (based on our data entry).
3. Select the Pareto diagram icon from the SPC menu to show the Pareto Options Screen.
4. Select one of the four options based on your data entry.



5. Fill in the Pareto Diagram Input form (the input form below is for Option 1; all are very similar).



   a. *Name of Chart:* the name that appears on the worksheet tab; this name must be unique
   b. *Category Range:* range containing categories (product in the above example); default is the first column of range selected on the worksheet
   c. *Frequency Range:* range containing the totaled frequencies; default is the second column of the range selected on the worksheet
   d. *Pareto Diagram Title:* the title to appear at the top of the chart; default is Pareto Diagram
   e. *Y Axis (Frequency) Label:* the label for the y axis; default is the cell above the selected frequency range on the worksheet (for data in columns)
   f. *X Axis (Category) Label:* the label for the x axis; default is the cell above the selected category range on the worksheet (for data in columns)
   g. *Include Cumulative Line?:* this option adds a cumulative line to the chart; default is No.
   h. *Categories in:* option for data being in columns or rows; programs selects default based on the data selected on the worksheet
   i. *Categories on:* option to put the categories on the x axis or y axis; default is the x axis; cannot put the categories on the y axis if the cumulative line option is set to yes.
   j. *Dates of Data Collection:* option to enter the start and end dates of data collection; not required for the chart.
6. Select OK and the Pareto diagram will be constructed. The diagram below used includes the cumulative line.



The Pareto diagram can be updated when new data is added by selecting the Update Chart toolbar on the SPC Menu. You may also change current options on the chart by selecting the Options icon from the SPC toolbar.

Histograms

A histogram is a bar chart that provides a snapshot in time of the variation in a process. It tells us how often a value or range of values occurred in a given time frame. A histogram will tell us the most frequently occurring value (the mode), the overall range, and the shape of the distribution (e.g., bell-shaped, skewed, bimodal, etc.). It is best to have 50 to 100 data points to construct a histogram, if possible. This program will construct a histogram from the raw data. It will automatically determine the number of classes (bars) as well as the class width. You have the opportunity to change the number of classes.

For more information on Pareto diagrams, please see our December 2005 and January 2006 newsletters on histograms on our website (http://www.spcforexcel.com/articleslist.htm). An example of how to make a histogram using the program is given below.

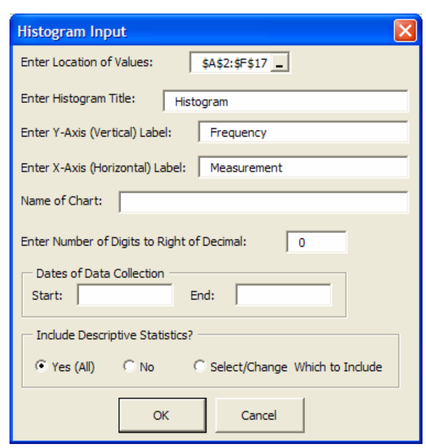
Example

A production superintendent is tracking yields from a reactor. He wants to create a histogram to determine the shape of the distribution. That data is given below.

1. Enter the data into a spreadsheet as shown below.

81	77	75	74	77	73
77	74	76	75	79	74
74	79	73	75	75	74
75	80	80	79	72	78
73	74	74	73	75	74
77	75	75	72	75	74
76	75	74	74	78	75
76	76	78	77	78	75
74	76	77	76	72	73
79	82	73	75	74	79
77	73	72	75	73	73
76	76	76	75	74	72
76	76	76	74	79	79
75	81	77	74	77	71
84	74	79	70	77	74
73	77	76	74	81	75

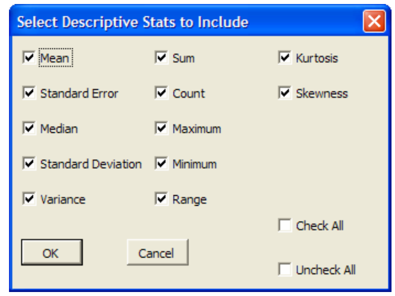
2. Select the histogram icon from the SPC Menu.
3. Fill in the Histogram Input form.
   a. *Enter location of Values:* This is the range containing the values for the histogram. The default range is the range selected on the worksheet before selecting the histogram option on the toolbar.
   b. *Enter Histogram Title:* This is the title that goes on the histogram chart. The default value is "Histogram."
   c. *Enter Y-Axis (Vertical Label):* This is the vertical axis label. The default is "Frequency."
   d. *Enter X-Axis (Horizontal Label):* This is the horizontal axis label. The default is "Measurement."
   e. *Name of Chart:* This is very important. This will be the name of the worksheet tab that contains the chart in your workbook.



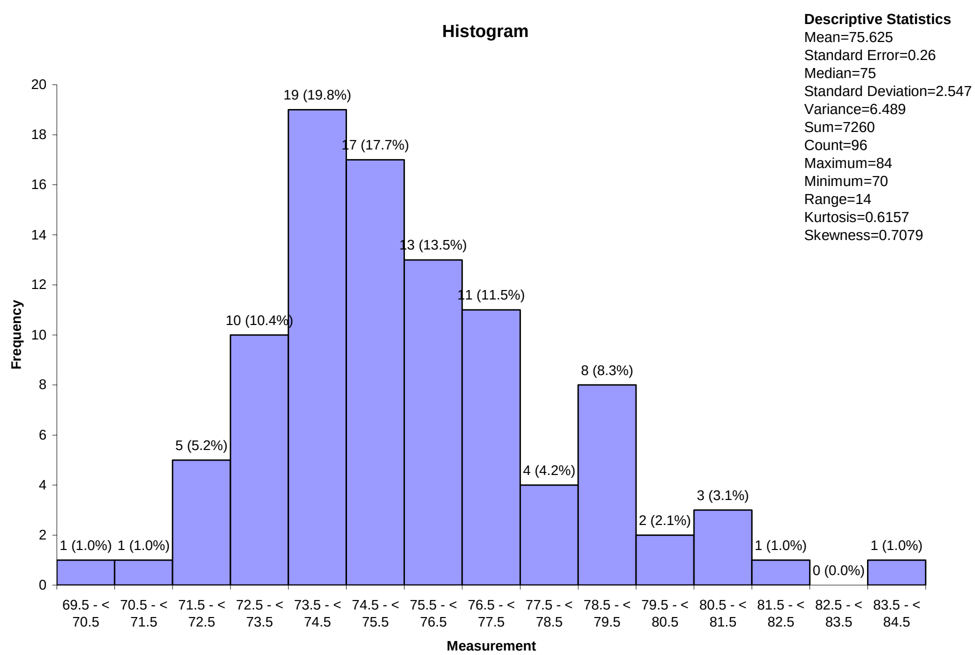
f. *Enter Number of Integers to Right of Decimal:* This is the rounding that is used in the data. For example, if the data contains whole numbers, this value is 0 (the default value). If the data has one decimal point to the right of the data (as shown in the data above), this value is 1. It is used to set the class boundaries.

g. *Dates of Data Collection:* Add the starting date and ending dates of data collection. These dates are optional. If entered, they will appear in a dialog box in the lower left-hand corner of the chart.

h. *Include Descriptive Statistics?*" If you want the descriptive statistics on the chart, select "Yes." The descriptive statistics include the average, standard deviation, count, etc. The default value is "Yes." There is also the option to "Select Which to Include." This option allows you to determine which of the descriptive statistics you want to include. If you select this option, you will see the dialog box to the right. Select which statistics you want to include. The statistics you select will remain the same if you update the histogram. You can "Check All" or "Uncheck All" if desired.

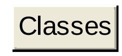


4. Select OK and the histogram will be created as shown below.



The number on the top of each bar represents the number in the class; the percent is the percentage of the total points in the class. The statistics are printed on the right hand side of the chart.

The number of classes (bars) on the histogram is determined automatically by the program. It is set as the square root of the number of data points in the range. Once the histogram is made, you can change the number of classes. There is a button in the upper left hand corner of the histogram chart that is used for this (you will see it when the histogram is first made).



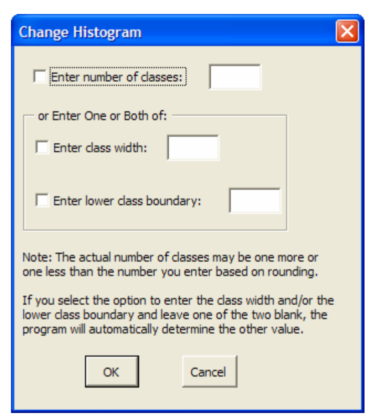
When you select this button on the chart, you will get the dialog box to the below. There are essentially two options:

- Enter the number of classes you want and select OK. The chart will then be displayed with the new number of classes.
- Enter the class width and enter the lower bound. This lets you set the starting point for the histogram (the lower bound) and the width of each class. The number of classes is set by these two values.

You also have the option to view the frequency distribution for the histogram. This is done by selecting the button with the caption "Freq. Dist." This button appears the first time the histogram is made. Selecting the button again hides the frequency distribution. The frequency distribution for this example is shown below.

Class Boundaries	Class Limits	Freq.	Rel. Freq.
69.5 - < 70.5	70 - 70	1	*1.0%*
70.5 - < 71.5	71 - 71	1	*1.0%*
71.5 - < 72.5	72 - 72	5	*5.2%*
72.5 - < 73.5	73 - 73	10	*10.4%*
73.5 - < 74.5	74 - 74	19	*19.8%*
74.5 - < 75.5	75 - 75	17	*17.7%*
75.5 - < 76.5	76 - 76	13	*13.5%*
76.5 - < 77.5	77 - 77	11	*11.5%*
77.5 - < 78.5	78 - 78	4	*4.2%*
78.5 - < 79.5	79 - 79	8	*8.3%*
79.5 - < 80.5	80 - 80	2	*2.1%*
80.5 - < 81.5	81 - 81	3	*3.1%*
81.5 - < 82.5	82 - 82	1	*1.0%*
82.5 - < 83.5	83 - 83	0	*0.0%*
83.5 - < 84.5	84 - 84	1	*1.0%*

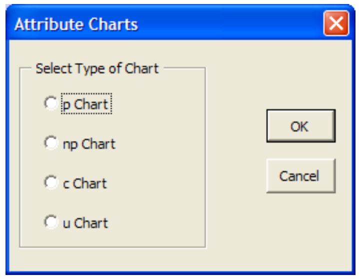
The histogram can be updated when new data is added by selecting the Update Chart toolbar on the SPC Menu. You may also change current options on the chart by selecting the Options icon from the SPC toolbar.



Attribute Control Charts

This program constructs p, np, c, and u attribute control charts. The data entry depends on the type of chart you are using.

You access this feature by selecting the attribute control chart icon ( ) on the SPC menu. You will see the form below. Select the type of chart you want to make. Each attribute control chart is described below.



p Charts

A p control chart is used to examine the variation in the proportion (or percentage) of defective items in a group of items. An item is defective if it fails to conform to some preset specification (operational definition). The p control chart is used with "yes/no" attributes data. This means that there are only two possible outcomes: either the item is defective or it is not defective. For example: either the phone is answered or it is not answered.

For more information on Pareto diagrams, please see our July 2005 newsletter on p control charts on our website (http://www.spcforexcel.com/articleslist.htm). An example of how to make a p control chart is given below.

Example

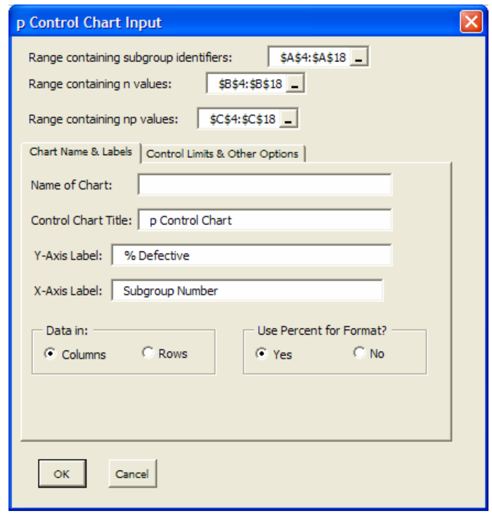
A company uses telemarketing to help boost sales. The sales manager is interested in tracking the % of telemarketing calls that result in an order each day. The data is has collected is given below. He wants to use that data to construct a p control chart. "n" is the subgroup size (the number of telemarketing calls made each day). "np" is the number of "defective" items -- in this case, the number of calls that result in an order. "p" is the proportion defective and is determined by p = np/n. For example, on the first day there were 40 telemarketing calls made (n = 40). Of these, 5 resulted in an order (np = 5). Thus, p = np/n = 5/40 = 0.125 or 12.5%. The values of p are plotted over time.

1. Enter the data into a spreadsheet as shown below.

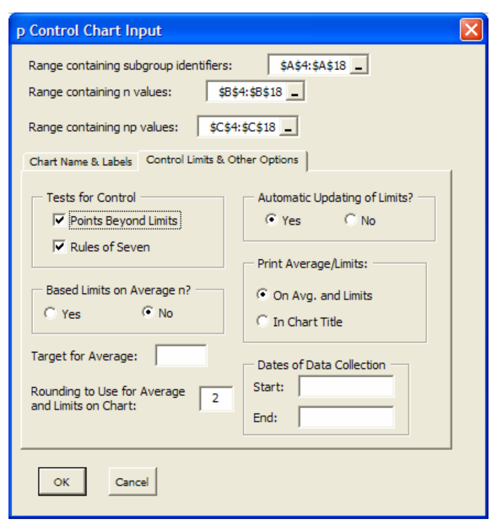
Date	Number of Telemarketing Calls (n)	Number that Result in an Order (np)
2/1/2008	40	5
2/2/2008	63	10
2/3/2008	47	12
2/4/2008	52	7
2/5/2008	34	3
2/6/2008	59	21
2/7/2008	36	12
2/8/2008	71	7
2/9/2008	53	11

2/10/2008	50	3
2/11/2008	41	12
2/12/2008	48	10
2/13/2008	67	5
2/14/2008	45	12
2/15/2008	54	18

2. Select the dates (the shaded area).
3. Select the following from the SPC Menu: Attribute Control Chart → p Control Chart
4. Fill out the p Control Chart Input form. The ranges are input at the top of the form; there are two pages on the form: Chart Names & Labels and Control Limits & Other Options.



   a. *Range containing the subgroup identifiers:* This is the range containing the subgroup numbers (dates in the above example). The default value is the range selected on the worksheet prior to selecting the attribute control option on the toolbar.
   b. *Range containing the n values:* This is the range containing the subgroup size (n). The default value is the range next to the subgroups unless you selected multiple ranges using the control key.
   c. *Range containing the np values:* This is the range containing the number non-conforming (np). The default value is the range next to the n values unless you selected multiple ranges using the control key.
   d. *Chart Names & Labels Page*
      i. *Name of Chart:* This is very important. Decide what you want to call the chart. This will be the name of the sheet that contains the chart in your workbook.
      ii. *Control Chart Title:* This is the title that goes on the control chart. The default value is "p Control Chart."
      iii. *Y-Axis Label:* This is the vertical axis label. The default value is "% Defective."
      iv. *X-Axis Label:* This is the horizontal axis label. The default value is "Subgroup Number"
      v. *Data in:* Select columns or rows depending on how the data is entered into the spreadsheet. The default depends on the range selection on the worksheet.
      vi. *Use Percent for Format?*: Select yes to format the chart as percent; no to format the chart as a general number.
   e. *Control Limits & Other Options Page (shown to the right).*



      i. *Test for Control:* There are two options: points beyond the limits and the rules of seven (seven in a row above or below the average, or seven in a row trending up or trending down).
      ii. *Automatic Update of Limits?*: This determines if the control limits are automatically updated when you add additional data to the chart. Select "Yes" if you want the control limits to automatically update; no if you don't want the limits to automatically update. The default is yes.

iii. *Based Limits on Average n?* Select no to change the limits each time the subgroup size changes. Select Yes to base the limits on the average subgroup size. The default value is No.

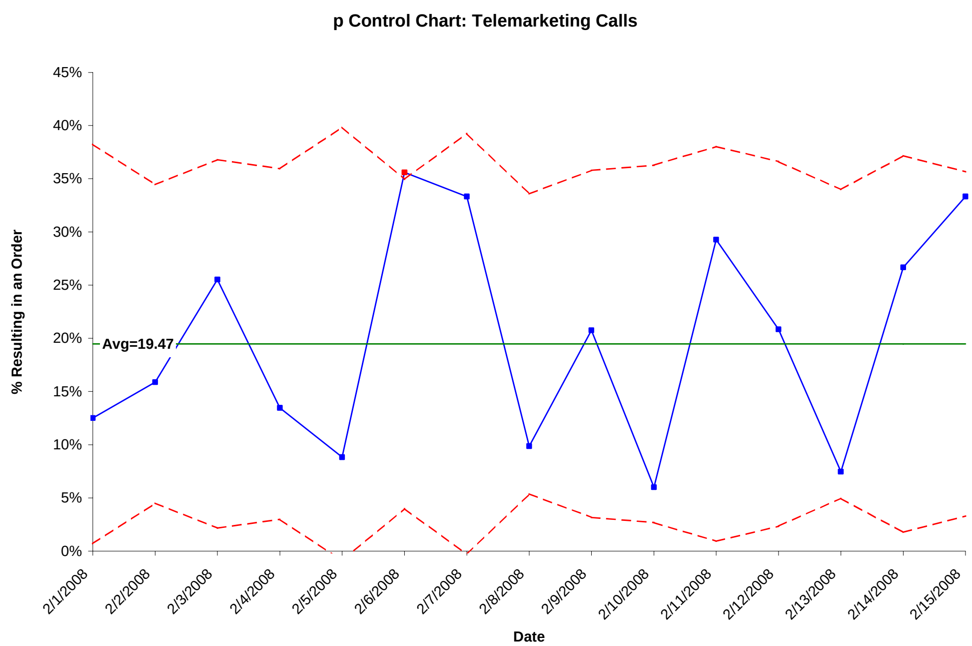
iv. *Print Average/Limits:* Selecting "On Avg. and Limits" will print these on the lines in the chart. Selecting "In Chart Title" will print the values in the chart title.

v. *Target for Average:* This is the target value for the variable. It is not required.

vi. *Dates of Data Collection:* Add the starting date and ending dates of data collection. These dates are optional. If entered, they will appear in a dialog box in the lower left-hand corner of the chart.

vii. *Rounding to Use in Titles:* This the rounding to use for the average and control limits printed in the title. The default value is determined by the program.

5. Select OK.
6. The p control chart is generated on a new chart sheet. The p control chart for the example data is shown below.



The values of p are plotted over time. The average ($\overline{p}$), the upper control limit (UCL) and the lower control limit (LCL) are calculated using the equations below. The average is plotted as a green solid line and the control limits are plotted as red dashed lines. The control limits in this example vary because the subgroup size varies. The values for the average and control limits (based on the average subgroup size, $\overline{n}$) are also printed on the chart or in the title depending on the option selected. Out of control points are in red.

$$\overline{p} = \frac{\sum np}{\sum n} \qquad \mathrm{UCL} = \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \qquad \mathrm{LCL} = \overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

The above control limits are not valid for the "small np case." This occurs when $n\overline{p} < 5$ or $n(1-\overline{p}) < 5$. In this case, the program automatically calculates the control limits using the binomial distribution. The p control chart can be updated when new data is added by selecting the Update Chart toolbar on the SPC Menu. You may also change current options on the chart by selecting the Options icon from the SPC toolbar. You can also split control limits, start the chart at a new point, and add comments (see the Single Point Actions section in this manual) as well as remove all out of control points or set the range to base the control limits on (see the All Points Action section in this manual).

np Control Charts

An np control chart is used to monitor the variation in the number of defective items in a group of items. *With this chart, the subgroup size (n), the number of items in the group, must be the same each time.* An item is defective if it fails to conform to some preset specification (operational definition). The np control chart is used with "yes/no" attributes data. This means that there are only two possible outcomes: either the item is defective or it is not defective. For example: either the phone is answered or it is not answered. An example of how to make an np control chart is shown below.

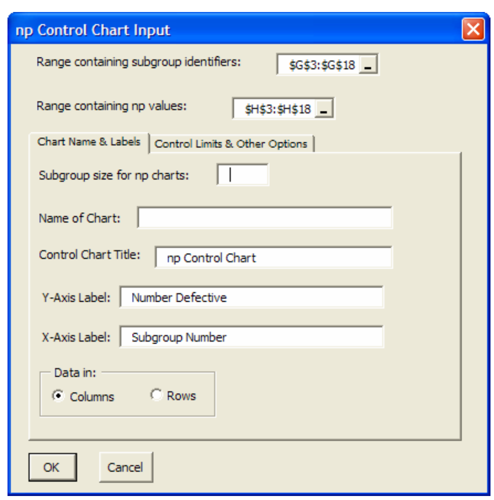
<u>Example</u>

An accounting supervisor is monitoring the number of defective invoices per day. A random sample of 100 invoices is taken each day. Each invoice is checked and rated as defective or not defective. The data is shown below. In this case, the subgroup size is constant (100). np is the number of defective items. For example, on day one, there were 22 defective invoices.

1. Enter the data into a spreadsheet as shown below.

Day Number	Number of Defective Invoices (np)
1	22
2	33
3	24
4	20
5	18
6	24
7	24
8	29
9	18
10	27
11	31
12	26
13	31
14	24
15	22

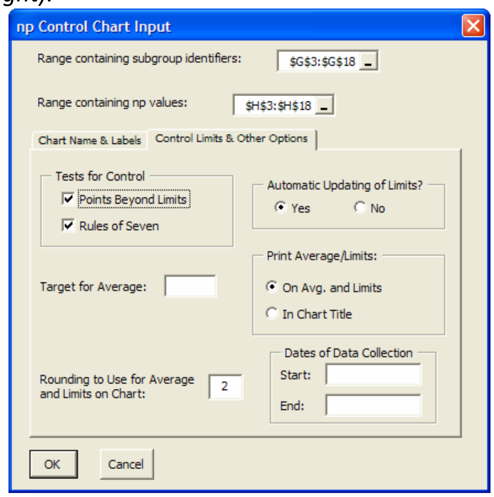
2. Select the data under Day Number (the shaded area).
3. Select the following from the SPC Menu: Attribute Control Chart → np Control Chart
4. Fill out the np Control Chart Input form. The ranges are input at the top of the form; there are two pages on the form: Chart Names & Labels and Control Limits & Other Options.
    a. *Range containing the subgroup identifiers:* This is the range containing the subgroup numbers (dates in the above example). The default value is the range selected on the worksheet prior to selecting the attribute control option on the toolbar.
    b. *Range containing the np values:* This is the range containing the number non-conforming (np). The default value is the range next to the subgroup identifiers unless you selected multiple ranges



using the control key.

c. *Chart Names & Labels Page*
   i. *Subgroup size for np chart:* Enter the constant subgroup size. It is required.
   ii. *Name of Chart:* This is very important. Decide what you want to call the chart. This will be the name of the sheet that contains the chart in your workbook.
   iii. *Control Chart Title:* This is the title that goes on the control chart. The default value is "p Control Chart."
   iv. *Y-Axis Label:* This is the vertical axis label. The default value is "% Defective."
   v. *X-Axis Label:* This is the horizontal axis label. The default value is "Subgroup Number"
   vi. *Data in:* Select columns or rows depending on how the data is entered into the spreadsheet. The default depends on the range selection on the worksheet.

d. *Control Limits & Other Options Page (shown to the right).*



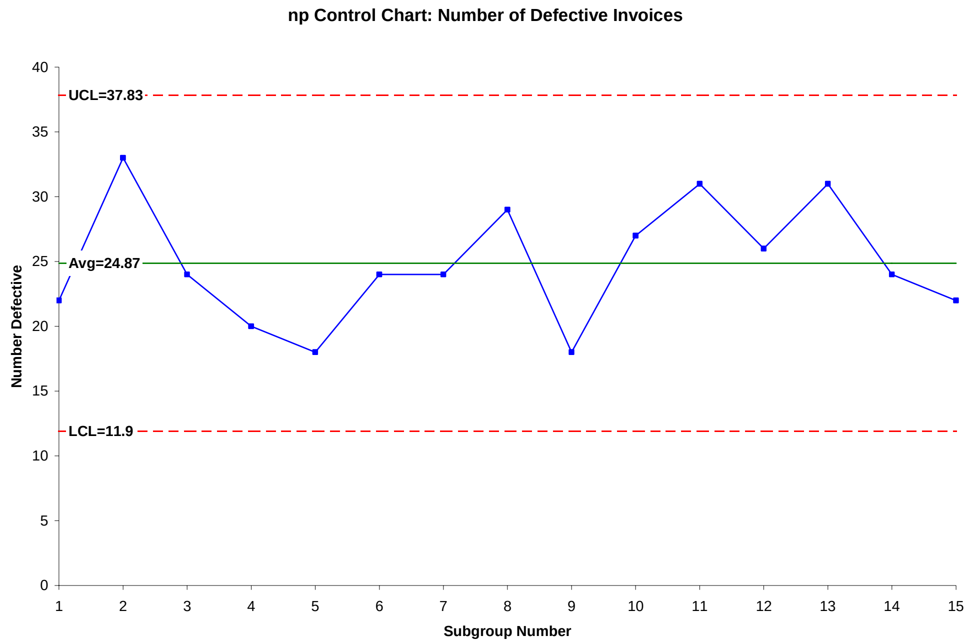
   i. *Test for Control:* There are two options: points beyond the limits and the rules of seven (seven in a row above or below the average, or seven in a row trending up or trending down).
   ii. *Automatic Update of Limits?*: This determines if the control limits are automatically updated when you add additional data to the chart. Select "Yes" if you want the control limits to automatically update; no if you don't want the limits to automatically update. The default is yes.
   iii. *Print Average/Limits:* Selecting "On Avg. and Limits" will print these on the lines in the chart. Selecting "In Chart Title" will print the values in the chart title.
   iv. *Target for Average:* This is the target value for the variable. It is not required.
   v. *Dates of Data Collection:* Add the starting date and ending dates of data collection. These dates are optional. If entered, they will appear in a dialog box in the lower left-hand corner of the chart.
   vi. *Rounding to Use in Titles:* This the rounding to use for the average and control limits printed in the title. The default value is determined by the program.

5. Select OK.
6. The np control chart is generated on a new chart sheet. The np control chart for the example data is shown below.

The values of np are plotted over time. The average ($n\overline{p}$), the upper control limit (UCL), and the lower control limit (LCL) are calculated using the equations below. k is the number of subgroups used in the calculations (k = 15 in this chart). The average is plotted as a solid green line and the control limits are plotted as red dashed lines. The values for the average and control limits, along with the subgroup size, are printed on the chart or in the chart title depending on the option selected. Out of control points are in red.

$$n\overline{p} = \frac{\sum np}{k} \qquad \overline{p} = \frac{n\overline{p}}{n} \qquad UCL = n\overline{p} + 3\sqrt{n\overline{p}(1-\overline{p})} \qquad LCL = n\overline{p} - 3\sqrt{n\overline{p}(1-\overline{p})}$$

The above control limits are not valid for the "small np case." This occurs when $n\overline{p} < 5$ or $n(1-\overline{p}) < 5$. In this case, the program automatically calculates the control limits using the binomial distribution.



The np control chart can be updated when new data is added by selecting the Update Chart toolbar on the SPC Menu. You may also change current options on the chart by selecting the Options icon from the SPC toolbar. You can also split control limits, start the chart at a new point, and add comments (see the Single Point Actions section in this manual) as well as remove all out of control points or set the range to base the control limits on (see the All Points Action section in this manual).

c Control Charts

A c control chart is used to monitor the variation in the number of defects. A defect occurs when something does not meet a preset specification (operational definition). A c control chart is used with counting type attributes data (e.g., 0, 1, 2, 3). These are whole numbers. In addition, to use a c control chart, two other conditions must be true:

- The opportunity for defects to occur must be large.
- The actual number that occurs must be small.

With a c control chart, we are often looking at an area, not a group of items. For example, we may use a c control chart to monitor injuries in a chemical plant. In this case, the subgroup is the plant. The opportunity for injuries to occur is large; the actual number that occurs is small relative to the opportunity. With a c control chart, the area of opportunity for defects to occur must be constant. For more information on c control charts, please see our July 2004 newsletter on c control charts on our website (http://www.spcforexcel.com/articleslist.htm). An example of how to construct a c control chart is given below.

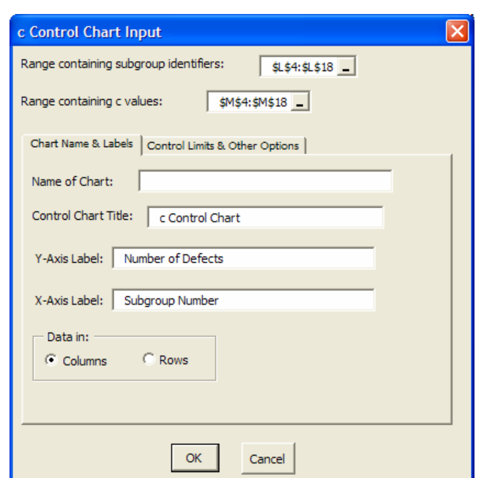
Example

A distributor is monitoring the number of returned goods per day. The data is shown below. "c" is the number of returned goods each day. For the first date, there were 20 goods returned.

1. Enter the data into a spreadsheet as shown below.

Day	Number of Returned Goods (c)
2/1/2008	20
2/2/2008	24
2/3/2008	14
2/4/2008	32
2/5/2008	28
2/6/2008	16
2/7/2008	19
2/8/2008	32
2/9/2008	27
2/10/2008	25
2/11/2008	24
2/12/2008	12
2/13/2008	17
2/14/2008	44
2/15/2008	13

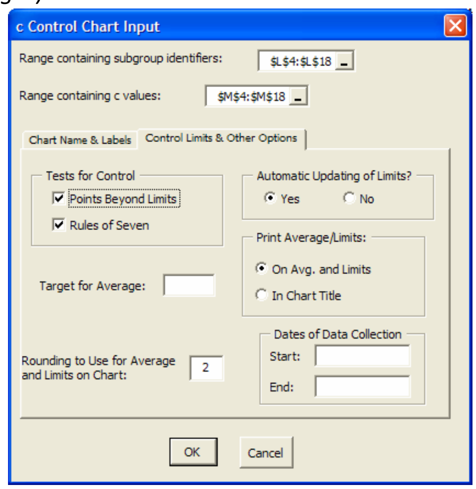
2. Select the data under (the shaded area).
3. Select the following from the SPC Menu: Attribute Control Chart → c Control Chart
4. Fill out the c Control Chart Input form. The ranges are input at the top of the form; there are two pages on the form: Chart Names & Labels and Control Limits & Other Options.
   a. *Range containing the subgroup identifiers:* This is the range containing the subgroup numbers (dates in the above example). The default value is the range selected on the worksheet prior to selecting the attribute control option on the toolbar.
   b. *Range containing the c values:* This is the range containing the number of defects (c). The default



value is the range next to the subgroup identifiers unless you selected multiple ranges using the control key.

c. *Chart Names & Labels Page*
    i. *Name of Chart:* This is very important. Decide what you want to call the chart. This will be the name of the sheet that contains the chart in your workbook.
    ii. *Control Chart Title:* This is the title that goes on the control chart. The default value is "p Control Chart."
    iii. *Y-Axis Label:* This is the vertical axis label. The default value is "% Defective."
    iv. *X-Axis Label:* This is the horizontal axis label. The default value is "Subgroup Number"
    v. *Data in:* Select columns or rows depending on how the data is entered into the spreadsheet. The default depends on the range selection on the worksheet.

d. *Control Limits & Other Options Page (shown to the right)* .



    i. *Test for Control:* There are two options: points beyond the limits and the rules of seven (seven in a row above or below the average, or seven in a row trending up or trending down).
    ii. *Automatic Update of Limits?*: This determines if the control limits are automatically updated when you add additional data to the chart. Select "Yes" if you want the control limits to automatically update; no if you don't want the limits to automatically update. The default is yes.
    iii. *Print Average/Limits:* Selecting "On Avg. and Limits" will print these on the lines in the chart. Selecting "In Chart Title" will print the values in the chart title.
    iv. *Target for Average:* This is the target value for the variable. It is not required.
    v. *Dates of Data Collection:* Add the starting date and ending dates of data collection. These dates are optional. If entered, they will appear in a dialog box in the lower left-hand corner of the chart.
    vi. *Rounding to Use in Titles:* This the rounding to use for the average and control limits printed in the title. The default value is determined by the program.

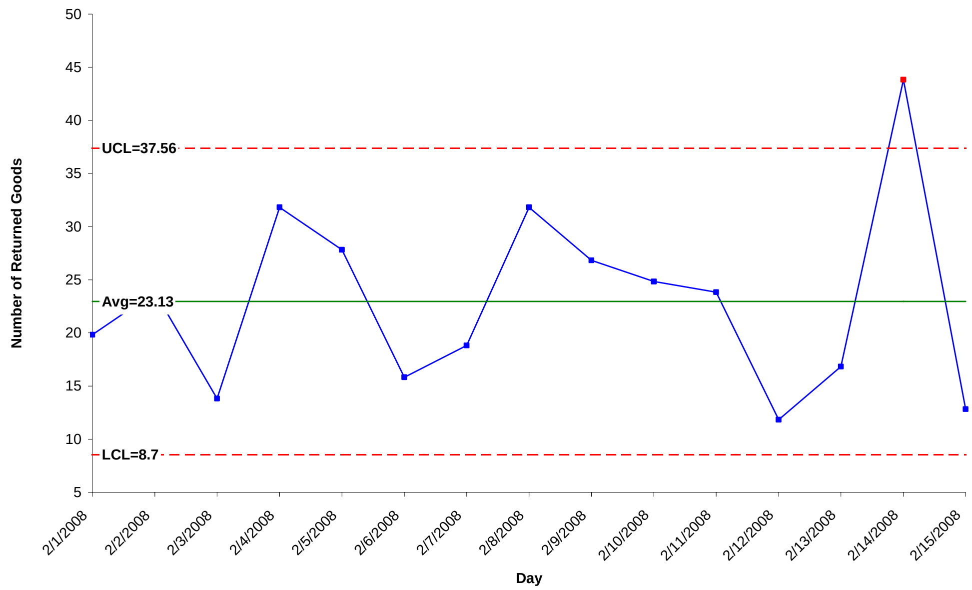
5. Select OK.
6. The c control chart is generated on a new chart sheet. The c control chart for the example data is shown below.

The values of c are plotted over time. The average ($\bar{c}$), the upper control limit (UCL), and the lower control limit (LCL) are calculated using the equations below. k is the number of subgroups used in the calculations (k = 15 in the example). The average is plotted as a solid green line and the control limits are plotted as red dashed lines. The values for the average and control limits, along with the subgroup size, are printed on the chart or in the chart title depending on the option selected. Out of control points are shown in red.

$$\bar{c} = \frac{\sum c}{k} \quad UCL = \bar{c} + 3\sqrt{\bar{c}} \qquad LCL = \bar{c} - 3\sqrt{\bar{c}}$$

These control limits are valid only if $\bar{c} > 3$. The program will automatically use the Poisson Distribution to determine the control limits if the average is less than 3.

c Control Chart



The np control chart can be updated when new data is added by selecting the Update Chart toolbar on the SPC Menu. You may also change current options on the chart by selecting the Options icon from the SPC toolbar. You can also split control limits, start the chart at a new point, and add comments (see the Single Point Actions section in this manual) as well as remove all out of control points or set the range to base the control limits on (see the All Points Action section in this manual).

u Control Charts

A u control chart is used to monitor the variation in the number of defects. A defect occurs when something does not meet a preset specification (operational definition). A u control chart is used with counting type attributes data (e.g., 0, 1, 2, 3). These are whole numbers. You can not have 1/2 defect. In addition, to use a u control chart, two other conditions must be true:

- The opportunity for defects to occur must be large.
- The actual number that occurs must be small.

With a u control chart, we are often looking at an area of opportunity for defects to occur. A u control is similar to a c control chart, except that the area of opportunity for defects to occur is not constant.

For more information on u control charts, please see our July 2007 newsletter on u control charts on our website (http://www.spcforexcel.com/articleslist.htm). An example of how to construct a u control chart is shown below.

Example

You have a new device that is being used in a hospital. It rarely produces one kind of infection, but you want to track how often that happens. You measure the number of days that the device is in use each month and the number of infections that occurred. The data is shown below. You decide to make your inspection unit 1000 days. So, the u control chart will be monitoring the number of infections per 1000 device-days. Using the data below, the program will automatically determine the subgroup size (n) and the values of u. For Jan, n = number of device-days/1000 = 2500/1000 = 2.5 and u = c/n = 5/2.5 = 2.0

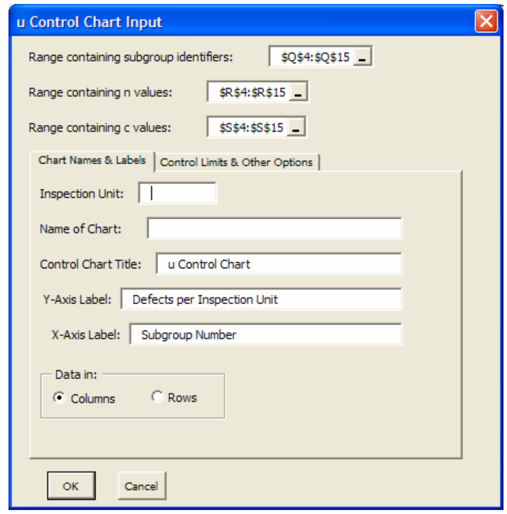
1. Enter the data into a spreadsheet as shown below.

Month	Number of Device-Days	No. of Infections
Jan	2500	5
Feb	4000	4
Mar	1000	3
Apr	3500	5
May	3000	3
Jun	4500	6
Jul	2500	4
Aug	3500	2
Sep	2000	6
Oct	3250	3
Nov	3500	4
Dec	3000	5

2. Select the dates (the shaded area).
3. Select the following from the SPC Menu: Attribute Control Chart → u Control Chart
4. Fill out the u Control Chart Input form shown on the next page.. The ranges are input at the top of the form; there are two pages on the form: Chart Names & Labels and Control Limits & Other Options.
    a. *Range containing the subgroup identifiers:* This is the range containing the subgroup numbers (dates in the above example). The default value is the range selected on the worksheet prior to selecting the attribute control option on the toolbar.
    b. *Range containing the n values:* This is the range containing the subgroup size (n). The default value is the range next to the subgroups unless you selected multiple ranges using the control key. Note that is "n" is the raw subgroup size and does not account for the size of an inspection unit.

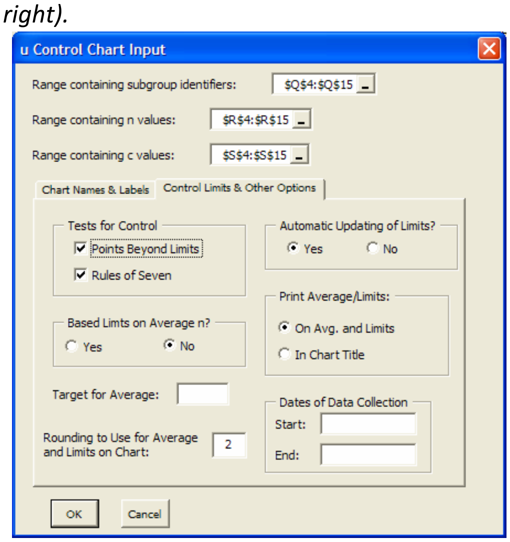
c. *Range containing the c values:* This is the range containing the number of defects (c). The default value is the range next to the n values unless you selected multiple ranges using the control key.

d. *Chart Names & Labels Page*



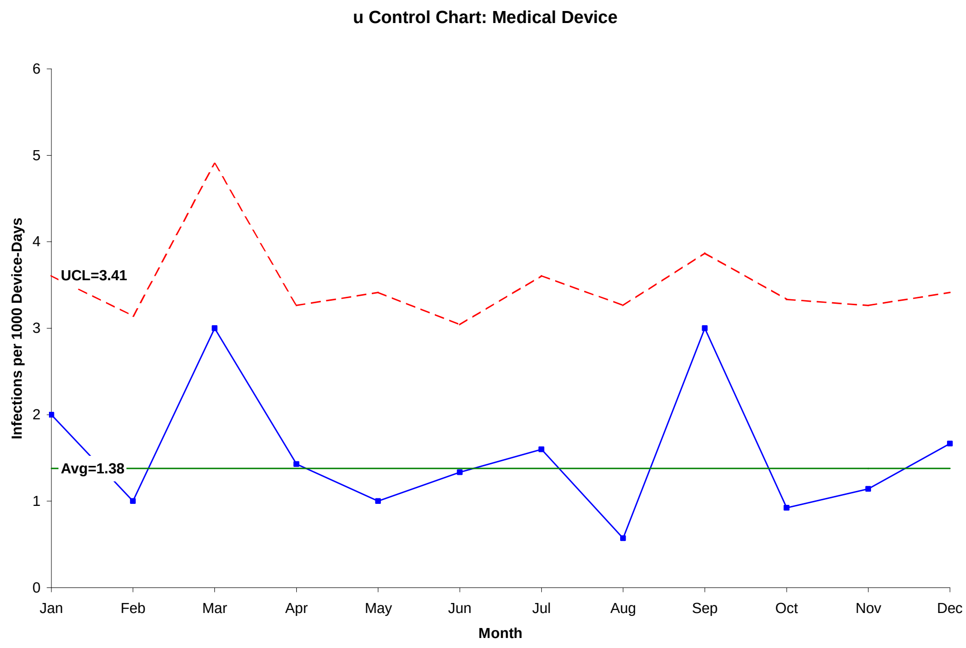
   vii. *Inspection Unit:* Enter the size of the inspection unit. In this example, we selected 1000. You can pick any number although the choice impacts the scaling of the chart.
   viii. *Name of Chart:* This is very important. Decide what you want to call the chart. This will be the name of the sheet that contains the chart in your workbook.
   ix. *Control Chart Title:* This is the title that goes on the control chart. The default value is "p Control Chart."
   x. *Y-Axis Label:* This is the vertical axis label. The default value is "% Defective."
   xi. *X-Axis Label:* This is the horizontal axis label. The default value is "Subgroup Number"
   xii. *Data in:* Select columns or rows depending on how the data is entered into the spreadsheet. The default depends on the range selection on the worksheet.

e. *Control Limits & Other Options Page (shown to the right).*



   xiii. *Test for Control:* There are two options: points beyond the limits and the rules of seven (seven in a row above or below the average, or seven in a row trending up or trending down).
   xiv. *Automatic Update of Limits?*: This determines if the control limits are automatically updated when you add additional data to the chart. Select "Yes" if you want the control limits to automatically update; no if you don't want the limits to automatically update. The default is yes.
   xv. *Based Limits on Average n?* Select no to change the limits each time the subgroup size changes. Select Yes to base the limits on the average subgroup size. The default value is No.
   xvi. *Print Average/Limits:* Selecting "On Avg. and Limits" will print these on the lines in the chart. Selecting "In Chart Title" will print the values in the chart title.
   xvii. *Target for Average:* This is the target value for the variable. It is not required.
   xviii. *Dates of Data Collection:* Add the starting date and ending dates of data collection. These dates are optional. If entered, they will appear in a dialog box in the lower left-hand corner of the chart.
   xix. *Rounding to Use in Titles:* This the rounding to use for the average and control limits printed in the title. The default value is determined by the program.

5. Select OK.
6. The u control chart is generated on a new chart sheet. The u control chart for the example data is shown below.



The u values are plotted over time. The average ($\overline{u}$), the upper control limit (UCL) and the lower control limit (LCL) are calculated using the equations below. There is no LCL in this example (it is negative). The average is plotted as a green solid line and the control limits are plotted as red dashed lines. The values for the average and control limits (based on the average subgroup size, $\overline{n}$) are printed on the chart or in the chart title depending on the option selected. Out of control points are in red.

$$\overline{u} = \frac{\sum c}{\sum n} \qquad UCL = \overline{u} + 3\sqrt{\frac{\overline{u}}{n}} \qquad LCL = \overline{u} - 3\sqrt{\frac{\overline{u}}{n}}$$

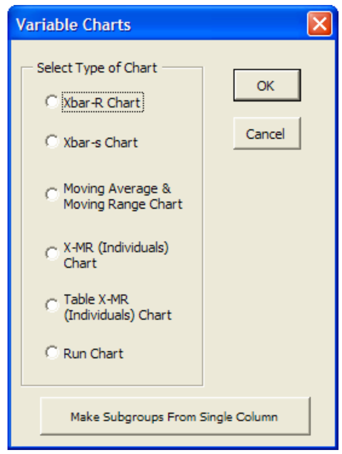
The control limits will be based on the actual subgroup size for each point. If the subgroup size varies, the control limits will also vary (as shown in the above example). These control limits are valid only if $\overline{c} > 3$. The program will automatically use the Poisson Distribution to determine the control limits if the average is less than 3.

The u control chart can be updated when new data is added by selecting the Update Chart toolbar on the SPC Menu. You may also change current options on the chart by selecting the Options icon from the SPC toolbar. You can also split control limits, start the chart at a new point, and add comments (see the Single Point Actions section in this manual) as well as remove all out of control points or set the range to base the control limits on (see the All Points Action section in this manual).

Variable Control Charts

This program constructs the variable control charts listed in the form below. The data entry depends on the type of control chart you are using. T here is also an option to make subgroups from data in a single column. This option, along with the variable control charts, is described in this section.

You access this feature by selecting the variable control chart icon (red) from the SPC menu. You will then see the form below. Select the type of chart you want to make. Each variable control chart is described below



Xbar-R Control Chart, Xbar-s Control Charts, and Moving Average/Moving Range Charts

An $\overline{X}$ -R control chart is used to examine the variation in variables data. Variables data are "measurements" (e.g., height, weight, time, dollars, and density). This control chart is used when you have lots of data and a method of rationally subgrouping the data. For more information on $\overline{X}$ -R control charts, please see our March 2005 and April 2005 newsletters on $\overline{X}$ -R control charts on our website (http://www.spcforexcel.com/articleslist.htm).

The $\overline{X}$ -s control chart is identical to the $\overline{X}$ -R control chart except that the subgroup standard deviation is used instead of the range. The steps in developing both control charts are the same in the program.

A moving average/moving range (MA/MR) chart is very similar to the $\overline{X}$ -R chart. The only major difference in constructing a MA/MR chart is in how the subgroups are formed. The MA/MR chart reuses data. You use a MA/MR chart when you have infrequent data that is not normally distributed. The steps in constructing the moving average/moving range chart is the same as the $\overline{X}$ -R control charts. Please refer to the end of this section to see how the subgroups are formed using a Moving Average/Moving Range chart.

An example of an $\overline{X}$ -R control chart is given below.

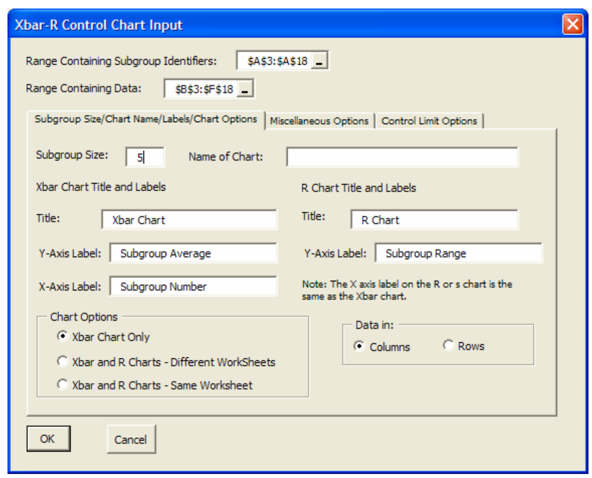
<u>Example</u>

A company wants to monitor sales. They have daily sales and decide to use an Xbar-R chart to monitor sales. The data is given below. A subgroup consists of the daily sales for one week.

1. Enter the data into a spreadsheet as shown below.

Week Number	Monday	Tuesday	Wednesday	Thursday	Friday
1	33.0	33.1	26.4	28.3	28.9
2	30.0	30.1	29.2	31.5	28.4
3	31.8	29.4	28.0	26.9	32.2
4	28.5	34.0	33.6	29.7	33.4
5	27.2	27.6	30.6	30.1	27.4
6	30.5	30.5	28.1	37.7	28.7
7	35.4	31.3	27.8	31.3	33.0
8	33.6	33.3	26.4	32.4	34.1
9	35.8	34.1	30.1	30.3	26.1
10	30.4	32.6	32.5	25.2	32.1
11	26.9	32.4	29.0	26.8	29.3
12	28.0	28.2	25.5	31.1	34.4
13	29.1	31.6	29.0	33.1	30.9
14	26.4	30.8	34.0	27.0	31.7
15	27.4	26.0	28.2	27.9	27.3

2. Select the data (the shaded area)
3. Select the following from the SPC Menu: Variable Control Chart → Xbar-R chart
4. Fill out the Xbar-R Control Chart Input form. The ranges are input at the top of the form; there are three pages on the form: Subgroup Size/Chart Name/Labels/Chart Options, Miscellaneous Options, and Control Limit Options.

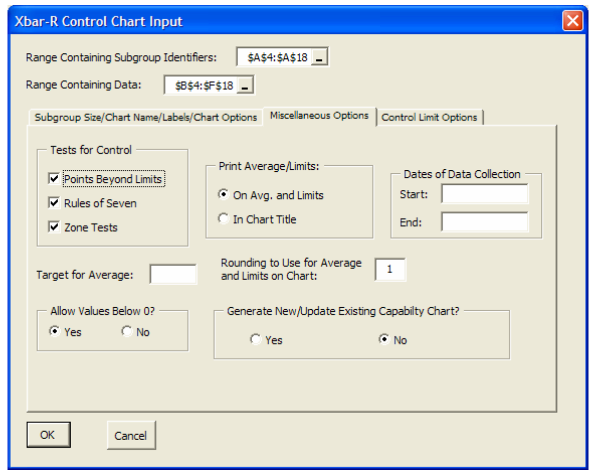


   a. *Range Containing the Subgroup Identifiers:* This is the range that contains the subgroup numbers. The default value is the first column in the range you selected prior to selecting the variable control chart option in the toolbar.
   b. *Range Containing the Data:* This is the range containing the data. The default range is the range you selected prior to selecting the variable control chart option excluding the first column in the range.
   c. *Subgroup Size/Chart Name/Labels/Chart Options Page*
      i. *Subgroup Size:* This is the subgroup size. The default value is the number of columns minus one in the range you selected prior to selecting the variable control chart option in the toolbar. THIS VALUE DETERMINES WHAT DATA IS INCLUDED.
      ii. *Name of Chart:* This is very important. Decide what you want to call the chart. This will be the name of the sheet that contains the chart in your workbook.
      iii. *Xbar Chart Title and Labels*
         - *Title:* This is the title that goes on the control chart. The default title is Xbar Chart.
         - *Y-Axis Label:* This is the vertical axis label. The default label is Subgroup Average.
         - *X-Axis Label:* This is the horizontal axis label. The default label is Subgroup Number.
      iv. *R Chart Title and Labels*
         - *Title:* This is the title that goes on the range chart. The default title is R Chart.
         - *Y*-Axis Label: This is the vertical axis label. The default label is Subgroup Range.

v. *Chart Options*
   - *Xbar Chart Only:* only the $\overline{X}$ chart is constructed. This is the default option
   - *Xbar and R Charts – Different Worksheets:* Both the $\overline{X}$ and range charts are constructed but on different worksheets.
   - *Xbar and R Charts – Same Worksheet:* Both the $\overline{X}$ and range charts are constructed on the same worksheet.

vi. *Data in:* Select columns or rows depending on how the data is entered into the spreadsheet. The program selects one or the other depending on the range selected on the worksheet.

d. *Miscellaneous Options Page*



i. *Tests for Control:* There are three options for interpreting the charts for control: points beyond the limits, the rules of seven (seven in a row above or below the average or trending up or down) and the zone tests (zones A, B, C, stratification, mixtures). The zone tests are not applied to range or standard deviation charts. If an out of control situation is detected, the points on the chart will be in red.

ii. *Target for Average:* This is the target value for the variable. It is not required.

iii. *Print Average/Limits:* There are two options.
   - *On Avg. and Limits*: This option prints the average and control limits values on the lines. This is the default option.
   - *In Chart Title:* This option prints the values in the control chart title.

iv. *Allow values below 0?:* Sometimes, it is not possible for the variable to have values below 0. If that is the case, select “No” for this option. The default value is “Yes.”

v. *Generate New/Update Existing Capability Chart?* Select Yes to do a process capability analysis. The default option is No. See the Process Capability section in this manual for more information.

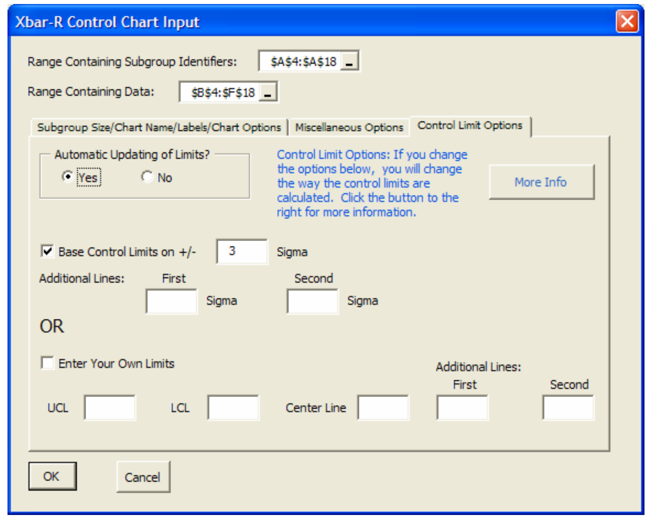
vi. *Dates of Data Collection:* Add the starting date and ending dates of data collection. These dates are optional. If entered, they will appear in a dialog box in the lower left-hand corner of the chart.

vii. *Rounding to Use in Titles:* This is the rounding to use for the average and control limits printed in the title. If the first cell in the range has been formatted, this format is used. If not, the value entered here is used for rounding. The program will estimate the rounding in the data.

e. *Control Limits Options Page*

i. *Update of Limits?*: This determines if the control limits are automatically updated when you add additional data to the chart. Select “Yes” if you want the control limits to automatically update; No if you don’t want the limits to automatically update. The default value is Yes.

ii. The rest of this dialog box is used only if you want to change the default way the program calculates the control limits. The program uses the equations given in this manual for the +/- three sigma limits. There are two other options you have:



- Base control limits on +/- "x" sigma: You can select the sigma limits you want to use in the control chart. The default value is 3. ***DO NOT CHANGE ANYTHING IF YOU WANT THE PROGRAM TO USE THE STANDARD CONTROL LIMIT EQUATIONS.*** You can also add two additional lines to the charts (above and below the average). Any values entered for these additional lines are ignored if the 3 sigma limits are being used. If you use any other value than 3 sigma for the control limits, the zone tests for out of control points will not be applied since it is no longer valid.
- Enter your own limits: You may also enter your own values for the X chart control limits. An additional two lines can also be added to the chart. The values entered here must be between the average and the UCL entered. The program will add them automatically to the area between the average and the LCL.

Select OK and the control charts are generated as shown on the next page.

The $\overline{X}$-R chart is really two charts. The top chart is the $\overline{X}$ chart. This chart looks at the variation in subgroup averages. The subgroup average is the average of the individual results in the subgroup. The bottom chart is the range chart. The subgroup range is the largest result minus the smallest result in the subgroup.

The values of $\overline{X}$ and the range are plotted over time. The average and control limits for both charts are calculated using the equations below.

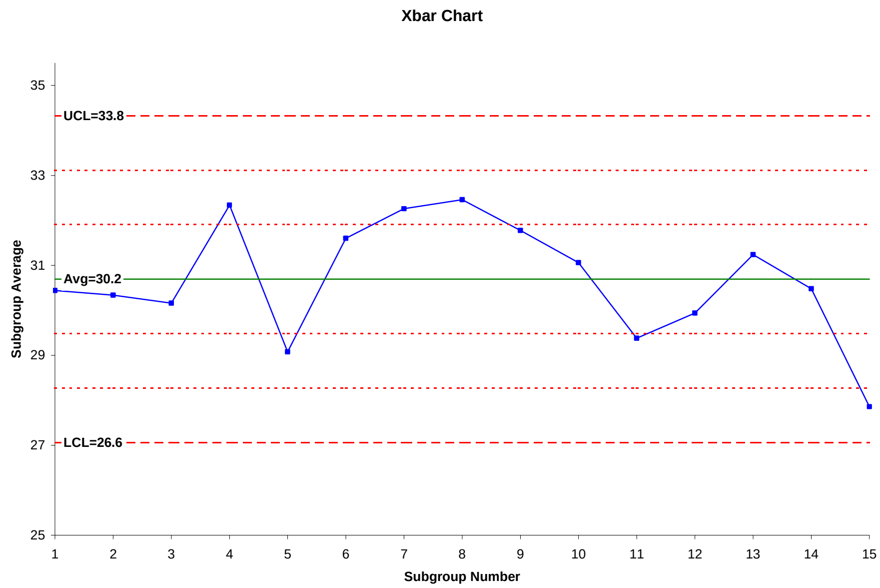
$\overline{X}$ Chart Equations:

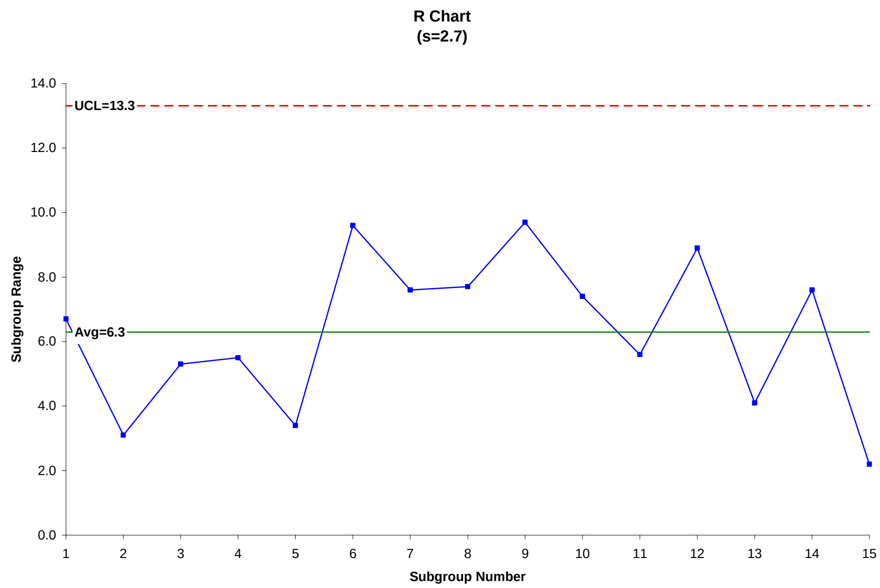
$$\overline{\overline{X}} = \frac{\sum \overline{X}}{k} \qquad UCL = \overline{\overline{X}} + A_2\overline{R} \qquad LCL = \overline{\overline{X}} - A_2\overline{R}$$

Range Chart Equations:

$$\overline{R} = \frac{\sum R}{k} \qquad UCL = D_4\overline{R} \qquad LCL = D_3\overline{R} \qquad \hat{\sigma}' = \frac{\overline{R}}{d_2}$$

The average is plotted as a green solid line and the control limits are plotted as red dashed lines on both charts. For the equations below, k is the number of subgroups. A_2, D_3, and D_4, d_2 are constants used in the calculations for charts. See the newsletter on our website for more information about these constants. The values for the average and control limits are printed on the respective charts





A $\overline{X}$ -s chart is very similar to the $\overline{X}$ -R chart. Instead of using the subgroup range, the $\overline{X}$ -s chart uses the subgroup standard deviation to determine process variability. It is usually used when your subgroup is greater than or equal to 10. For the equations below, k is the number of subgroups. A_3, B_3, and B_4, C_4 are constants used in the calculations for the charts.

$\overline{X}$ Chart Equations:

$$\overline{\overline{X}} = \frac{\sum \overline{X}}{k} \qquad UCL = \overline{\overline{X}} + A_3\overline{s} \qquad LCL = \overline{\overline{X}} - A_3\overline{s}$$

s Chart Equation:

$$\bar{s} = \frac{\sum s}{k} \qquad UCL = B_4\bar{s} \qquad LCL = B_3\bar{s} \qquad \hat{\sigma}' = \frac{\bar{s}}{c_4}$$

The control limit equations for the Moving Average/Moving Range chart are identical to the $\bar{X}$-R chart. The difference in these two types of chart is how you subgroup the data. The MA/MR chart reuses data. For example, the data below shows the accounts receivable for each week. The data could be regrouped into subgroup sizes of three using a MA/MR chart. The first subgroup for the MA/MR chart is formed using the first three results (for the weeks of 2/5, 2/12 and 2/19. The second subgroup for the MA/MR chart uses the weeks of 2/12 and 2/19 and then adds in the week of 2/26. This continues for each of the remaining samples. You use a MA/MR chart when you have infrequent data that is not normally distributed.

Week of	Accounts Receivable		Subgroup Number	1	2	3
2/5	110		1	110	104	98
2/12	104		2	104	98	112
2/19	98		3	98	112	113
2/26	112		4	112	113	100
3/5	113		5	113	100	89
3/12	100		6	100	89	113
3/19	89		7	89	113	109
3/26	113		8	113	109	105
4/2	109		9	109	105	108
4/9	105		10	105	108	95
4/16	108		11	108	95	101
4/23	95		12	95	101	98
4/30	101		13	101	98	100
5/7	98		14	98	100	105
5/14	100		15	100	105	103
5/21	105		16	105	103	99
5/28	103		17	103	99	112
6/4	99		18	99	112	98
6/11	112					
6/18	98					

These control charts can be updated when new data is added by selecting the Update Chart toolbar on the SPC Menu. You may also change current options on the chart by selecting the Options icon from the SPC toolbar. You can also split control limits, start the chart at a new point, and add comments (see the Single Point Actions section in this manual) as well as remove all out of control points or set the range to base the control limits on (see the All Points Action section in this manual).

X-mR (Individuals) Control Charts

An individuals control chart (with a moving range of two) is used to examine the variation in variables data. Variables data are "measurements" (e.g., height, weight, time, dollars, and density). *This chart is used when you have limited data* (for example, one data point per day or per week). It is also useful when data are difficult to obtain. *To use this chart, the individual measurements should be normally distributed*, i.e., a histogram of the individual measurements is bell-shaped.

For more information on X-mR control charts, please see our October 2006 newsletter on X-mR control charts on our website (http://www.spcforexcel.com/articleslist.htm). An example of an individuals control chart is given below.

Example

In this example, the dollar value of accounts receivable past due 90 days is being monitored. We want to use an X-mR control chart to analyze the process. The data collected is shown below.

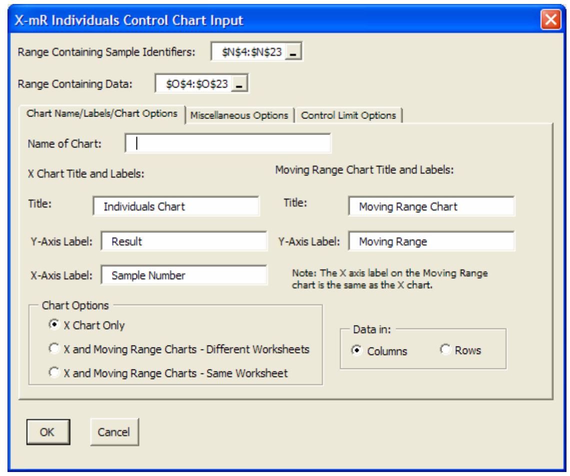
1. Enter the data into a spreadsheet as shown below.

Week of	Accounts Receivable
2/5/2008	110
2/12/2008	104
2/19/2008	98
2/26/2008	112
3/4/2008	113
3/11/2008	100
3/18/2008	89
3/25/2008	113
4/1/2008	109
4/8/2008	105
4/15/2008	108
4/22/2008	95
4/29/2008	101
5/6/2008	98
5/13/2008	100
5/20/2008	105
5/27/2008	103
6/3/2008	99
6/10/2008	112
6/17/2008	98

2. Select the weeks (the shaded area)
3. Select the following from the SPC Menu: Variable Control Chart → Xm-R chart
4. Fill out the X-mR Control Chart Input form below. The ranges are input at the top of the form; there are three pages on the form: Subgroup Size/Chart Name/Labels/Chart Options, Miscellaneous Options, and Control Limit Options.
    f. *Range Containing the Subgroup Identifiers:* This is the range that contains the subgroup numbers. The default value is the first column in the range you selected prior to selecting the variable control chart option in the toolbar.
    g. *Range Containing the Data:* This is the range containing the data. The default range is the range you selected prior to selecting the variable control chart option excluding the first column in the range.

h. *Subgroup Size/Chart Name/Labels/Chart Options Page*

   i. *Name of Chart:* This is very important. Decide what you want to call the chart. This will be the name of the sheet that contains the chart in your workbook.



   ii. *X Chart Title and Labels*
      - *Title:* This is the title that goes on the control chart. The default title is X Chart.
      - *Y-Axis Label:* This is the vertical axis label. The default label is Subgroup Average.
      - *X-Axis Label:* This is the horizontal axis label. The default label is Subgroup Number.

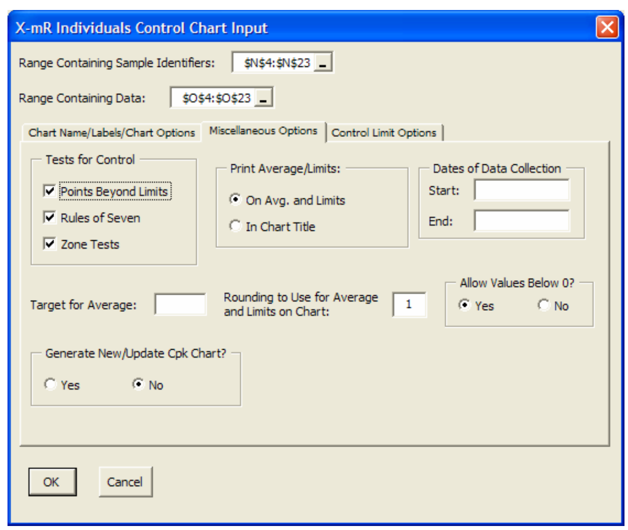
   iii. R Chart Title and Labels
      - *Title:* This is the title that goes on the range chart. The default title is R Chart.
      - *Y*-Axis Label: This is the vertical axis label. The default label is Subgroup Range.

   iv. *Chart Options*
      - *X Chart Only:* only the X chart is constructed. This is the default option
      - *X and Moving Range Charts – Different Worksheets:* Both the X and moving range charts are constructed but on different worksheets.
      - *X and Moving Range Charts – Same Worksheet:* Both the X and moving range charts are constructed on the same worksheet.

   v. Data in: Select columns or rows depending on how the data is entered into the spreadsheet. The program selects one or the other depending on the range selected on the worksheet.

i. Miscellaneous Options Page

   i. Tests for Control: There are three options for interpreting the charts for control: points beyond the limits, the rules of seven (seven in a row above or below the average or trending up or down) and the zone tests (zones A, B, C, stratification, mixtures). The zone tests are not applied to range or standard deviation charts. If an out of control situation is detected, the points on the chart will be in red.



   ii. *Print Average/Limits:* There are two options.
      - *On Avg. and Limits*: This option prints the average and control limits values on the lines. This is the default option.
      - *In Chart Title:* This option prints the values in the control chart title.

iii. *Dates of Data Collection:* Add the starting date and ending dates of data collection. These dates are optional. If entered, they will appear in a dialog box in the lower left-hand corner of the chart.

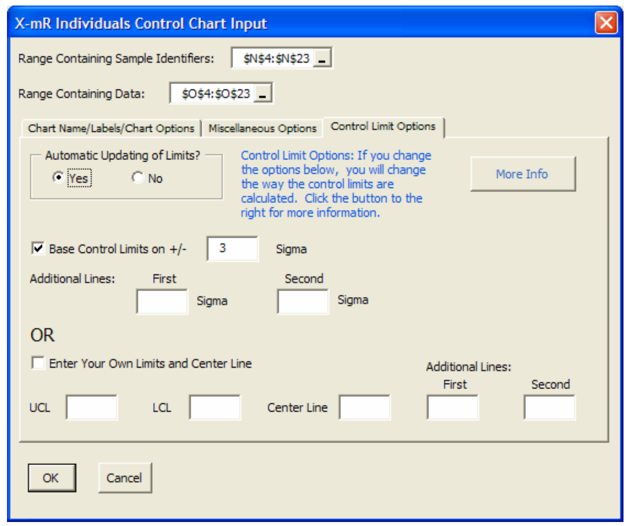
iv. *Target for Average:* This is the target value for the variable. It is not required.

v. *Rounding to Use in Titles:* This is the rounding to use for the average and control limits printed in the title. If the first cell in the range has been formatted, this format is used. If not, the value entered here is used for rounding. The program will estimate the rounding in the data.

vi. *Allow values below 0?:* Sometimes, it is not possible for the variable to have values below 0. If that is the case, select "No" for this option. The default value is "Yes."

vii. *Generate New/Update Existing Capability Chart?* Select Yes to do a process capability analysis. The default option is No. See the Process Capability section in this manual for more information.

j. *Control Limits Options Page*



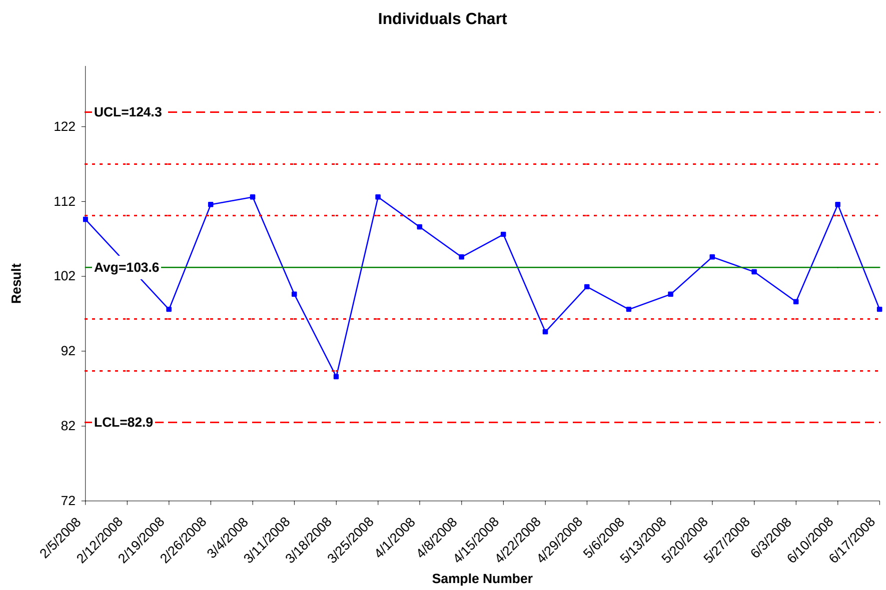
i. *Automatic Updating of Limits?*: This determines if the control limits are automatically updated when you add additional data to the chart. Select "Yes" if you want the control limits to automatically update; No if you don't want the limits to automatically update. The default value is Yes.

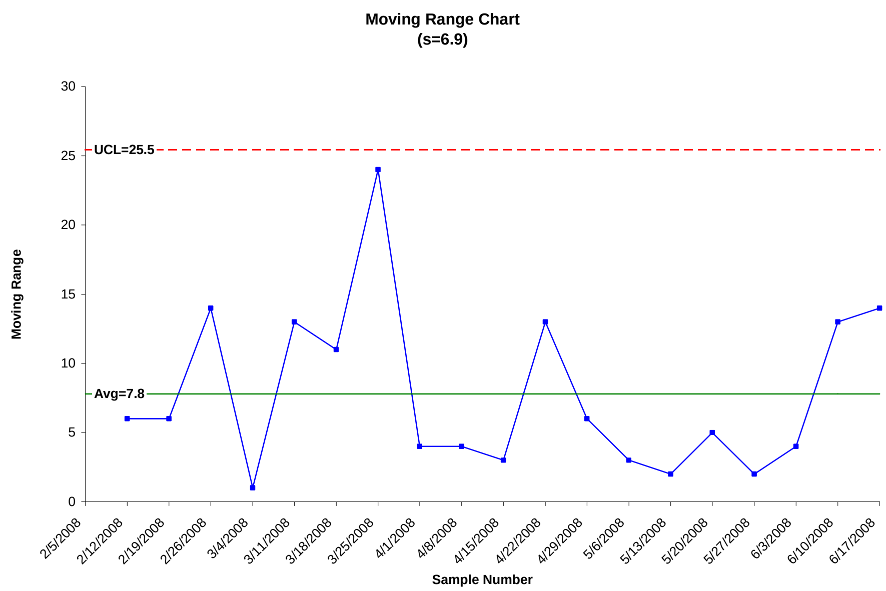
ii. The rest of this dialog box is used only if you want to change the default way the program calculates the control limits. The program uses the equations given in this manual for the +/- three sigma limits. There are two other options you have:

- Base control limits on +/- "x" sigma: You can select the sigma limits you want to use in the control chart. The default value is 3. **DO *NOT CHANGE ANYTHING IF YOU WANT THE PROGRAM TO USE THE STANDARD CONTROL LIMIT EQUATIONS.*** You can also add two additional lines to the charts (above and below the average). Any values entered for these additional lines are ignored if the 3 sigma limits are being used. If you use any other value than 3 sigma for the control limits, the zone tests for out of control points will not be applied since it is no longer valid.
- Enter your own limits: You may also enter your own values for the X chart control limits. An additional two lines can also be added to the chart. The values entered here must be between the average and the UCL entered. The program will add them automatically to the area between the average and the LCL.

5. Select OK.
6. The X-mR chart is generated as shown on the next page.

An X-mR control chart is really two charts. The top chart is the X chart where the individual result (accounts receivable past due 90 days for a week) is plotted. For example, the first point corresponds to $110,000 in past due receivables for the first week (2/6). The second point corresponds to $104,000 for the second week (2/13). The bottom chart is the moving range chart. The moving range between consecutive points is plotted on this chart. For example, the range between accounts receivable past due for 90 days between the week of 2/6 and 2/13 is $110,000 - $104,000 = $6,000. There is no range corresponding to the first data point on the X chart.





The values of X and the moving range are plotted over time. The average and control limits for both charts are calculated using the equations below. The average is plotted as a green solid line and the control limits are plotted as red dashed lines on both charts. For the equations below, k is the number of samples (individual X values).

X Chart Equations:

$$\overline{X} = \frac{\sum X}{k} \qquad UCL = \overline{X} + 2.66\overline{R} \qquad LCL = \overline{X} - 2.66\overline{R}$$

Moving Range Chart Equations:

$$\overline{R} = \frac{\sum R}{k-1} \qquad UCL = 3.27\overline{R} \qquad LCL = None \qquad \hat{\sigma}' = \frac{\overline{R}}{1.128}$$

The values for the average and control limits are printed on the respective charts. The estimated standard deviation is printed on the range chart with the estimated standard deviation = s.

The X-mR chart can be updated when new data is added by selecting the Update Chart toolbar on the SPC Menu. You may also change current options on the chart by selecting the Options icon from the SPC toolbar. You can also split control limits, start the chart at a new point, and add comments (see the Single Point Actions section in this manual) as well as remove all out of control points or set the range to base the control limits on (see the All Points Action section in this manual).

Table X-MR (Individuals) Charts

This option is used to generate multiple individual control charts at the same time. You can either generate the charts one at a time (loops through the dialog box each time) or all at once using chart names and labels that already are entered on the worksheet.

Generate Charts One at a Time

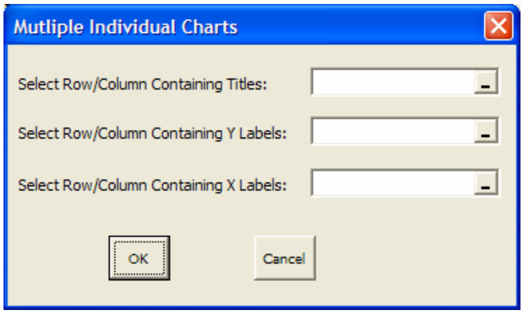
Sample	Result 1	Result 2
1	98.5	93.61
2	101.22	106.38
3	105.99	108.67
4	89.08	98.83
5	105.48	94.57
6	96.55	91.55
7	90.77	95.11
8	96.13	89.41
9	97.16	97.98
10	100.67	98.17

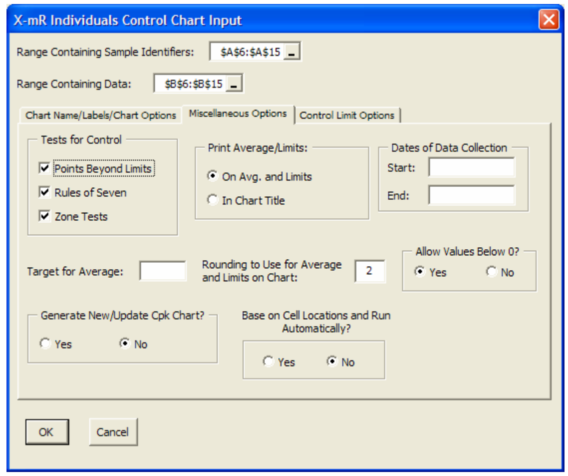
1. Enter the data into a spreadsheet as shown to the right.
   a. The sample numbers are in one column.
   b. The results are in the adjacent columns.
2. Select the sample numbers (shaded area).
3. Select the following from the SPC Menu: Variable Control Chart → Table X-mR (Individuals) Chart
4. Fill out the X-mR Control Chart Input form.
   a. This is the same form described in detail the X-mR chart section of this manual.
5. Select OK.
6. The first chart will be constructed. The program then moves to the next column and shows the X-mR Chart Input form again.
7. This continues until a blank cell is found for sample 1.

Generate Charts All at Once

Name	Chart 1	Chart 2
Y Label	Y	Y
X Label	X	X
Sample	Result 1	Result 2
1	98.5	93.61
2	101.22	106.38
3	105.99	108.67
4	89.08	98.83
5	105.48	94.57
6	96.55	91.55
7	90.77	95.11
8	96.13	89.41
9	97.16	97.98
10	100.67	98.17

1. Enter the data into a spreadsheet as shown to the right.
   a. The sample numbers are in one column.
   b. The results are in the adjacent columns.
   c. The chart names are in one row. These names must be unique and cannot match any existing worksheet tab names.
   d. The Y and X labels must be in one row.
2. Select the sample numbers (shaded area).
3. Select the following from the SPC Menu: Variable Control Chart → Table X-mR (Individuals) Chart
4. Select the following from the SPC Menu: Variable Control Chart → Table X-mR (Individuals) Chart
5. Fill out the X-mR Control Chart Input form.
   a. This is the same form described in detail the X-mR chart section of this manual.
   b. All the charts will be based on the information you enter into the form.
   c. Leave the name of the chart blank.
6. Select the Miscellaneous Options Page (see below).
   a. In the lower right hand side of the dialog box is "Base Labels on Cell Locations and Run Automatically." Select Yes.
   b. You will see the form shown on the next page.
      i. Select the row containing the title (name of chart), the Y labels, and the X labels. Then select OK.
   c. This returns you to the individuals control chart form.
   d. Select OK. This will generate all the charts automatically.





Charts made with the Table X-mR can be updated when new data is added by selecting the Update Chart toolbar on the SPC Menu. You may also change current options on the chart by selecting the Options icon from the SPC toolbar. You can also split control limits, start the chart at a new point, and add comments (see the Single Point Actions section in this manual) as well as remove all out of control points or set the range to base the control limits on (see the All Points Action section in this manual).

Run Charts

A run chart plots the points over time and adds an average line to the chart. It is not a control chart as such but you can apply the rule of seven tests to the run chart. An example of how to make a run chart is shown below.

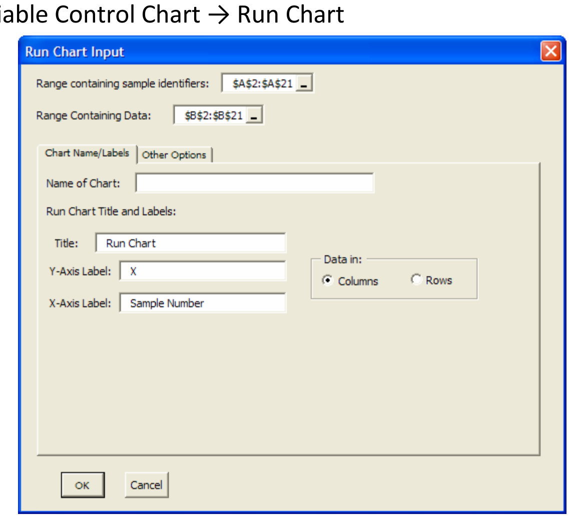
<u>Example</u>

In this example, the dollar value of accounts receivable past due 90 days is being monitored. We want to use a run chart t to analyze the process. The data collected is shown below.

1. Enter the data into a spreadsheet as shown below.

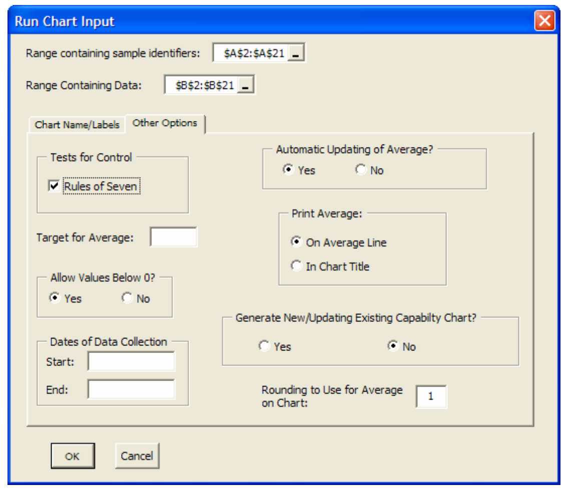
Week of	Accounts Receivable
2/5/2008	110
2/12/2008	104
2/19/2008	98
2/26/2008	112
3/4/2008	113
3/11/2008	100
3/18/2008	89
3/25/2008	113
4/1/2008	109
4/8/2008	105
4/15/2008	108
4/22/2008	95
4/29/2008	101
5/6/2008	98
5/13/2008	100
5/20/2008	105
5/27/2008	103
6/3/2008	99
6/10/2008	112
6/17/2008	98

2. Select the weeks (the shaded area)
3. Select the following from the SPC Menu: Variable Control Chart → Run Chart
4. Fill out the Run Control Chart Input form. The ranges are input at the top of the form; there are two pages on the form: Chart Names/Labels and Other Options.
    a. Range Containing the Sample Identifiers: This is the range that contains the sample identifiers. The default value is the first column in the range you selected prior to selecting the variable control chart option in the toolbar.
    b. Range Containing the Data: This is the range containing the data. The default range is the range you selected prior to selecting the variable control chart option excluding the first column in the range.



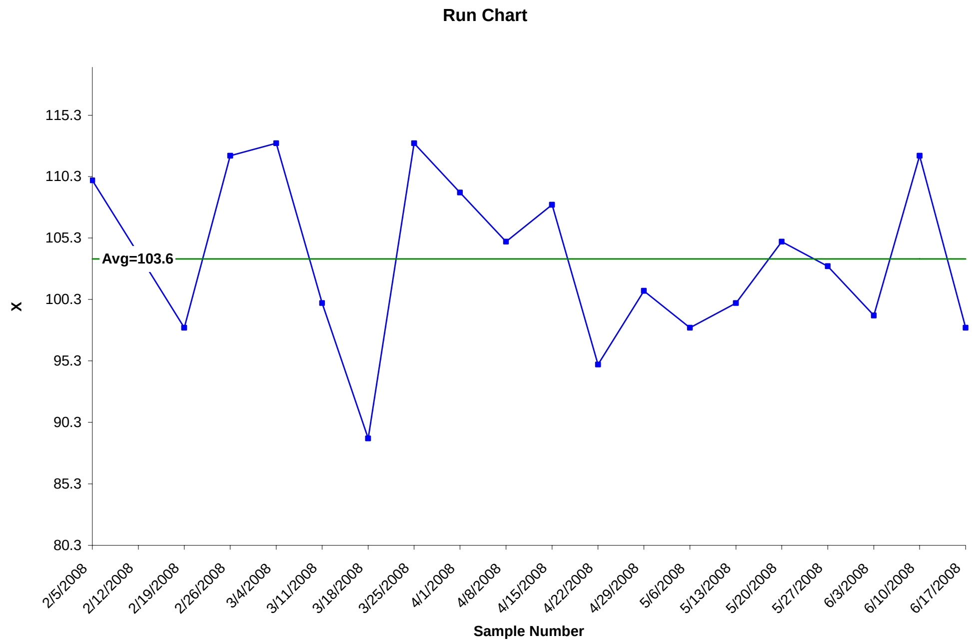
c. Chart Name/Labels Page
    i. Name of Chart: This is very important. Decide what you want to call the chart. This will be the name of the sheet that contains the chart in your workbook.
    ii. Title: This is the title that goes on the l chart. The default title is Run Chart.
    iii. Y-Axis Label: This is the vertical axis label. The default label is X.
    iv. X-Axis Label: This is the horizontal axis label. The default label is Sample Number.
    v. Data in: Select columns or rows depending on how the data is entered into the spreadsheet. The program selects one or the other depending on the range selected on the worksheet.

d. Other Options Page



    i. Tests for Control: There is one option for interpreting the chart for control: the rules of seven (seven in a row above or below the average or trending up or down). If an out of control situation is detected, the points on the chart will be in red.
    ii. Automatic Updating of Average: Yes is the default which updates the average when new data is added to the chart; to prevent the average from changing, select the No option.
    iii. Target for Average: This is the target value for the variable. It is not required.
    iv. Print Average: There are two options.
    v. On Average Line: This option prints the average value on the average line. This is the default option.
    vi. In Chart Title: This option prints the value in the chart title.
    vii. Allow values below 0?: Sometimes, it is not possible for the variable to have values below 0. If that is the case, select "No" for this option. The default value is "Yes."
    viii. Dates of Data Collection: Add the starting date and ending dates of data collection. These dates are optional. If entered, they will appear in a dialog box in the lower left-hand corner of the chart.
    ix. Generate New/Update Existing Capability Chart? Select Yes to do a process capability analysis. The default option is No. See the Process Capability section in this manual for more information.
    x. Rounding to Use in Titles: This is the rounding to use for the average. If the first cell in the range has been formatted, this format is used. If not, the value entered here is used for rounding. The program will estimate the rounding in the data.

5. Select OK.
6. The run chart below will be generated.

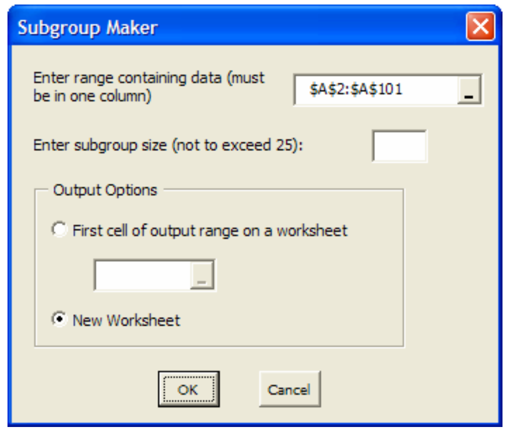


Run c harts can be updated when new data is added by selecting the Update Chart toolbar on the SPC Menu. You may also change current options on the chart by selecting the Options icon from the SPC toolbar. You can also split the average, start the chart at a new point, and add comments (see the Single Point Actions section in this manual) or set the range to base the control limits on (see the All Points Action section in this manual).

Subgroup Maker: Make Subgroups from Column of Numbers

The program has the option to make subgroups from a single column of numbers as shown below.

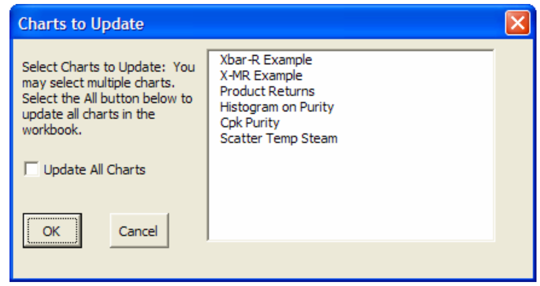
1. Enter the data into a single column.
2. Select the data you want to make into subgroups (the data only, not any sample identifier).
3. Select the following from the SPC Menu: Variable Control Chart → Make Subgroups from Single Column"
4. Fill out the Subgroup Maker form.
    a. *Enter range containing data:* This is the range containing the data; default is the range selected on the worksheet.
    b. Enter the subgroup size, not to exceed 25: This is the subgroup size you want from 2 to 25.
    c. Output Options:
        - *First cell of output range on a worksheet:* enter the cell location on the worksheet where you want the subgroups formed.
        - *New worksheet:* select this option to put the subgroups on a new worksheet.



5. *Select OK:* The subgroups will be generated and place based on your output option. You will then get the Variables Chart dialog box to select the type of chart you want to make and follow the instructions for that chart.

Updating Charts

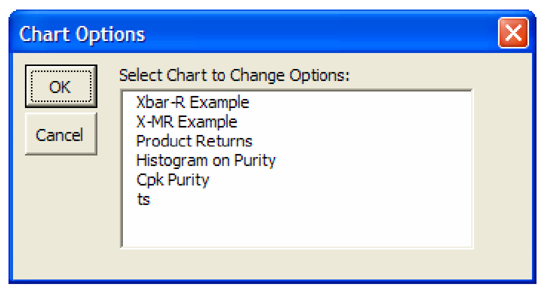
All charts can be updated easily after new data has been entered into the spreadsheet. There is no need to select anything on the worksheet. The program automatically checks to see what new data has been entered. Once you have entered the new data, select the update option from the SPC toolbar. You will get the dialog shown to the right.



The dialog box lists all the available charts for updating in the workbook. You may select multiple charts at once. You can also check the "Update All Charts" option to update all the charts in the workbook.

Changing Chart Options

You make changes to the chart using the options button on the SPC toolbar. When you select this button, you get the dialog box to the right. Select the chart whose options you want to change. The dialog box for that chart then is shown. You can change anything except the name of the chart.

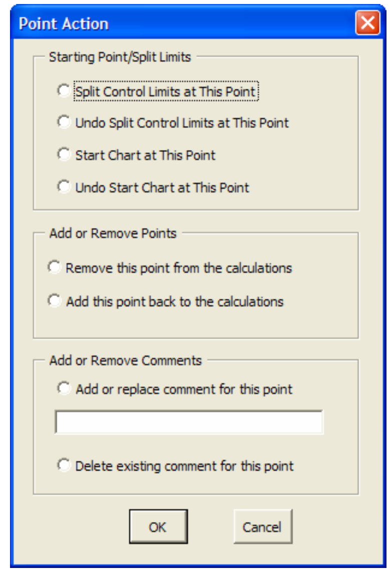


Note on chart title, y-axis label, and x-axis label: To make permanent changes to the chart title or the labels on either axis, you must go through the options shown here. Making changes directly on the chart will not permanently change the title and labels. If you make the changes on the chart and then update the chart, the program will use the stored values.

Single Point Actions

This option is used for action on a single point on a chart including:

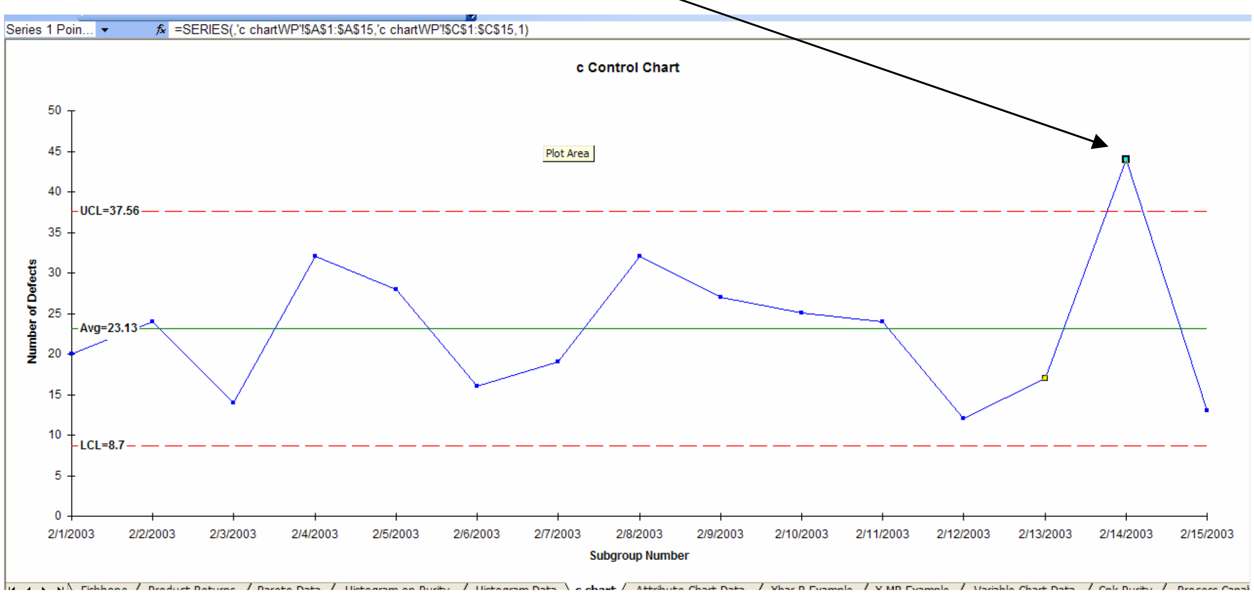
- Splitting control limits at a point
- Removing the split at the point
- Starting the chart at the point
- Undo starting the chart at the point
- Removing a point from the calculations
- Adding a point back to the calculations
- Adding or replacing a comment to the point
- Deleting the existing comment for the point



You must select a point on the chart ***prior*** to selecting the single point option from the menu. To select a single point:

- Select the series by pointing the mouse at the series and clicking
- Select the point you want using the mouse.

Below is an example of a chart with one point selected (this is the c chart data from the example workbook).

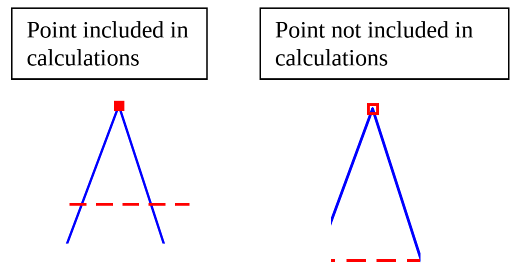


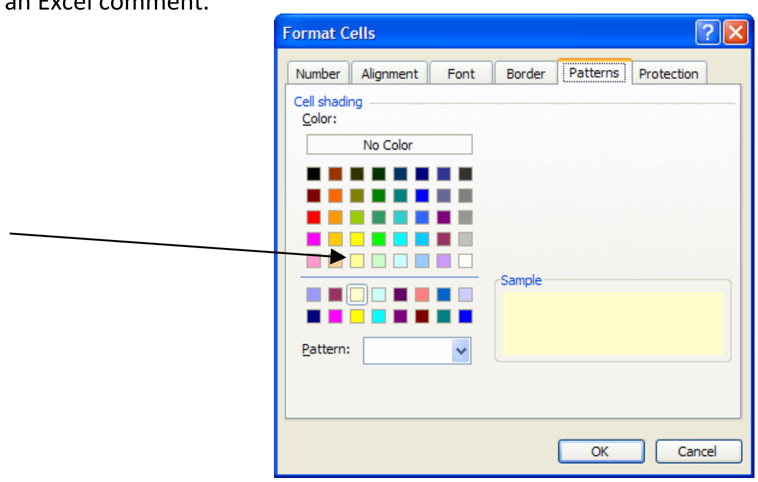
After selecting the point, select the single point action option () from the SPC menu. The form above will appear. Select the option you want and then select OK. The average and limits are recalculated based on your option and the chart is re-made. See the notes on the following page for more information.

Notes on Single Point Actions

Some of the actions on this option will produce minor changes in your workbook. These changes along with other issues on single point actions are described below.

- Control limits can be split multiple times within a chart.
  - ***This will produce a change to your worksheet.*** The cell containing the point will be in italics. This will occur in the first column or row of data.
- ***Starting the chart at a new point will produce a change on your worksheet.*** The cell containing that point will be in bold. This will occur for in the first column or row of data.
- If a point is removed from the calculations, it is still plotted. Its appearance will change - it will just be outlined as shown to the right.
  - ***This will also produce a change to your worksheet.*** The cell containing the point will be shaded a light tan. The shading will occur in the first column or row of data. The shading used is shown to the right.
- Comments added to or deleted from the points must be added through this option. Adding or deleting them manually to the chart will not store the result.
  - ***This will also produce a change to your worksheet.*** The comments will be added to the cell containing that point as an Excel comment.





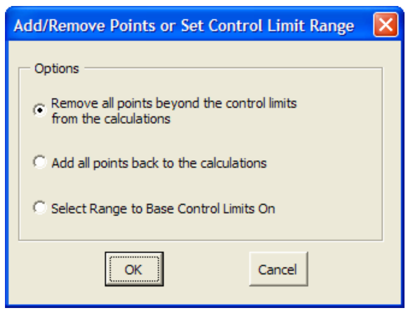
All Points Action

When you select the All Points Action from the SPC toolbar, you will see the dialog box to the right. With this option, you can:

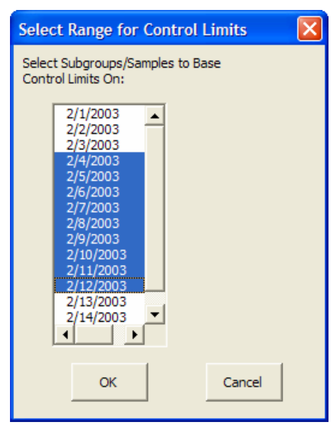
- Remove or add back to the calculations all points *beyond the control limits*.
- Select the range to base control limits on.

A chart must be selected before this option can be used. If you have the two charts on the same worksheet (e.g., $\overline{X}$ -R chart), you must select the $\overline{X}$ chart. You cannot delete points by selecting the range chart.

Once you have selected the chart, select the all points option ( ) on the SPC menu. You will get the form to the right. Select the option you want.



If you select the option "Select Range to Base Control Limits On," you will see a box like the one to the right. This box will contain the subgroup identifiers from your chart. Select the subgroup you want to include in the calculation of averages and control limits. Selecting this option will set the option to automatic update the averages and limits to No.



Note: When removing all points beyond the limits, the average and limits are recalculated. It is possible that additional points will now be beyond the limits. You may have to run this several times to remove all points beyond the limits.

Scatter Diagrams

A scatter diagram is used to show the relationship between two kinds of data. It could be the relationship between a cause and an effect, between one cause and another, or even between one cause and two others. You have the option to select the type of scatter diagram as well as the option to add labels to the chart. An example of a scatter diagram is shown below.

For more information on scatter diagrams, please see our February newsletter on our website (http://www.spcforexcel.com/articleslist.htm).

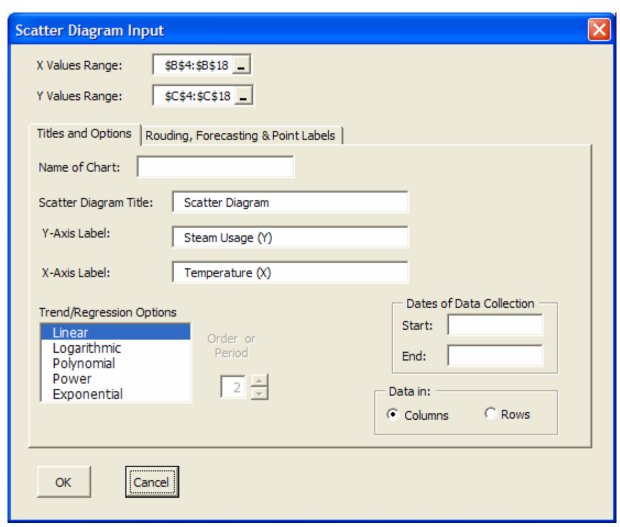
Example

An engineer is examining that data that that relates steam usage in a plant to the atmospheric temperature. The question being answered here is "Does the atmospheric temperature have an effect on steam usage in the plant." For 25 days, data were collected on steam usage and temperature. There are 25 sets of data point. The data is shown below.

1. Enter the data into a spreadsheet as shown below. The data does not have to be in adjacent columns.

Sample No.	Temperature (X)	Steam Usage (Y)
1	35.3	10.98
2	29.7	11.13
3	30.8	12.51
4	58.8	8.4
5	61.4	9.27
6	71.3	8.73
7	74.4	6.36
8	76.7	8.5
9	70.7	7.82
10	57.5	9.14
11	46.4	8.24
12	28.9	12.19
13	28.1	11.88
14	39.1	9.57
15	46.8	10.94

2. Select the data the shaded area, not the sample numbers or headings. We will use the sample headings for labels as shown below.
3. Select the Scatter diagram () option from the SPC Menu.
4. Fill out the Scatter Diagram Input form. The input for the ranges is at the top of the form. There are two pages on the form: Input Data/Titles and Rounding/Forecasting/Point Labels.
    a. *X Values Range:* This is the range containing the X values. The default value is the range you selected prior to selecting the scatter diagram option in the toolbar.



b. *Y Values Range:* This is the range containing the Y values. The default value is the range next to the X values or the second range selected on the worksheet.

5. Titles and Options Page

a. *Name of Chart:* This is very important. Decide what you want to call the chart. This will be the name of the sheet that contains the chart in your workbook.

b. *Scatter Diagram Title:* This is the title that goes on the scatter diagram. The default value is "Scatter Diagram."

c. *Y-Axis Label:* This is the vertical axis label. The default value is the cell contents above the X values range.

d. *X-Axis Label:* This is the horizontal axis label. The default value is the cell contents above the Y values range.

e. *Trend/Regression Options:* Select the regression you want. The options are:

i. Linear

ii. Logarithmic

iii. Polynomial (activates order or period option)

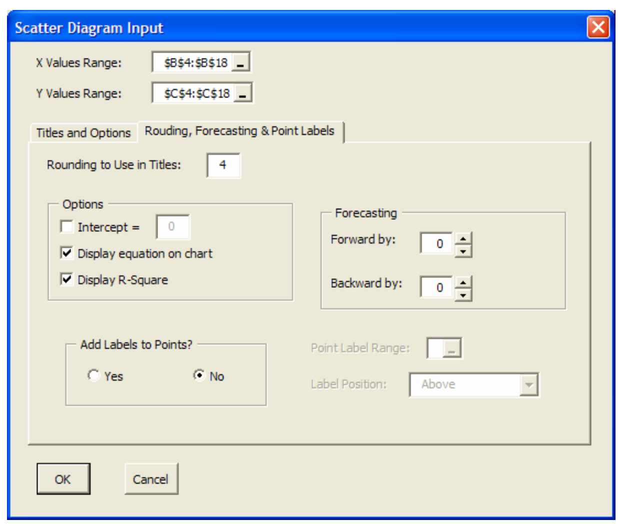
iv. Power

v. Exponential

f. *Dates of Data Collection:* Add the starting date and ending dates of data collection. These dates are optional. If entered, they will appear in a dialog box in the lower left-hand corner of the chart.

g. Data in: Select columns or rows depending on how the data is entered into the spreadsheet. The program selects one or the other depending on the range selected prior to selecting the scatter diagram option on the SPC toolbar.

6. Rounding, Forecasting and Point Labels Page



a. *Rounding to Use in Titles:* This the rounding to use for the probability printed in the title. The default value is 4.

b. Options: There are three options to consider:

i. *Intercept:* check this box if you want to force the y-intercept through 0 or another value

ii. *Display equation on chart:* select this option to display the equation on the chart.

iii. Display R-Square: select this option to display R^2 on the chart.

c. *Forecasting:* You can forecast forward or backward by changing the 0 values in the appropriate boxes.

d. *Add Labels to Point:* Select "Yes" if you want to add labels to the points.

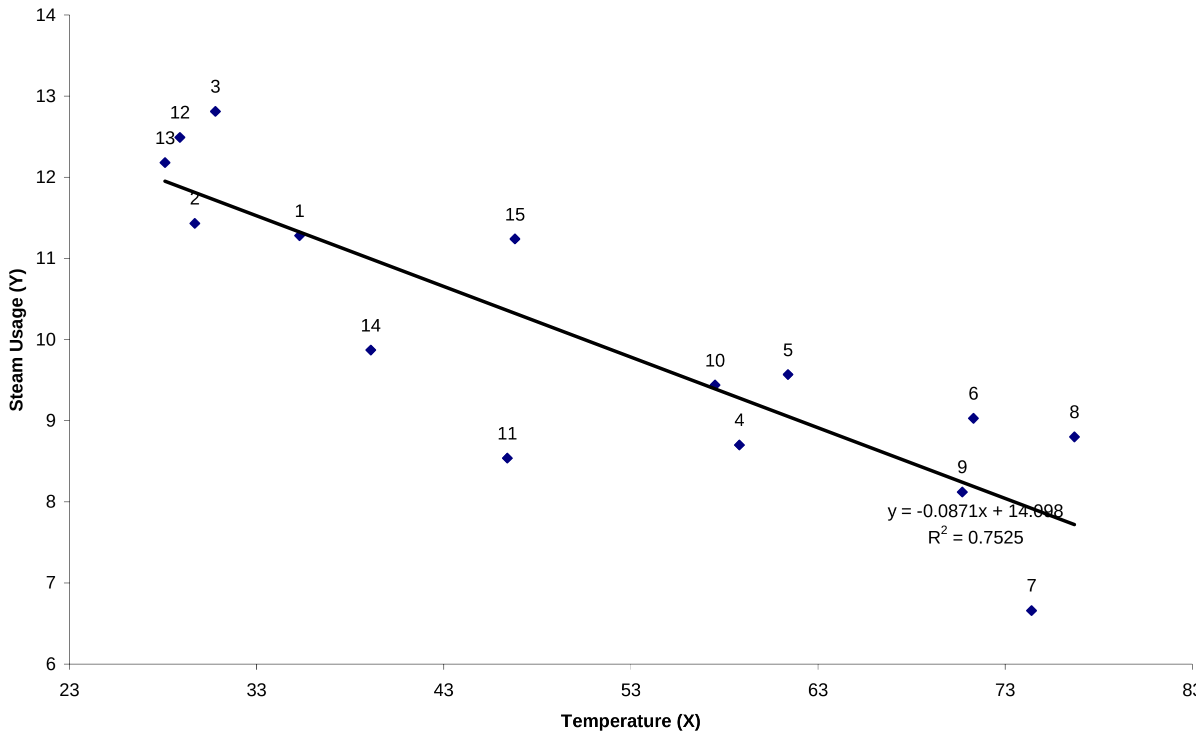
i. *Point Range Label:* This is the range containing the point label ranges and should be equal to the number of points on the chart.

ii. *Label Position:* This determines where to put the labels (above, below, left, center, and right) relative to the point.

7. Select OK

8. The scatter diagram is generated. The scatter diagram for the example data is shown on the next page. The sample numbers were used as labels.

Scatter Diagram (Significant, p = 0)



Each set of data points is charted on a scatter. As can be seen in the figure, there is a negative correlation in the data. This means that as temperature decreases, the steam usage in the plant tends to increase. There can be a positive correlation, a negative correlation, or no correlation.

The program determines if the relationship between the two variables is statistically significant at a probability of 0.05. If there is a significant relationship, the resulting probability is given in the title (in this case, the round goes to p =0). The equation is of the form:

$$y = b_1x + b_0$$

where y is the variable on the y axis, x is the variable on the x axis, b_1 is the slope of the line, and b_0 is where the line crosses the y axis. In the example above, the slope is -0.0871. This means that for each unit increase in x (one degree of temperature in this case), the y value (steam usage in this case) decreases by -0.0871.

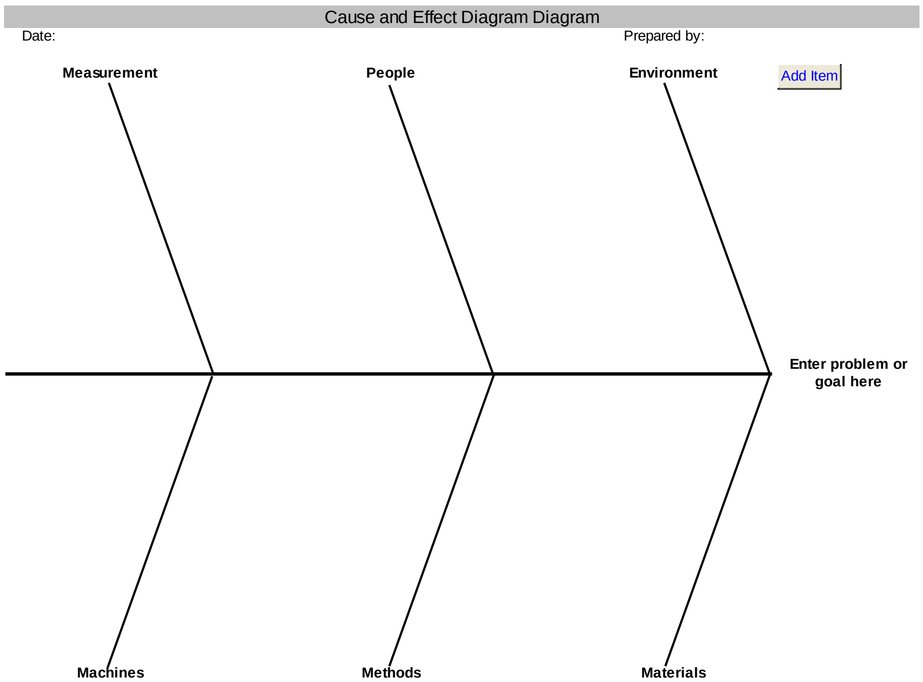
The value of R^2 in the chart is the % of variation in y that is explained by x. In this example, 75% of the variation in steam usage is explained by the variation in temperature.

The scatter diagram can be updated with new data. This program also handles multiple linear regression. You may run this technique to get much more information.

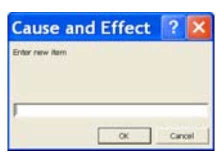
Cause and Effect Diagrams

This program contains a cause and effect (fishbone) diagram. To use this feature, select the cause and effect option on the SPC menu. A blank cause and effect diagram will be inserted into your workbook.

For more information on scatter diagrams, please see our October 2005 and November 2005 newsletters on our website (http://www.spcforexcel.com/articleslist.htm).



You can change the main category headings (measurement, people, etc). You can enter the problem or goal at the head of the fishbone (see above). You will see a button that says "Add Item." Select this button and you will get the form below.

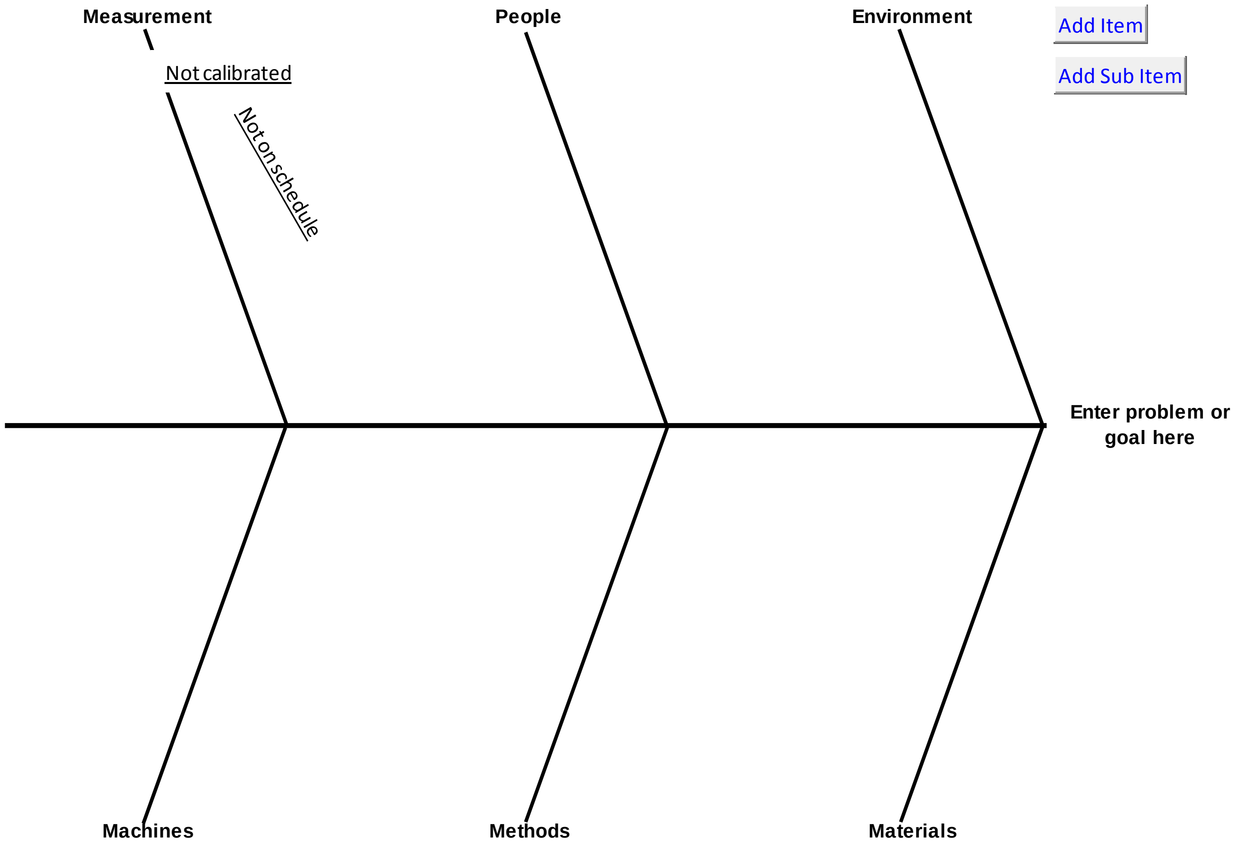


Enter an item to add to the cause and effect diagram. For example, you might enter "Not calibrated." Then select OK. This item is then shown on the chart and you may move it to any location you want as is shown on the next page.

For Excel 2007 users, you will have a button that says "Add Sub Item". Select this button and you will get the same form as above. However, this item will be drawn at any angle and is designed to fit under a previous entry. For example, you might enter "Not on schedule" as a sub item under "Not calibrated" as is shown on the next page.

Cause and Effect Diagram Diagram

Date: Prepared by:



Process Capability

Process capability answers the question: Is the process capable of meeting specifications? Specifications can be set by customers. Specifications could also be standards set by management for a process. For example, the standard for days sales outstanding might be set by leadership to be less than 46 days. One measure of process capability is the Cpk index. Another is Ppk. To determine the process capability, the individual sample results should be normally distributed (the histogram is a bell shaped curve) and the process should be in statistical control.

For more information on process capability, please see our October, November, and December newsletters on our website (http://www.spcforexcel.com/articleslist.htm). An example of process capability is given below.

Note: Process capability depends on the standard deviation. There are several ways the standard deviation can be calculated. For Cpk, the standard deviation is the estimated standard deviation from a range chart. If you data is in one column (or row), a moving range of two is used to calculated an average range and then estimate the standard deviation. If the data is in multiple columns, the program uses the range chart with a subgroup size based on the number of columns to calculate the average range and then estimate the standard deviation. Ppk is determined using the calculated standard deviation based on all the data. Visit the newsletters above for more information.

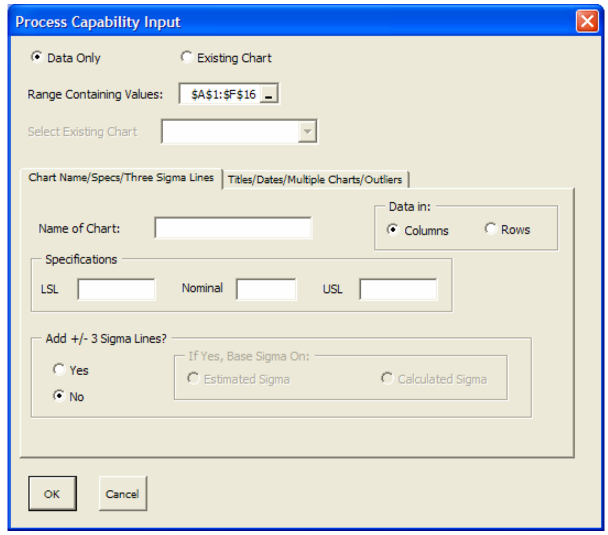
Example

A product has a lower specification limit of 60 and a lower specification limit of 80. Data has been collected and is given below. Is the process capable of meeting the specifications?

1. Enter the data into a spreadsheet as shown below.

81	77	75	74	77	73
77	74	76	75	79	74
74	79	73	75	75	74
75	80	80	79	72	78
73	74	74	73	75	74
77	75	75	72	75	74
76	75	74	74	78	75
76	76	78	77	78	75
74	76	77	76	72	73
79	82	73	75	74	79
77	73	72	75	73	73
76	76	76	75	74	72
76	76	76	74	79	79
75	81	77	74	77	71
84	74	79	70	77	74
73	77	76	74	81	75

2. Select the data on the spreadsheet (the shaded area)
3. Select the process capability icon () from the SPC menu.
4. Fill out the Process Capability Input form. The ranges are input at the top of the form; there are two pages on the form: Chart Names/Specs/Three Sigma Lines and Titles/Dates/Multiple Charts/Outliers



a. *Data or Existing chart?:* Select the option you want. "Data Only" is the default option. When "Data Only" is selected, the "Range Containing Values" is enabled.

b. *Range Containing Values:* This is the range containing the values on which to do the process capability analysis. The default value is the range selected on the spreadsheet before selecting the process capability option. If "Existing Chart" is selected, the "Select Existing Chart" list box is enabled and a list of available charts is given in the list box. Select the chart you want to do the process capability analysis on.

c. Chart Names/Specs/Three Sigma Lines Page

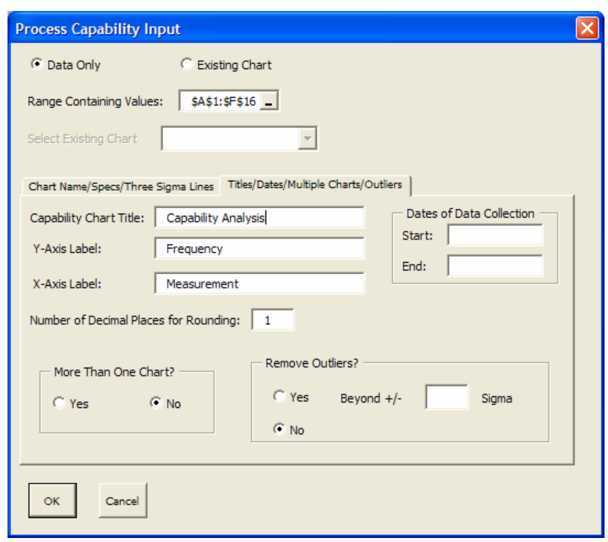
    i. *Name of Chart:* This is very important. Decide what you want to call the chart. This will be the name of the sheet that contains the chart in your workbook. If you select the "Existing Chart" option, the chart automatically will be the name of the existing chart worksheet with Cpk added.

    ii. *Data in:* Select columns or rows depending on how the data is entered into the spreadsheet. The program selects one or the other depending on the range selected prior to selecting the process capability option on the SPC toolbar.

    iii. *Specifications:* Enter the upper specification limit (USL), the lower specification limit (LSL) and the nominal, the target (if desired). Only one specification limit is required.

    iv. *Add +/- 3 Sigma Limits:* In addition to the specifications, you can add the +/- three sigma limits to the chart. The default is No. If you select Yes, you can chose sigma to the estimated sigma from the range chart or the calculated standard deviation of all the data.

d. Titles/Dates/Multiple Charts/Outliers



    i. *Capability Chart Title:* This is the title that goes on the chart. The default value is "Capability Analysis."

    ii. *Y-Axis Label:* This is the vertical axis label. The default value is "Frequency."

    iii. *X-Axis Label:* This is the horizontal axis label. The default value is "Measurement."

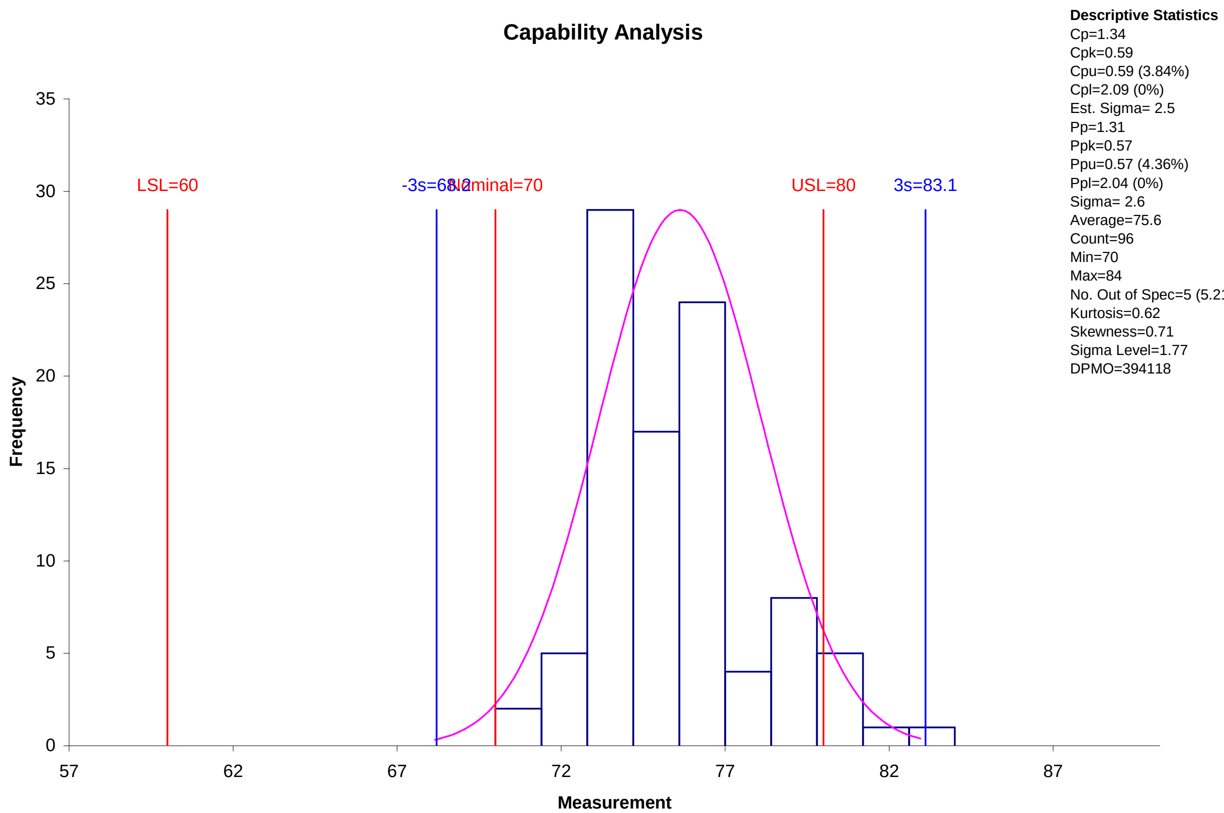
    iv. *Number of Decimal Places for Rounding:* This is the rounding to use for the values in the titles on the chart.

    v. *Dates of Data Collection:* Add the starting date and ending dates of data collection. These dates are optional. If entered, they will appear in a dialog box in the lower left-hand corner of the chart.

    vi. *More Than One Chart?* Select "Yes" if you want to make multiple process capability charts by looping through the dialog box. The program assumes that the next set of data for the process capability analysis is adjacent to the current set. Use one row or one column of data if you are selecting this option. “No” is the default value.

vii. *Remove Outliers?* Select “Yes” if you want to remove outliers from the calculations. Enter the number of standard deviations you want to remove outliers beyond (e.g., beyond +/- 6sigma). The default option is “No.” You have the option of selecting the standard deviation based on the range chart or the calculated standard deviation from all the data.

5. Select OK.
6. The process capability chart will be generated as shown on the next page (using LSL = 60, nominal = 70 and USL = 80 with the three sigma limits added.)



The statistics printed on the right include the following:

- Cp: = (USL-LSL)/6 $\hat{\sigma}'$ where $\hat{\sigma}'$ is the estimated standard deviation from a range or s chart
- Cpk: the minimum of Cpu or Cpl
- Cpu: the capability based on the USL = (USL-$\bar{X}$)/3$\hat{\sigma}'$ where $\bar{X}$ is the overall average (the number in parentheses is the theoretical % greater than the USL)
- Cpl: the capability based on the LSL = ($\bar{X}$-LSL)/3$\hat{\sigma}'$ (the number in parentheses is the theoretical % less than the LSL)
- Est. Sigma = $\hat{\sigma}'$
- Pp: = (USL-LSL)/6s where s is the standard deviation of the measurements
- Ppk: the minimum of Ppu or Ppl

- Ppu: the capability based on the USL = (USL-$\overline{X}$)/3s where $\overline{X}$ is the overall average (the number in parentheses is the theoretical % greater than the USL)
- Ppl: the capability based on the LSL = ($\overline{X}$-LSL)/3s (the number in parentheses is the theoretical % less than the LSL)
- Sigma: = s
- Average: = $\overline{X}$
- Count: = number of data points in the analysis
- No. Out of Spec: = actual number out of specification (number in parentheses is the % out)
- Kurtosis: a measure of the shape of the distribution; please see our April 2008 newsletter on the website for more information (http://www.spcforexcel.com/articleslist.htm).
- Skewness: a measure of asymmetry; please see our April 2008 newsletter on the website for more information (http://www.spcforexcel.com/articleslist.htm).
- Sigma Level: A statistical term that measures how much a process varies from perfection, based on the number of defects per million units.
    - One Sigma = 690,000 per million units
    - Two Sigma = 308,000 per million units
    - Three Sigma = 66,800 per million units
    - Four Sigma = 6,210 per million units
    - Five Sigma = 230 per million units
    - Six Sigma = 3.4 per million units
- DPMO: Defects per million opportunities

Charts made with the process capability analysis tool can be updated when new data is added by selecting the Update Chart toolbar on the SPC Menu. You may also change current options on the chart by selecting the Options icon from the SPC toolbar.

Multiple Process Capability Analysis

This option is used to automatically generate multiple process capability analysis, to remove outliers, adjust specification limits, and generate a summary process capability table.

For more information on process capability, please see our October, November, and December newsletters on our website (http://www.spcforexcel.com/articleslist.htm). Also, please review the information in the Process Capability Analysis portion of this manual. An example of the multiple process capability is given below.

<u>Example</u>

A company has five different parameters that are measured on a part. They want to conduct a process analysis capability on each parameter. The data is shown below. Each chart name represents a different parameter.

1. Enter the data into a spreadsheet as shown below. There must be a row containing the ***unique*** "Name of Chart." This becomes the worksheet tab name. In addition, there must be a row containing the LSL and/or USL. The row containing the nominal value is optional.

Name of Chart	Chart 1	Chart 2	Chart 3	Chart 4	Chart 5
LSL	70	68	75	60	65
Nominal	100	100	97.5	100	95
USL	130	132	120	140	125
	90	92	96	109	111
	92	112	101	102	99
	109	82	117	103	90
	89	91	91	116	117
	92	109	105	83	95
	97	107	112	76	105
	112	85	97	108	75
	95	99	101	115	94
	95	108	105	108	94
	102	99	85	112	101
	108	97	89	99	107
	101	93	111	83	106
	116	89	104	105	102
	110	118	102	98	99
	81	87	117	114	107
	94	101	97	116	95
	91	103	106	100	99
	108	97	106	108	87
	95	87	84	116	105
	100	96	113	114	112

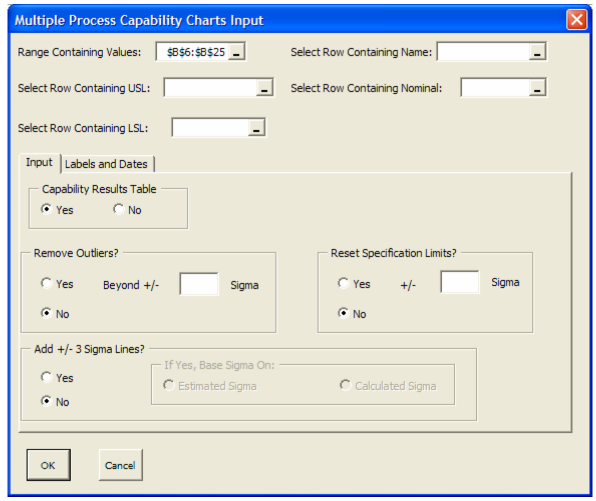
2. Select the shaded data above (the data for the first process capability analysis).
3. Select the Multiple Process Capability icon ( ) from the SPC menu.
4. Fill in the Multiple Process Capability Charts Input form. The top part of the form is for inputting the ranges containing the data, chart names, LSL, nominal, and USL. There are two other pages on the form: Input and Labels and Dates.
    a. *Range Containing Values:* This is the range of the data for the first process capability chart. The default value is the selected area on the spreadsheet. The data must be in columns for this feature.

b. *Select Row Containing Name:* Select a cell in the row or the row itself that contains the unique name of the chart. This name will be on the worksheet tab containing the process capability chart.

c. *Select Row Containing USL:* Select a cell in the row or the row that contains the USL values.

d. *Select Row Containing Nominal:* Select a cell in the row or the row that contains the nominal values.

e. *Select Row Containing LSL:* Select a cell in the row or the row that contains the LSL values.



f. Input Page

i. *Capability Results Table:* If "Yes" is selected, a table summarizing the process capability for all charts will be generated. An example is shown at the end of this section.

ii. *Remove Outliers?* Select "Yes" if you want to remove outliers from the calculations. Enter the number of standard deviations you want to remove outliers beyond (e.g., beyond +/- 6 sigma). The default option is "No."

iii. *Reset Specifications Limits?* Select "Yes" if you want the program to replace the existing specification limits with new limits set at the value of +/- sigma you enter. This is useful if you are trying to set specification limits, e.g., for prototype data.

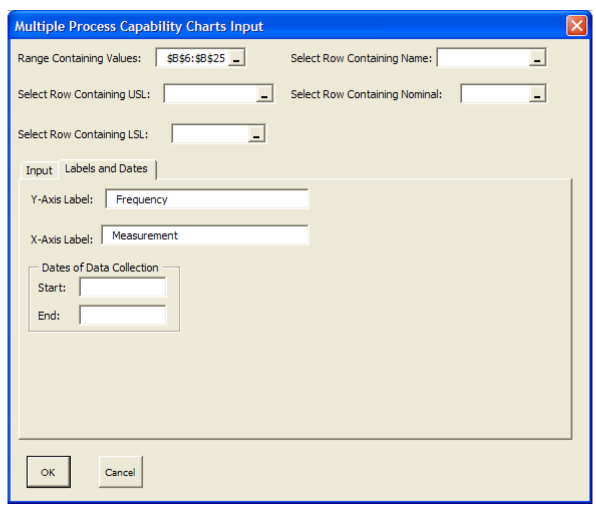
iv. *Add +/- 3 Sigma Limits:* In addition to the specifications, you can add the +/- three sigma limits to the chart. The default is No. If you select Yes, you can chose sigma to the estimated sigma from the range chart or the calculated standard deviation of all the data.

g. *Labels and Date Page*

i. *Y-Axis Label:* This is the vertical axis label. The default value is "Frequency."

ii. *X-Axis Label:* This is the horizontal axis label. The default value is "Measurement."

iii. *Dates of Data Collection:* Add the starting date and ending dates of data collection. These dates are optional. If entered, they will appear in a dialog box in the lower left-hand corner of the chart.



5. Select OK. The process capability charts will be generated. The program moves through each column until it finds a blank cell for the first sample. Each chart is placed on a new chart sheet. The process capability table is generated if that option was selected. The table for this data is shown below.

Name	Cp	Cpk	Cpu	Cpl	Est. Sigma	Pp	Ppk	Ppu	Ppl	Sigma	Average	Count	Minimum	Maximum	Kurtosis	Skewness	LSL	USL
Chart 1	1.07	1.03	1.11	1.03	9.33	1.09	1.05	1.13	1.05	9.17	98.85	20	81	116	-0.65	0.19	70	130
Chart 2	0.94	0.87	1.01	0.87	11.38	1.1	1.02	1.18	1.02	9.72	97.6	20	82	118	-0.53	0.37	68	132
Chart 3	0.67	0.54	0.54	0.8	11.24	0.77	0.62	0.62	0.92	9.73	101.95	20	84	117	-0.58	-0.28	75	120
Chart 4	1.29	1.16	1.16	1.43	10.31	1.13	1.01	1.01	1.25	11.78	104.25	20	76	116	0.74	-1.2	60	140
Chart 5	1.02	0.85	0.85	1.19	9.85	1.04	0.87	0.87	1.21	9.63	100	20	75	117	1.16	-0.68	65	125

Stacked and UnStacked Data Entry

Several of the techniques in the Six Sigma Tools section have the option of using “stacked” or “unstacked” data. This simply refers to how the data is entered into the spreadsheet.

Stacked Data Entry

The stacked data entry is shown below. There are five different processes or treatments. These are labeled A – E with the labels in the same row. The observations for each treatment are listed in the column. For example, for treatment A, the observations are 250, 260, 230, and 270.

A	B	C	D	E
250	310	250	340	250
260	330	230	270	240
230	280	220	300	270
270	360	260	320	290

Unstacked Data Entry

The unstacked data entry is shown below. The treatment levels are in one column and the responses are in another column.

Treatment Levels	Responses
A	270
E	290
D	320
A	250
A	260
C	260
B	360
D	270
B	330
E	270
C	230
A	230
E	240
D	340
C	250
B	310
B	280
D	300

Single/Two Samples

One Sample z and t Tests for a Mean

This test is performed on the mean of one sample. The test will gives you a confidence interval for the mean and has the option to test the hypothesis that the mean is equal to, less than, or greater than a specified mean. You have the following two options:

- t test: this test uses the t distribution and is usually used when with smaller sample sizes although it can be used with any sample size; this test uses the standard deviation calculated from the data.
- z test: this test uses the z distribution and is usually used with larger sample sizes (> 30); there two options for the standard deviation with the z distribution:
  - Option 1: enter the standard deviation; the user enters a known standard deviation and this value is used in the calculations
  - Option 2: calculate the standard deviation based on the data; the program will calculate the standard deviation based from the data and use this value in the calculations.

There are three options for the type of tests (where μ_0 is the specified mean):

- Two-sided:
  - Null hypotheses: $\mu = \mu_0$
  - Alternative hypotheses: $\mu \neq \mu_0$
- Lower one-sided:
  - Null hypotheses: $\mu = \mu_0$
  - Alternative hypotheses: $\mu < \mu_0$
- Upper one-sided:
  - Null hypotheses: $\mu = \mu_0$
  - Alternative hypotheses: $\mu > \mu_0$

The data can be in a range in a spreadsheet or the user can enter an average, standard deviation, and sample size for the test. The example below demonstrates how to do this test using data in a range in the spreadsheet.

Example[2]

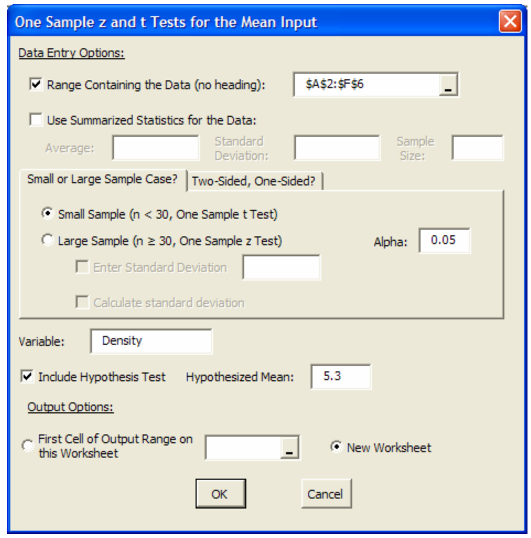
The data below represents estimates of the earth's density taken in 1798 by British scientist Henry Cavendish. There are 29 measurements and are expressed as multiples of the density of water. We want to use the t test to determine a 95% confidence interval around the mean density and to test the null hypothesis of $\mu = 5.3$ with the alternate hypothesis being $\mu \neq 5.3$.

1. Enter the data into a spreadsheet as shown below. The data can be one column or in multiple columns as shown below.

Density					
5.53	5.07	5.46	5.79	5.58	5.5
5.47	5.29	5.55	5.75	5.27	5.57
4.88	5.34	5.34	5.29	5.85	5.42
5.62	5.26	5.3	5.1	5.65	5.61
5.63	5.44	5.36	5.86	5.39	

[2] Example from Statistics and Data Analysis, by Ajit Tamhane and Dorothy Dunlop, Prentice-Hall, 2000.

2. Select the shaded data (not the column heading).
3. Select the Six Sigma Tools icon from the SPC Menu to show the form "Other Six Sigma Tools."
4. Select the Single/Two Sample page (this is the default page).
5. Select the One Sample t and z Tests for a Mean option.
6. Fill in the One Sample t and z Tests for a Mean Input form.



   a. *Data Entry Options:* There are two data entry options.
      i. *Range Containing the Data (no heading):* this is the worksheet range containing the data; default is the worksheet range highlighted on the worksheet.
      ii. *Use Summarized Statistis for the Data:* this option allows you to enter a calculated average, calculated standard deviation, and sample size.
   b. *Small or Large Sample Case?*: this is one of two tabs in the center of the form; this determines the case to use for the calculations; the default is based on the sample size; if the size is less than 30, the small sample case is selected; greater than or equal to 30, the large sample cause is selected. You may override this selection by selecting which test you want to use.
   c. *Alpha:* this is the confidence level; 1- alpha is the confidence interval.
   d. *Two-Sided, One-Sided:* this is the second of two tabs; select the tab and you will have the three options: two-sided, lower one-sided, or upper-one sided; two-sided is the default; select which type of test you want
   e. *Variable:* this is the name of the variable; the default is what is in the cell above the first cell in the selected range on the spreadsheet
   f. *Include Hypothesis Test:* this option will include a hypothesis test in addition to the confidence interval.
   g. *Hypothesized Mean:* this is the mean that you specify (in this example, 5.3).
   h. *Output Options:* There are two output options:
      i. *First Cell of Output Range on an Existing Worksheet:* select this option if you want the output on an existing worksheet; then select the first cell of the range where you want the output placed; a check is made to ensure that no existing data is overwritten in the worksheet.
      ii. *New Worksheet:* select this option to put the output on a new worksheet in the workbook.
7. Select OK when you have completed the input form.
8. The output for the density data is given below along with an explanation of the results.

Output

95% Two-Sided Hypothesis Test on a Mean		Explanation of Results (not part of output)
H_0: μ = 5.3		Null hypothesis
H_1: μ <> 5.3		Alternate hypothesis (two-sided in example)
Variable	Density	Name of variable
Mean	5.4541	Calculated average of data
Standard Deviation (s)	0.230	Calculated standard deviation
Sample Size	29	Number of data points
Standard Error of Mean	0.0427	Calculated as standard deviation/sqrt(count)
Degrees of Freedom	28	Calculated degrees of freedom = n -1
Alpha	0.05	Value of alpha entered by user
$t_{(0.025, 28)}$	2.048	t value based on alpha and degrees of freedom
Lower Confidence Limit	5.3666	Calculated lower confidence interval
Upper Confidence Limit	5.5416	Calculated upper confidence interval
Hypothesized Mean	5.3	Specified mean entered by user
t	3.608	Calculated t value
p Value	0.0012	Calculated p value (in red if p value ≤ alpha)
The null hypothesis is rejected.		Decision based on p value and alpha
There is evidence that the mean is not equal to 5.3.		
Reject if:		Reject if p value ≤ alpha
If p value (0.0012) <= alpha (0.05)		
If |t| > $t_{(0.025, 28)}$		

Notes:

- If the hypothesis test option is not selected, the output below the Upper Confidence Limit is not included.
- If the large sample size is selected, the z distribution will be used instead of the t distribution; the "t" in the output above will be replaced by "z" and the results will reflect the z distribution.
- If the large sample size is selected and a standard deviation entered, the output above will have "assumed standard deviation (σ)" instead of "standard deviation (s)."
- If the option for a one-sided test is selected, there is only one confidence limit given.

Formulas

The formulas for the small sample test (t) are given below:

Two-sided t confidence interval: $\bar{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$

Upper one-sided t confidence interval: $\mu \ge \bar{x} - t_{n-1,\alpha}\frac{s}{\sqrt{n}}$

Lower one-sided t confidence interval: $\mu \le \bar{x} + t_{n-1,\alpha}\frac{s}{\sqrt{n}}$

Calculated t value for hypothesis test: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

The formulas for the large sample test (z) are given below:

Two-sided z confidence interval: $\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Upper one-sided z confidence interval: $\mu \ge \bar{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}}$

Lower one-sided z confidence interval: $\mu \le \bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}$

Calculated z value for hypothesis test: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

One Sample Variance

This test is performed on the variance of one sample. The test will gives you a confidence interval for the standard deviation and the variance. It has the option to test the hypothesis that the standard deviation/variance is equal to, less than, or greater than a specified standard deviation/variance.

There are three options for the type of tests (where σ_0^2 is the specified variance):

- Two-sided:
  - Null hypotheses: $\sigma^2 = \sigma_0^2$
  - Alternative hypotheses: $\sigma^2 \neq \sigma_0^2$
- Lower one-sided:
  - Null hypotheses: $\sigma^2 = \sigma_0^2$
  - Alternative hypotheses: $\sigma^2 < \sigma_0^2$
- Upper one-sided:
  - Null hypotheses: $\sigma^2 = \sigma_0^2$
  - Alternative hypotheses: $\sigma^2 > \sigma_0^2$

The data can be in a range in a spreadsheet or the user can enter a standard deviation and sample size for the test. An example of how to do this test using SPC for MS Excel is given below. We will use the earth's density data from the One Sample z and t Tests section of these instructions.

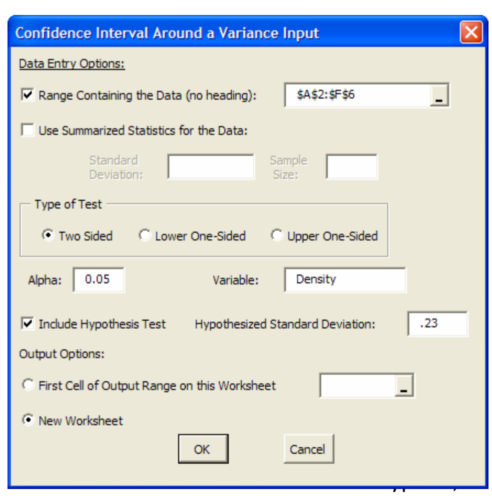
<u>Example</u>

In this example, we want to find a confidence interval around a variance as well as perform a hypothesis test to see if the standard deviation is equal to 0.23.

1. Enter the data into a spreadsheet as shown to the right and select the data (not the headings). The data can be in one column.

Density					
5.53	5.07	5.46	5.79	5.58	5.5
5.47	5.29	5.55	5.75	5.27	5.57
4.88	5.34	5.34	5.29	5.85	5.42
5.62	5.26	5.3	5.1	5.65	5.61
5.63	5.44	5.36	5.86	5.39	

2. Select the Six Sigma Tools icon from the SPC Menu to show the form "Other Six Sigma Tools."
3. Select the Single/Two Sample page (this is the default page).
4. Select the One Sample Variance Option
5. Fill in the One Sample Variance Input form.
   a. *Data Entry Options:* There are two data entry options.
      i. *Range Containing the Data (no heading):* this is the worksheet range containing the data; default is the worksheet range highlighted on the worksheet.
      ii. *Use Summarized Statistis for the Data:* this option allows you to enter a calculated average, calculated standard deviation, and sample size.



   b. *Type of Test:* select the type of test you want to perform: two-sided, lower one-sided, or upper-one sided; two-sided is the default.
   c. *Alpha:* this is the confidence level; 1- alpha is the confidence interval.
   d. *Variable:* this is the name of the variable; the default is what is in the cell above the first cell in the selected range on the spreadsheet
   e. *Include Hypothesis Test:* this option will include a hypothesis test in addition to the confidence interval.
   f. *Hypothesized Standard Deviation:* this is the standard deviation that you specify (in this example, .23).
   g. *Output Options:* There are two output options:
      i. *First Cell of Output Range on an Existing Worksheet:* select this option if you want the output on an existing worksheet; then select the first cell of the range where you want the output placed; a check is made to ensure that no existing data is overwritten in the worksheet.
      ii. *New Worksheet:* select this option to put the output on a new worksheet in the workbook.
6. Select OK when you have completed the input form.
7. The output for the density data is given below along with an explanation of the results.

Output

95% Two-Sided Hypothesis Test on a Variance		Explanation (not in output)
H_0: $\sigma^2 = 0.0529$		Null hypothesis
H_1: σ^2 <> 0.0529		Alternate hypothesis (two-sided)
Variable	Density	Name of variable
Standard Deviation	0.230	Calculated standard deviation
Variance	0.0529	Calculated variance
Sample Size	29	Number of data points
Degrees of Freedom	28	Degrees of freedom = n - 1
Alpha	0.05	Value of alpha entered by user
Variance LCL	0.0333	(1-alpha)% lower confidence interval for the variance
Variance UCL	0.0968	(1-alpha)% upper confidence interval for the variance
Standard Deviation LCL	0.183	(1-alpha)% lower confidence interval for the standard deviation
Standard Deviation UCL	0.311	(1-alpha)% upper confidence interval for the standard deviation
Hypothesized Variance	0.0529	Square of specified standard deviation entered by user
Hypothesized Standard Deviation	0.23	Specified standard deviation by user
$\chi^2_{(0.025,28)}$	15.308	Lower χ^2 value based on alpha and the degrees of freedom
$\chi^2_{(1-0.025,28)}$	44.461	Upper χ^2 value based on alpha and the degrees of freedom
χ^2	28.010	Calculated χ^2 value
p value	0.9279	Calculated p value (in red if p value ≤ alpha)
The null hypothesis is not rejected.		Decision based on alpha and p value
There is no evidence that the variance does not equal 0.0529.		Reject if p value ≤ alpha

Formulas

Two-sided confidence interval: $s\sqrt{\frac{n-1}{\chi^2_{n-1,\alpha/2}}} \le \sigma^2 \le s\sqrt{\frac{n-1}{\chi^2_{n-1,1-\alpha/2}}}$

One-sided upper confidence interval: $\sigma^2 \le s\sqrt{\frac{n-1}{\chi^2_{n-1,1-\alpha}}}$

One-sided lower confidence interval: $\sigma^2 \ge s\sqrt{\frac{n-1}{\chi^2_{n-1,\alpha}}}$

Calculated chi-square statistic for the hypothesis test: $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$

Two Sample z and t Tests for Difference in Two Means

This test compares the means of two samples. The test will provide a confidence interval for the difference in two means and has the option for hypothesis testing as well. The test accounts automatically for variances that are equal or unequal.

. You have the following two options:

- t test: this test uses the t distribution and is usually used when with smaller sample sizes although it can be used with any sample size; this test uses the standard deviation calculated from the data.
- z test: this test uses the z distribution and is usually used with larger sample sizes (> 30); there two options for the standard deviation with the z distribution:
    - Option 1: enter the standard deviation; the user enters a known standard deviation for both methods and these values are used in the calculations
    - Option 2: calculate the standard deviation based on the data; the program will calculate the standard deviation for both methods based from the data and use these values in the calculations.

There are three options for the type of tests (where δ_0 is the specified difference in the two means):

- Two-sided:
    - Null hypotheses: $\mu_1 - \mu_2 = \delta_0$
    - Alternative hypotheses: $\mu_1 - \mu_2 \neq \delta_0$
- Lower one-sided:
    - Null hypotheses: $\mu_1 - \mu_2 = \delta_0$
    - Alternative hypotheses: $\mu_1 - \mu_2 < \delta_0$
- Upper one-sided:
    - Null hypotheses: $\mu_1 - \mu_2 = \delta_0$
    - Alternative hypotheses: $\mu_1 - \mu_2 > \delta_0$

The data can be in a range in a spreadsheet or the user can enter an average, standard deviation, and sample size for each method. The example below demonstrates how to do this test using the SPC for MS Excel using data in a range in the spreadsheet.

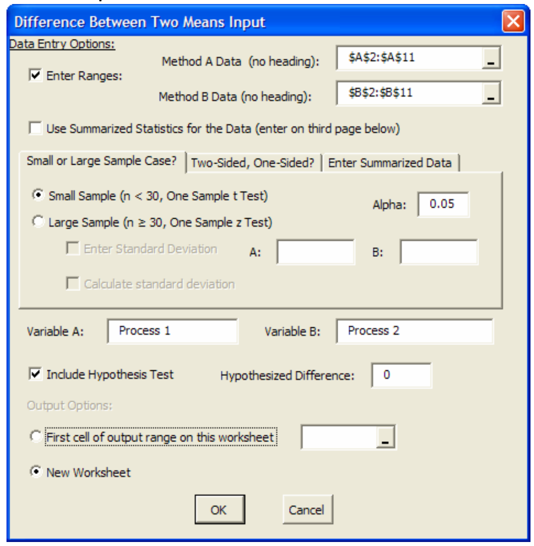
Example

A company uses two different processes to put sand into bags. The question is whether each process places the same weight into the bags. Ten bags from each process were weighed. The data is shown below. We want to see if there is a difference in the means from the two processes.

1. Enter the data into a spreadsheet as shown below.

Process 1	Process 2
50.1	50.1
49.9	49.7
50.0	50.0
49.9	49.4
50.2	49.4
50	49.4
50.4	49.7
49.9	49.4
49.7	49.4
49.7	49.3

9. Select the shaded data (not the column heading).
10. Select the Six Sigma Tools icon from the SPC Menu to show the form “Other Six Sigma Tools.”
11. Select the Single/Two Sample page (this is the default page).
12. Select the Two Sample z and t Tests for Difference in Two Means option.
13. Fill in the Two Sample z and t Tests for Difference in Two Means Input form.



    a. *Data Entry Options:* There are two data entry options.
        i. *Enter Ranges:* this option allows you to enter the ranges on the worksheet containing the data for Methods A and B; default is the worksheet range highlighted on the worksheet.
        ii. *Use Summarized Statistis for the Data:* this option allows you to enter a calculated average, calculated standard deviation, and sample size for each method.
    b. *Small or Large Sample Case?*: this is one of three tabs in the center of the form; this determines the case to use for the calculations; the default is based on the sample size; if the size is less than 30, the small sample case is selected; greater than or equal to 30, the large sample cause is selected. You may override this selection by selecting which test you want to use.
    c. *Alpha:* this is the confidence level; 1- alpha is the confidence interval.
    d. *Two-Sided, One-Sided:* this is the second of three tabs; select the tab and you will have the three options: two-sided, lower one-sided, or upper-one sided; two-sided is the default; select which type of test you want
    e. *Enter Summarized Data:* this is the third tab; select this tab if you want to enter the average, standard deviation, and sample size for each method.
    f. *Variable A:* this is the name of Method A; the default is what is in the cell above the first cell in the selected range on the spreadsheet
    g. *Variable B:* this is the name of Method A; the default is what is in the cell above the first cell in the second selected range on the spreadsheet
    h. *Include Hypothesis Test:* this option will include a hypothesis test in addition to the confidence interval.
    i. *Hypothesized Difference:* this is the difference in means that you specify (in this example, 0).
    j. *Output Options:* There are two output options:
        i. *First Cell of Output Range on an Existing Worksheet:* select this option if you want the output on an existing worksheet; then select the first cell of the range where you want the output placed; a check is made to ensure that no existing data is overwritten in the worksheet.
        ii. *New Worksheet:* select this option to put the output on a new worksheet in the workbook.
14. Select OK when you have completed the input form.
15. The output for the test is given below along with an explanation of the results.

Output

95% Two-Sided Hypothesis Test for the Difference in Two Means			Explanation (not in output)
H_0: μ_1 - μ_2 = 0			Null hypothesis
H_0: μ_1 - μ_2 <> 0			Alternate hypothesis (two-sided in this example)
	Process 1	*Process 2*	
Mean	49.980	49.580	Calculated means of two methods
Standard Deviation	0.215	0.282	Calculated standard deviation of two methods
Variance	0.0462	0.0796	Calculated variance of two methods
Sample Size	10	10	Sample size for each method
Difference in Means	0.400		Method A Mean – Method B Mean
Equal Variances?	Yes		Yes is variances are equal; no if not
Pooled Variance	0.0629		Pooled variance based on equal or unequal variances
Pooled Standard Deviation	0.251		Pooled standard deviation based on equal or unequal variances
Degrees of Freedom	18		Degrees of freedom (depend on equal or unequal variance)
Alpha	0.05		Alpha entered by user
$t_{(0.025, 18)}$	2.101		t value based on alpha and degrees of freedom
Lower Confidence Level	0.164		100(1-alpha)% lower confidence interval
Upper Confidence Level	0.636		100(1-alpha)% upper confidence interval
t	3.567		Calculated t value
p Value	0.0022		Calculated p value (in red if p value ≤ alpha)
The null hypothesis is rejected.			Decision based on alpha and p value
There is evidence that the difference in means is not equal to 0.			Reject if p value ≤ alpha
Reject if:			
If p value (0.0022) <= alpha (0.05)			
If $\|t\| > t_{(0.025, 18)}$			

Notes:

- The difference in means is the mean of Method A – mean of Method B
- The test uses the F test to determine if the variances are equal or not equal. This changes the way the pooled variance and degrees of freedom are calculated (see formulas below).

Formulas

The formulas used in the calculations for the small sample case (t) are given below.

Two-sided t confidence interval: $\bar{x}_1 - \bar{x}_2 - t_{n_1+n_2-2,\alpha/2} s_{diff} \le \mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + t_{n_1+n_2-2,\alpha/2} s_{diff}$

Upper one-sided t confidence interval: $\mu_1 - \mu_2 \ge \bar{x}_1 - \bar{x}_2 - t_{n_1+n_2-2,\alpha/2} s_{diff}$

Lower one-sided t confidence interval: $\mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + t_{n_1+n_2-2,\alpha/2} s_{diff}$

If the variances are equal:

$$s_{diff} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

Degrees of freedom = $n_1 + n_2 - 2$

If the variances are not equal:

$$s_{diff} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Degrees of freedom = $\dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\left(\dfrac{s_1^2/n_1}{n_1 - 1}\right) + \left(\dfrac{s_2^2/n_2}{n2 - 1}\right)}$

Calculated t value for hypothesis test: $t = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{s_{diff}}$

The formulas used in the large sample case (z) are given below.

Two-sided z confidence interval: $\bar{x}_1 - \bar{x}_2 - z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \le \mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Upper one-sided z confidence interval: $\mu_1 - \mu_2 \ge \bar{x}_1 - \bar{x}_2 - z_\alpha\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Lower one-sided z confidence interval: $\mu_1 - \mu_2 \le \bar{x}_1 - \bar{x}_2 + z_\alpha\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Calculated z value for hypothesis test: $z = \dfrac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

Paired Sample Comparison t Test

The paired sample comparison t test is used when the samples are not independent. For example, you might want to compare two analytical test methods. You take a series of samples; mix them well; divide them in half; and run one-half in one analytical test method and the other half in the second analytical test. The samples are not independent in this case. You can use this test to determine if there is a significant difference in the two tests.

There are three options for the type of tests (where δ_0 is the specified difference):

- Two-sided:
    - Null hypotheses: $\mu_1 - \mu_2 = \delta_0$
    - Alternative hypotheses: $\mu_1 - \mu_2 \neq \delta_0$
- Lower one-sided:
    - Null hypotheses: $\mu_1 - \mu_2 = \delta_0$
    - Alternative hypotheses: $\mu_1 - \mu_2 < \delta_0$
- Upper one-sided:
    - Null hypotheses: $\mu_1 - \mu_2 = \delta_0$
    - Alternative hypotheses: $\mu_1 - \mu_2 > \delta_0$

The data can be in a range in a spreadsheet or the user can enter an average, standard deviation, and sample size for each method. The example below demonstrates how to do this test using the SPC for MS Excel using data in a range in the spreadsheet.

Example[3]

A medical researcher wants to compare two methods of measuring cardiac output. Method A is the standard method and is considered very accurate. But this method is invasive. Method B is less accurate but not invasive. Cardiac output from 26 patients was measured using both methods. We want to sue the paired sample t test to determine if there is a significant difference between the two methods.

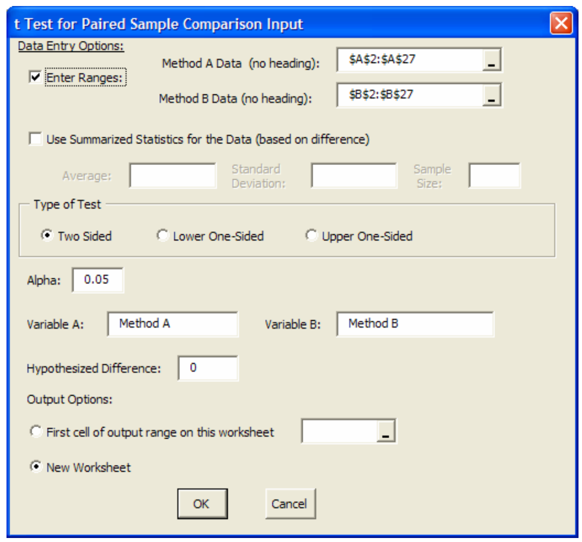
1. Enter the data for the two methods into a spreadsheet as shown below. The data does not have to be in adjacent columns.

Method A	Method B
6.3	5.2
6.3	6.6
3.5	2.3
5.1	4.4
5.5	4.1
7.7	6.4
6.3	5.7
2.8	2.3
3.4	3.2
5.7	5.2
5.6	4.9
6.2	6.1
6.6	6.3
7.7	7.4
7.4	7.4
5.6	4.9
6.3	5.4
8.4	8.4

[3] Example from Statistics and Data Analysis, by Ajit Tamhane and Dorothy Dunlop, Prentice-Hall, 2000.

5.6	5.1
4.8	4.4
4.3	4.3
4.2	4.1
3.3	2.2
3.8	4
5.7	5.8
4.1	4

2. Highlight the data (not the headings)
3. Select the Six Sigma Tools icon from the SPC Menu to show the form "Other Six Sigma Tools."
4. Select the Single/Two Sample page (this is the default page).
5. Select the t Test for Paired Sample Comparison option.
6. Fill in the t Test for Paired Sample Comparison Input form.



   a. *Data Entry Options:* There are two data entry options.
      i. *Enter Ranges:* this option allows you to enter the ranges on the worksheet containing the data for Methods A and B; default is the worksheet range highlighted on the worksheet.
      ii. *Use Summarized Statistis for the Data:* this option allows you to enter a calculated average, calculated standard deviation, and sample size for the differences.
   b. *Two-Sided, One-Sided:* you will have the three options: two-sided, lower one-sided, or upper-one sided; two-sided is the default; select which type of test you want
   c. *Alpha:* this is the confidence level; 1- alpha is the confidence interval.
   d. *Variable A:* this is the name of Method A; the default is what is in the cell above the first cell in the selected range on the spreadsheet
   e. *Variable B:* this is the name of Method A; the default is what is in the cell above the first cell in the second selected range on the spreadsheet
   f. *Include Hypothesis Test:* this option will include a hypothesis test in addition to the confidence interval.
   g. *Hypothesized Difference:* this is the difference in means that you specify (in this example, 0).
   h. *Output Options:* There are two output options:
      i. *First Cell of Output Range on an Existing Worksheet:* select this option if you want the output on an existing worksheet; then select the first cell of the range where you want the output placed; a check is made to ensure that no existing data is overwritten in the worksheet.
      ii. *New Worksheet:* select this option to put the output on a new worksheet in the workbook.
7. Select OK when you have completed the input form.
8. The output for the test is given below along with an explanation of the results.

Output

95% Two-Sided Hypothesis Test for Paired Sample Comparisons				Explanation (not in output)
H_0: $\delta_0 = 0$				Null hypothesis
H_1: δ_0 <> 0				Alternate hypothesis (two-sided in this example)
	Difference	Method A	Method B	
Mean	0.465	5.469	5.004	Calculated means, difference is Method A – Method B
Standard Deviation	0.482	1.486	1.580	Calculated standard deviations
Variance	0.233	2.209	2.496	Calculated observations
Sample Size	26			Number of paired samples
Degrees of Freedom	25			Degrees of freedom
Alpha	0.05			Alpha entered by user
$t_{(0.025, 25)}$	2.060			t value based on alpha and degrees of freedom
Upper Confidence Level	0.271			100(1-alpha)% upper confidence limit
Lower Confidence Level	0.660			100(1-alpha)% lower confidence limit
t	4.919			Calculated t value
p value	0.0000			Calculated p value (in red if p value ≤ alpha)
The null hypothesis is rejected.				Decision based on p value and alpha
There is evidence that the mean difference is not equal to 0.				Reject if p value ≤ alpha
Reject if:				
If p value (0) <= alpha (0.05)				
If |p value| > $t_{(0.025, 25)}$				

Formulas:

Let $\overline{d}$ = average of the sample difference between Method A and Method B and let s_d be the standard deviation of those differences.

Two-sided t confidence interval: $\overline{d} - t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}} \le \delta_0 \le \overline{d} + t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}}$

Upper one-sided t confidence interval: $\delta_0 \ge \overline{d} - t_{n-1,\alpha}\frac{s_d}{\sqrt{n}}$

Lower one-sided t confidence interval: $\delta_0 \le \overline{d} + t_{n-1,\alpha}\frac{s_d}{\sqrt{n}}$

Calculated t value for hypothesis test: $t = \frac{\overline{d} - \delta_0}{s_d / \sqrt{n}}$

One Proportion Test

This test is performed with binary data when you want to examine the absence or presence of a specific attribute. This type of test is very common in elections when polling people to find out if they support person A or person B.

The test gives two options based on the size of the sample: large or small sample. A different approach is used based on which case is selected. You should select the large sample case when $n\bar{p} \geq 10$ with n = sample size and $\bar{p}$ is the average proportion of events.

There are three options for the type of tests (where p_0 is the specified proportion):

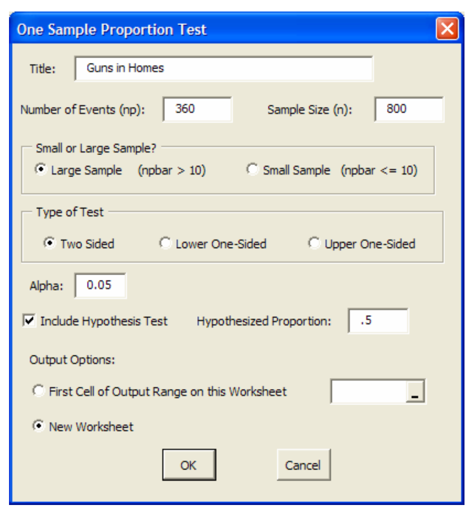
- Two-sided:
  - Null hypotheses: $p = p_0$
  - Alternative hypotheses: $p \neq p_0$
- Lower one-sided:
  - Null hypotheses: $p = p_0$
  - Alternative hypotheses: $p < p_0$
- Upper one-sided:
  - Null hypotheses: $p = p_0$
  - Alternative hypotheses: $p > p_0$

The example below demonstrates how to do this test using the SPC for MS Excel.

<u>Example</u>[4]

A magazine conducted a telephone survey of 800 adults and asked if they had guns in the home. 45% of the respondents said yes. We want to develop a 95% confidence interval for the proportion of homes with guns. We also want a hypothesis test to see if 50% of the homes have guns. In this example, n = 800 and $\bar{p}$ = .45. Thus, $n\bar{p}$ =800(.45) = 360. We will use the large sample option.

1. Select the Six Sigma Tools icon from the SPC Menu to show the form "Other Six Sigma Tools."
2. Select the Single/Two Sample page (this is the default page).
3. Select the One Proportion Test option.
4. Fill in the One Proportion Test Input form.
    a. *Title:* enter the title of the test (optional)
    b. *Number of Events (np):* this is the number of events that occurred (e.g., the number of adults that said they had a gun in the home).
    c. *Sample size (n):* this is the sample size (e.g., the total number of adults who responded).
    d. *Small or Large Sample?*: select the large sample size if the number of events is greater than 10; select the small sample size if it is less than 10.
    e. *Type of Test:* option for two-sided, lower one-sided, or upper one-sided test; two-sided test is the default.
    f. *Alpha:* the confidence level; 1- alpha is the confidence interval.
    g. *Hypothesized Proportion:* this is the proportion you specify (in this example, 0.5).
    h. *Output Options:* There are two output options:



[4] Example from <u>Statistics and Data Analysis</u>, by Ajit Tamhane and Dorothy Dunlop, Prentice-Hall, 2000.

i. *First Cell of Output Range on an Existing Worksheet:* select this option if you want the output on an existing worksheet; then select the first cell of the range where you want the output placed; a check is made to ensure that no existing data is overwritten in the worksheet.
ii. *New Worksheet:* select this option to put the output on a new worksheet in the workbook.

5. Select OK when you have completed the input form.
6. The output for the test is given below along with an explanation of the results.

Output

95% Two-Sided Hypothesis Test for a Sample Proportion		Explanation (not in output)
Houses with Guns		
H_0: p = 0.5		Null hypothesis
H_1: p <> 0.5		Alternate hypothesis (two-sided in this example)
Sample Case	Large	Sample case selected by user
Number of Events	360	Number of events entered by user
Sample Size	800	Sample size entered by user
Sample p	0.45	Sample p = events/sample size
Sample q	0.55	Sample q = 1 - p
Alpha	0.05	Alpha entered by user
$z_{(0.025)}$	1.960	Calculated z value based on alpha
Lower Confidence Limit	0.416	100(1-alpha)% upper confidence interval
Upper Confidence Limit	0.485	100(1-alpha)% lower confidence interval
z	-2.828	Calculated z value
p Value	0.0047	Calculated p value (in red if p value ≤ alpha)
The null hypothesis is rejected.		Decision based on p value and alpha
There is evidence that the proportion is not equal to 0.5.		Reject if p value ≤ alpha
Reject if:		
If p value (0.0047) <= alpha (0.05)		
If |z| > $z_{(0.025)}$		

Formulas

The following formulas are used for the large sample case using p = proportion of events and q = proportion of non-events:

Two-side confidence interval: $$\frac{p+\frac{z_{\alpha/2}^2}{2n}-\sqrt{\frac{pqz_{\alpha/2}^2}{n}+\frac{z_{\alpha/2}^4}{4n^2}}}{1+\frac{z_{\alpha/2}^2}{n}} \le p \le \frac{p+\frac{z_{\alpha/2}^2}{2n}+\sqrt{\frac{pqz_{\alpha/2}^2}{n}+\frac{z_{\alpha/2}^4}{4n^2}}}{1+\frac{z_{\alpha/2}^2}{n}}$$

One-sided upper confidence: $$p \le \frac{p+\frac{z_{\alpha}^2}{2n}+\sqrt{\frac{pqz_{\alpha}^2}{n}+\frac{z_{\alpha}^4}{4n^2}}}{1+\frac{z_{\alpha}^2}{n}}$$

One-sided lower confidence interval: $p \geq \dfrac{p + \dfrac{z_\alpha^2}{2n} - \sqrt{\dfrac{pqz_\alpha^2}{n} + \dfrac{z_\alpha^4}{4n^2}}}{1 + \dfrac{z_\alpha^2}{n}}$

Calculated z value for hypothesis test: $z = \dfrac{p - p_0}{\sqrt{p_0 q_0 / n}}$

The small sample case uses the F distribution to develop the confidence intervals. The probability is based on the binomial distribution.

Define the following using np = number of events and n = sample size:

ν_1 = 2np, ν_2 = 2(n – np +1), ν_3 = 2(np+1), and ν_4 = 2(n-np)

Two-side confidence interval for small case:

Lower bound = $\dfrac{\nu_1 u}{\nu_2 + \nu_1 u}$ *where* $u = F_{(\nu_1, \nu_2, 0.025)}$

Upper bound = $\dfrac{\nu_3 w}{\nu_4 + \nu_3 w}$ *where* $w = F_{(\nu_1, \nu_2, 0.975)}$

Upper one-sided confidence interval

Upper bound = $\dfrac{\nu_3 w}{\nu_4 + \nu_3 w}$ *where* $w = F_{(\nu_1, \nu_2, 0.95)}$

Lower one-sided confidence interval

Lower bound = $\dfrac{\nu_1 u}{\nu_2 + \nu_1 u}$ *where* $u = F_{(\nu_1, \nu_2, 0.05)}$

Two Proportions Test

This test is used to compare the proportions or rates of two populations with binary outcomes. For example, you could be testing two different drugs to determine if one is better than another at curing a certain illness. The Two Proportions Test can be used to determine this.

There are three options for the type of tests (where δ_0 is the specified difference in proportions):

- Two-sided:
  - Null hypotheses: $p_1 - p_2 = \delta_0$
  - Alternative hypotheses: $p_1 - p_2 \neq \delta_0$
- Lower one-sided:
  - Null hypotheses: $p_1 - p_2 = \delta_0$
  - Alternative hypotheses: $p_1 - p_2 < \delta_0$
- Upper one-sided:
  - Null hypotheses: $p_1 - p_2 = \delta_0$
  - Alternative hypotheses: $p_1 - p_2 > \delta_0$

The example below demonstrates how to do this test using the SPC for MS Excel.

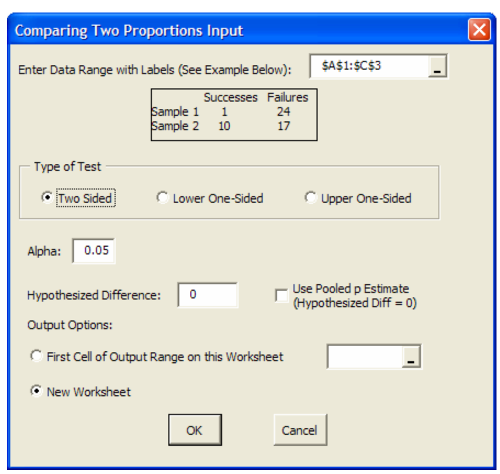
Example[5]

A test was performed to compare two drug therapies on leukemia. The results are shown below. We want to use these results to conduct a two parametric test.

1. Enter the data in a spreadsheet as shown below. Note that the number of events (successes in this case) and the number non-events (failures in this case) are entered on the spreadsheet.

	Success	Failure
Drug 0	14	7
Drug 1	38	4

2. Select the data and the row and columns headings.
3. Select the Six Sigma Tools icon from the SPC Menu to show the form “Other Six Sigma Tools.”
4. Select the Single/Two Sample page (this is the default page).
5. Select the Two Proportions Test option.
6. Fill in the Two Proportions Test Input form.
   a. *Enter Data Range with Labels:* this is the spreadsheet range containing the data; the default is the selected range on the spreadsheet
   b. *Type of Test:* option for two-sided, lower one-sided, or upper one-sided test; two-sided test is the default.
   c. *Alpha:* the confidence level; 1- alpha is the confidence interval.
   d. *Hypothesized Difference:* this is the difference in proportions you specify (in this example, 0).
   e. *Use Pooled p Estimate:* this options pools the successful proportions and uses this value in the z calculation; can only be used if the hypothesized difference is 0.



[5] Example from Statistics and Data Analysis, by Ajit Tamhane and Dorothy Dunlop, Prentice-Hall, 2000.

f. *Output Options:* There are two output options:

   i. *First Cell of Output Range on an Existing Worksheet:* select this option if you want the output on an existing worksheet; then select the first cell of the range where you want the output placed; a check is made to ensure that no existing data is overwritten in the worksheet.

   ii. *New Worksheet:* select this option to put the output on a new worksheet in the workbook.

7. Select OK when you have completed the input form.
8. The output for the test is given below along with an explanation of the results

Output

95% Two-Sided Hypothesis Test for the Difference in Two Proportions			Explanation (not in output)
H_0: p_1 - p_2 = 0			Null hypothesis
H_1: p_1 - p_2 <> 0			Alternate hypothesis (two-sided in this example)
	Drug 0	Drug 1	
Number of Events	14	38	Number of events from data
Sample Size	21	42	Calculated sample size from data
Sample p	0.667	0.905	Calculated sample proportion of events
Sample q	0.333	0.095	Calculated sample proportion of non-events
Use Pooled p Estimate?	Yes		Yes, if pooled estimate is used
Pooled p Estimate	0.825		Calculated pooled estimate of proportion of events
Difference in Proportions	-0.238		Calculated difference in proportion of events
Alpha	0.05		Alpha entered by user
$z_{(0.025)}$	1.960		z value based on alpha
Lower Confidence Limit	-0.458		100(1-alpha)% lower confidence limit
Upper Confidence Limit	-0.018		100(1-alpha)% upper confidence limit
z	-2.347		Calculated z value
p Value	0.0189		Calculated p value (in red if p value ≤ alpha)
Fisher's Exact Test p	0.0323		Will be included if hypothesized difference is 0
The null hypothesis is rejected.			Decision based on p value and alpha
There is evidence that the difference in proportions is not equal to 0.			Reject if p value ≤ alpha
Reject if:			
If p value (0.0189) <= alpha (0.05)			
If |z| > $z_{(0.025)}$			

Formulas

$\hat{p}$ = proportion of events in the sample

$\hat{q}$ = proportion of non-events in the sample = 1 - $\hat{p}$

n = sample size

Two sided conference interval: $\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \le p_1 - p_2 \le \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

Upper one-sided confidence interval: $p_1 - p_2 \ge \hat{p}_1 - \hat{p}_2 - z_\alpha\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

Lower one-sided confidence interval: $p_1 - p_2 \le \hat{p}_1 - \hat{p}_2 + z_\alpha \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

Calculated z value when not using pooled estimate of p: $z = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$

Pooled estimate of p: $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$

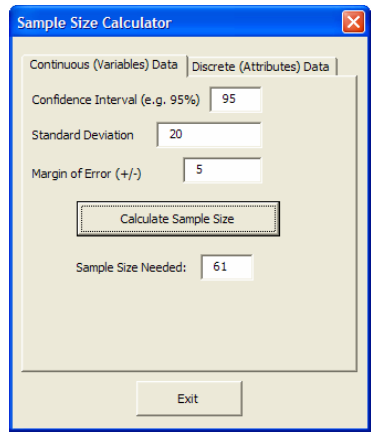
Calculated z value when using pooled estimate of p: $z = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{p}\hat{q}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

Sample Size Calculator

To access the sample size calculator:

1. Select the Six Sigma Tools icon from the SPC Menu.
2. Select the Sample Size Calculator option on the form Other Six Sigma Tools

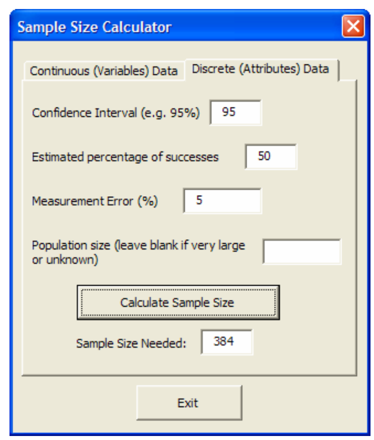
When you select this option, you will see the dialog box to the right. The first page is for variables data. The second page is for attributes data.



For variables data, select the Continuous (Variables) Data tab and enter the following:

- Confidence interval
- Standard deviation
- Margin of error (+/-)

In the example, we have a 95% confidence interval. The process has a standard deviation of 20. The margin of error is how large a difference in means that we would like to be able to detect. In this example, it is 5. Then select "Calculate Sample Size," and the sample size needed will appear in the bottom box. In this example, 61 samples are needed. This test is based on the assumption of the normal distribution.



For attributes data, select the Discrete (Attributes) Data tab and enter the following:

- Confidence interval
- Estimated percentage of successes
- Measurement error (%)
- Population size (if known)

In the example, we want a 95% confidence level. 50% is our estimate of successes. We want results that are accurate within +/- 5%. Then select "Calculate Sample Size," and the sample size needed will appear in the bottom box. In this example, 384 is the sample size we need.

Multiple Means and Variances

Fisher's LSD Method for Means

The Fisher Least Significant Difference Method is used to compare means from multiple processes. The method compares all pairs of means. It controls the error rate (α) for each individual pairwise comparison but does not control the family error rate. Both error rates are given in the output. The output from this method also includes the ANOVA table, the table of mean comparisons, and a graph containing all pairs of means.

There are two options for entering the data: stacked or unstacked. Stacked data has each treatment in a single column. Unstacked data has the treatment labels in one column and the results in the adjacent column. Please refer to the Stacked and Unstacked Data Entry section for more information. The example below uses stacked data. The example below shows how to use this method in SPC for MS Excel.

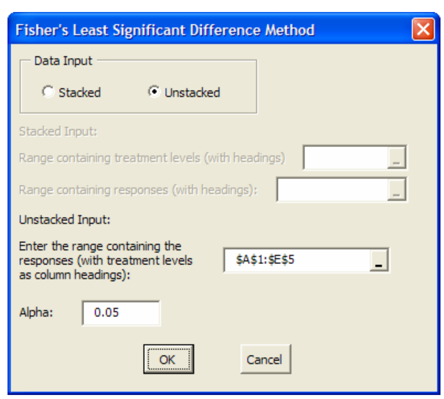
<u>Example</u>[6]

Five different treatments (A to E) were used to control the weight of a coating in grams. There were four samples for each treatment. We want to use Fisher's LSD method to determine if there are any significant differences in the treatments. The data is shown below.

1. Enter the data into a spreadsheet as shown below.
    a. The column headings represent the treatment labels with the responses for each treatment under the label. The number of results for each treatment does not have to be equal.

A	B	C	D	E
250	310	250	340	250
260	330	230	270	240
230	280	220	300	270
270	360	260	320	290

2. Select the data and the column headings.
3. Select the following from the SPC Menu:
    Six Sigma Tools → Other Six Sigma Tools → Multiple Means and Variances → Fisher's LSD Method for Means
4. Fill in the Fisher's Least Significant Difference Method Input form.
    a. *Data Input:* there are two options: stacked and unstacked; if the range selected on the worksheet contains more than two columns, the default is unstacked; this can be changed; the layout of the data entry is shown above.
    b. *Alpha:* the confidence level; 1- alpha is the confidence interval.
5. Select OK when you have completed the input form.
6. The output for the method is described below.



[6] Example data from Wheeler, Donald, <u>Advanced Topics in Statistical Process Control</u>, SPC Press, Inc, 1995.

Output

Two additional worksheets are added to the workbook for this analysis. One contains the data, the ANOVA table, and the means pairs table. The other is graph depicting all pairs of means.

The data is summarized as stacked data along with some statistics for each treatment level as shown below. The statistics include the number of results (count), the average, the standard deviation, and the variance for each treatment level. The value of alpha is also given.

	A	B	C	D	E
	270	360	260	320	290
	250	330	230	270	270
	260	310	250	340	240
	230	280	220	300	250
Count	4	4	4	4	4
Average	252.5	320	240	307.5	262.5
Std. Dev.	17.078	33.665	18.257	29.861	22.174
Variance	291.667	1133.333	333.333	891.667	491.667
Alpha	0.05				

The ANOVA table is given next as shown below. This table provides the sum of squares of the treatments, error and total as well as the degrees of freedom. The mean square results are given along with the F value. The key number is the p value. If this is less than or equal to alpha, it is turned red.

ANOVA					
Source	Sum of Squares	Degrees of Freedom	Mean Square	F	p Value
Treatments	19830	4	4957.5	7.89	0.0012
Error	9425	15	628.3333		
Total	29255	19			

The table of means is then given. This table is shown below. The family error rate and the individual rate are given. These are expressed as confidence limits.

Fisher Least Significant Difference (LSD) Method					
Family Conf. Int.=74.23%, Individual Conf. Int.=95%					
Comparisons	Diff. in Means	LSD	LCon	UCon	Sig Diff.?
A - B	-67.5	37.779	-105.279	-29.721	Yes
A - C	12.5	37.779	-25.279	50.279	No
A - D	-55	37.779	-92.779	-17.221	Yes
A - E	-10	37.779	-47.779	27.779	No
B - C	80	37.779	42.221	117.779	Yes
B - D	12.5	37.779	-25.279	50.279	No
B - E	57.5	37.779	19.721	95.279	Yes
C - D	-67.5	37.779	-105.279	-29.721	Yes
C - E	-22.5	37.779	-60.279	15.279	No
D - E	45	37.779	7.221	82.779	Yes

There is evidence that some pairs of means are different.

The columns in the table are:

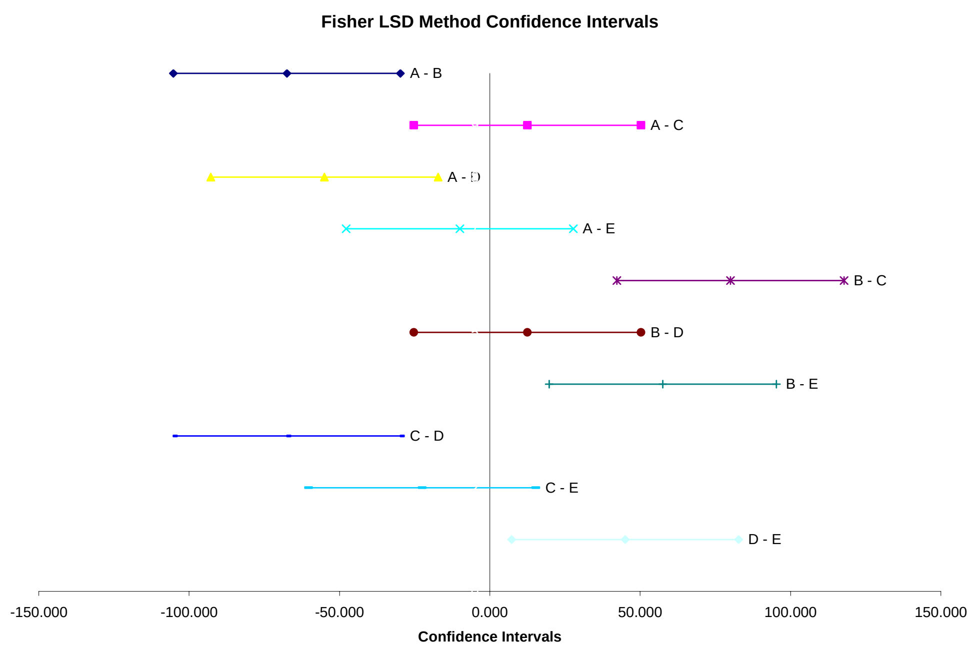
- Comparisons: all possible combination of pairs of means
- Diff in Means: the difference in the two means

- LSD: Fisher's Least Significant Difference which is given by:[7]

$$LSD = t_{\alpha/2,N-a}\sqrt{MS_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

  where N = number of total results, a = number of treatment levels, MS_E is the mean square error and n is the sample size for the individual treatment levels.
- LCON: the 100(1-α)% lower confidence interval = difference in means – LSD
- UCON: the 100(1-α)% upper confidence interval = difference in means + LSD
- Sig Diff?: conclusion if there is a significant difference between the two means; if there is a significant difference, Yes is printed and turned to red; otherwise, No is printed. There is a significant difference if the absolute value of the difference in means is greater than the LSD.

The second worksheet added to the workbook contains the confidence interval charts for the pairs of means as shown below. Each confidence interval is labeled. Those that do not include 0 are significantly different.



[7] Montgomery, Douglas C., Design and Analysis of Experiments, Sixth Edition, John Wiley & Sons, Inc, 2005

Tukey's Method for Means

The Tukey's Method is used to compare means from multiple processes. The method compares all pairs of means. It controls the family error rate (α). Both the family error and the individual error rates are given in the output. The output from this method also includes the ANOVA table, the table of mean comparisons, and a graph containing all pairs of means.

There are two options for entering the data: stacked or unstacked. Stacked data has each treatment in a single column. Unstacked data has the treatment labels in one column and the results in the adjacent column. Please refer to the Stacked and Unstacked Data Entry section for more information. The example below uses stacked data. The example below shows how to use this method in SPC for MS Excel. This is the same data that was used for the Fisher LSD method.

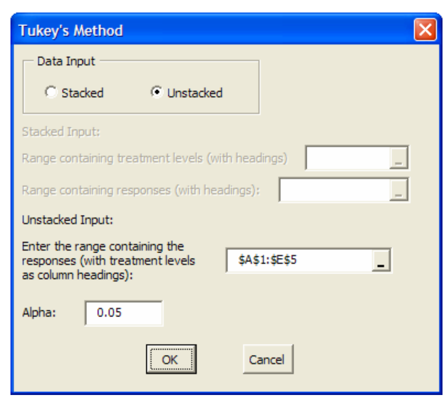
Example[8]

Five different treatments (A to E) were used to control the weight of a coating in grams. There were four samples for each treatment. We want to use Tukey's method to determine if there are any significant differences in the treatments. The data is shown below.

1. Enter the data into a spreadsheet as shown below.
    a. .The column headings represent the treatment labels with the responses for each treatment under the label. The number of results for each treatment does not have to be equal.

A	B	C	D	E
250	310	250	340	250
260	330	230	270	240
230	280	220	300	270
270	360	260	320	290

2. Select the data and the column headings.
3. Select the following from the SPC Menu:
    Six Sigma Tools → Other Six Sigma Tools → Multiple Means and Variances → Tukey's Method for Means
4. Fill in the Tukey's Method Input form.
    a. *Data Input:* there are two options: stacked and unstacked; if the range selected on the worksheet contains more than two columns, the default is unstacked; this can be changed; the layout of the data entry is shown above.
    b. *Alpha:* the confidence level; 1- alpha is the confidence interval.
5. Select OK when you have completed the input form.
6. The output for the method is described below.



[8] Example data from Wheeler, Donald, Advanced Topics in Statistical Process Control, SPC Press, Inc, 1995.

Output

Two additional worksheets are added to the workbook for this analysis. One contains the data, the ANOVA table, and the means pairs table. The other is graph depicting all pairs of means.

The data is summarized as stacked data along with some statistics for each treatment level as shown below. The statistics include the number of results (count), the average, the standard deviation, and the variance for each treatment level. The value of alpha is also given.

	A	B	C	D	E
	270	360	260	320	290
	250	330	230	270	270
	260	310	250	340	240
	230	280	220	300	250
Count	4	4	4	4	4
Average	252.5	320	240	307.5	262.5
Std. Dev.	17.078	33.665	18.257	29.861	22.174
Variance	291.667	1133.333	333.333	891.667	491.667
Alpha	0.05				

The ANOVA table is given next as shown below. This table provides the sum of squares of the treatments, error and total as well as the degrees of freedom. The mean square results are given along with the F value. The key number is the p value. If this is less than or equal to alpha, it is turned red.

ANOVA					
Source	Sum of Squares	Degrees of Freedom	Mean Square	F	p Value
Treatments	19830	4	4957.5	7.89	0.0012
Error	9425	15	628.3333		
Total	29255	19			

The table of means is then given. This table is shown below. The family error rate and the individual rate are given. These are expressed as confidence limits.

Tukey's Test					
Family Conf. Int.=95%, Individual Conf. Int:=99.25%, q(a,f,p)=4.367					
Comparisons	Diff. in Means	Critical Value	LCon	UCon	Sig Diff.?
A - B	-67.5	54.733	-122.233	-12.767	Yes
A - C	12.5	54.733	-42.233	67.233	No
A - D	-55	54.733	-109.733	-0.267	Yes
A - E	-10	54.733	-64.733	44.733	No
B - C	80	54.733	25.267	134.733	Yes
B - D	12.5	54.733	-42.233	67.233	No
B - E	57.5	54.733	2.767	112.233	Yes
C - D	-67.5	54.733	-122.233	-12.767	Yes
C - E	-22.5	54.733	-77.233	32.233	No
D - E	45	54.733	-9.733	99.733	No
There is evidence that some pairs of means are different.					

The columns in the table are:

- Comparisons: all possible combination of pairs of means

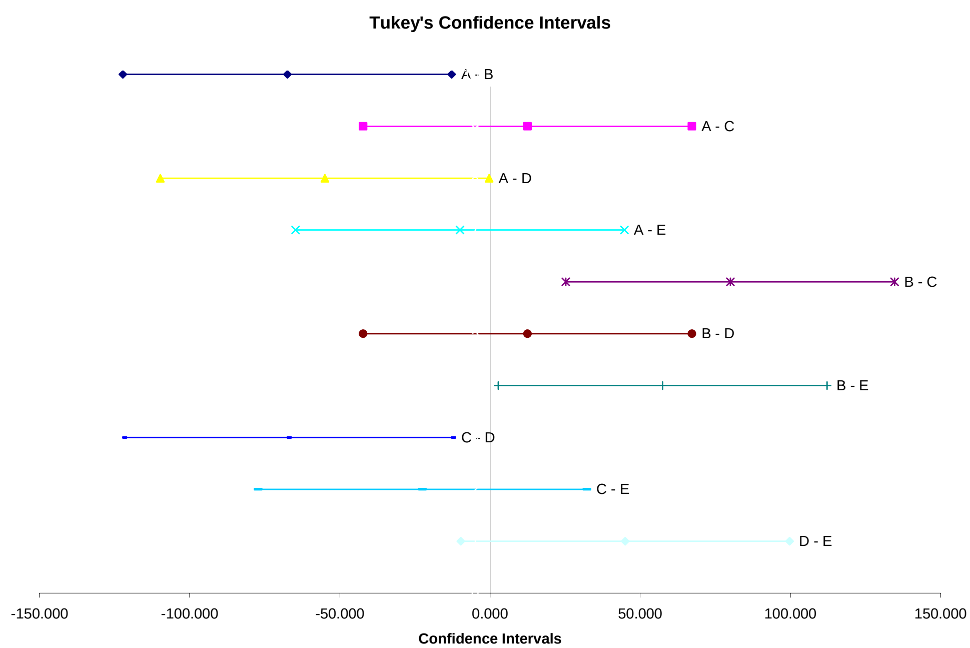
- Diff in Means: the difference in the two means
- Critical Value:[9]

$$T_\alpha = \frac{q_\alpha(a,f)}{\sqrt{2}}\sqrt{MS_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

where $q_\alpha(a,f)$ = the studentized range statistic, f = degrees of freedom associated with MS_E, a = number of treatment levels, MS_E = the mean square error, and n = the sample size for the individual treatment levels.

- LCON: the 100(1-α)% lower confidence interval = difference in means – T_α
- UCON: the 100(1-α)% upper confidence interval = difference in means + T_α
- Sig Diff?: conclusion if there is a significant difference between the two means; if there is a significant difference, Yes is printed and turned to red; otherwise, No is printed. There is a significant difference if the absolute value of the difference in means is greater than the critical value.

The second worksheet added to the workbook contains the confidence interval charts for the pairs of means as shown below. Each confidence interval is labeled. Those that do not include 0 are significantly different.



[9] Montgomery, Douglas C., Design and Analysis of Experiments, Sixth Edition, John Wiley & Sons, Inc, 2005

Bonferroni’s Method for Means

Bonferroni’s Method is used to compare means from multiple processes. The method compares all pairs of means. It controls the family error rate (α). Both the family error and the individual error rates are given in the output. The output from this method also includes the ANOVA table, the table of mean comparisons, and a graph containing all pairs of means.

There are two options for entering the data: stacked or unstacked. Stacked data has each treatment in a single column. Unstacked data has the treatment labels in one column and the results in the adjacent column. Please refer to the Stacked and Unstacked Data Entry section for more information. The example below uses stacked data. The example below shows how to use this method in SPC for MS Excel. This is the same data that was used for the Fisher LSD method.

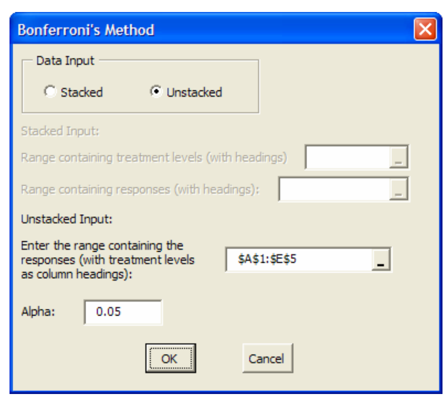
Example[10]

Five different treatments (A to E) were used to control the weight of a coating in grams. There were four samples for each treatment. We want to use Bonferroni’s method to determine if there are any significant differences in the treatments. The data is shown below.

1. Enter the data into a spreadsheet as shown below.
    a. The column headings represent the treatment labels with the responses for each treatment under the label. The number of results for each treatment does not have to be equal.

A	B	C	D	E
250	310	250	340	250
260	330	230	270	240
230	280	220	300	270
270	360	260	320	290

2. Select the data and the column headings.
3. Select the following from the SPC Menu:
    Six Sigma Tools → Other Six Sigma Tools → Multiple Means and Variances → Bonferroni’s Method for Means
4. Fill in the Bonferroni’s Method Input form.
    c. *Data Input:* there are two options: stacked and unstacked; if the range selected on the worksheet contains more than two columns, the default is unstacked; this can be changed; the layout of the data entry is shown above.
    d. *Alpha:* the confidence level; 1- alpha is the confidence interval.
5. Select OK when you have completed the input form.
6. The output for the method is described below.



[10] Example data from Wheeler, Donald, Advanced Topics in Statistical Process Control, SPC Press, Inc, 1995.

Output

Two additional worksheets are added to the workbook for this analysis. One contains the data, the ANOVA table, and the means pairs table. The other is graph depicting all pairs of means.

The data is summarized as stacked data along with some statistics for each treatment level as shown below. The statistics include the number of results (count), the average, the standard deviation, and the variance for each treatment level. The value of alpha is also given.

	A	B	C	D	E
	270	360	260	320	290
	250	330	230	270	270
	260	310	250	340	240
	230	280	220	300	250
Count	4	4	4	4	4
Average	252.5	320	240	307.5	262.5
Std. Dev.	17.078	33.665	18.257	29.861	22.174
Variance	291.667	1133.333	333.333	891.667	491.667
Alpha	0.05				

The ANOVA table is given next as shown below. This table provides the sum of squares of the treatments, error and total as well as the degrees of freedom. The mean square results are given along with the F value. The key number is the p value. If this is less than or equal to alpha, it is turned red.

ANOVA					
Source	Sum of Squares	Degrees of Freedom	Mean Square	F	p Value
Treatments	19830	4	4957.5	7.89	0.0012
Error	9425	15	628.3333		
Total	29255	19			

The table of means is then given. This table is shown below. The family error rate and the individual rate are given. These are expressed as confidence limits.

Bonferroni's Method					
Family Conf. Int.=95%, Individual Conf. Int.=99%					
Comparisons	Diff. in Means	Critical Value	LCon	UCon	Sig Diff?
A - B	-67.5	58.244	-125.744	-9.256	Yes
A - C	12.5	58.244	-45.744	70.744	No
A - D	-55	58.244	-113.244	3.244	No
A - E	-10	58.244	-68.244	48.244	No
B - C	80	58.244	21.756	138.244	Yes
B - D	12.5	58.244	-45.744	70.744	No
B - E	57.5	58.244	-0.744	115.744	No
C - D	-67.5	58.244	-125.744	-9.256	Yes
C - E	-22.5	58.244	-80.744	35.744	No
D - E	45	58.244	-13.244	103.244	No
There is evidence that some pairs of means are different.					

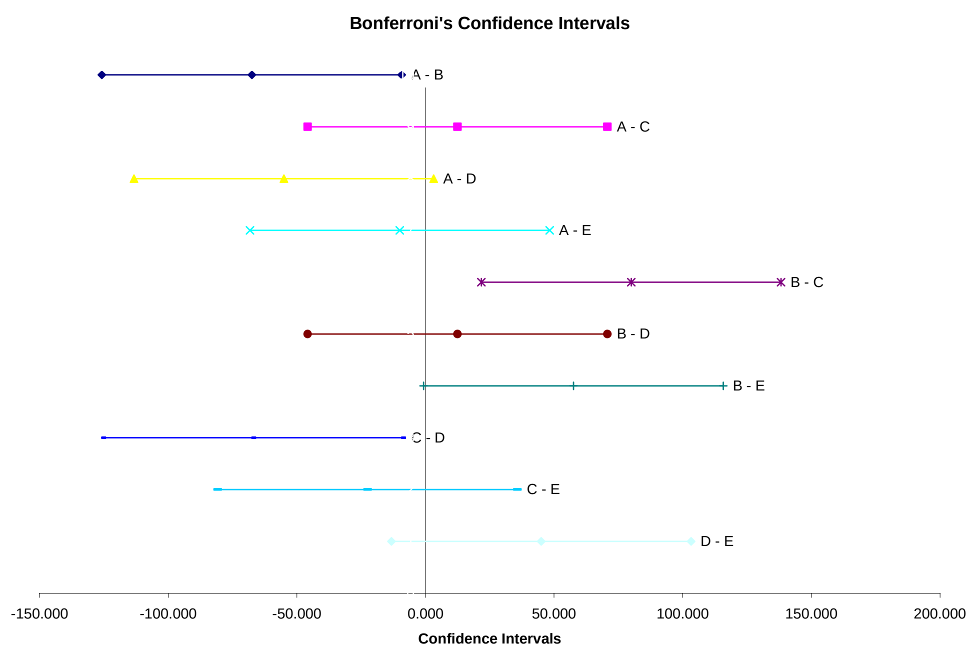
The columns in the table are:

- Comparisons: all possible combination of pairs of means
- Diff in Means: the difference in the two means
- Critical Value:

$$CV = t_{\upsilon,\alpha/k}\sqrt{MS_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

  where t is the t value for υ degrees of freedom and a confidence coefficient of α/k where k is the number of comparisons, MS_E = the mean square error, and n = the sample size for the individual treatment levels.
- LCON: the 100(1-α)% lower confidence interval = difference in means – CV
- UCON: the 100(1-α)% upper confidence interval = difference in means + CV
- Sig Diff?: conclusion if there is a significant difference between the two means; if there is a significant difference, Yes is printed and turned to red; otherwise, No is printed. There is a significant difference if the absolute value of the difference in means is greater than the critical value.

The second worksheet added to the workbook contains the confidence interval charts for the pairs of means as shown below. Each confidence interval is labeled. Those that do not include 0 are significantly different.



Analysis of Means

Analysis of means is a graphical and statistical way of comparing k treatments means with the overall mean using control charts. The method used in this program is described in the book *Advanced Topics in Statistical Control* by Dr. Donald J. Wheeler. The maximum number of treatments is 25. The maximum subgroup size for each treatment is also 25.

There are two options for entering the data: stacked or unstacked. Stacked data has each treatment in a single column. Unstacked data has the treatment labels in one column and the results in the adjacent column. Please refer to the Stacked and Unstacked Entry section in this manual for more information. The example below uses stacked data. The example below shows how to use this method in SPC for MS Excel. This is the same data that was used for the Fisher LSD method.

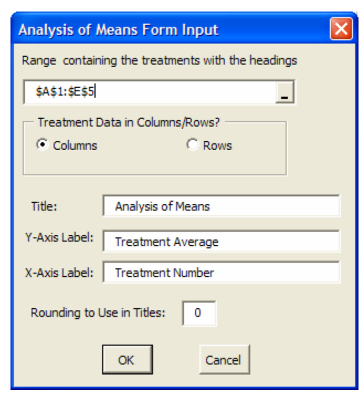
<u>Example</u>[11]

Five different treatments (A to E) were used to control the weight of a coating in grams. There were four samples for each treatment. We want to use Analysis of Means to determine if there are any significant differences in the treatments. The data is shown below.

1. Enter the data into a spreadsheet as shown below.
    a. The column headings represent the treatment labels with the responses for each treatment under the label. The number of results for each treatment does not have to be equal.

A	B	C	D	E
250	310	250	340	250
260	330	230	270	240
230	280	220	300	270
270	360	260	320	290

2. Select the data and the column headings.
3. Select the following from the SPC Menu:
    Six Sigma Tools → Other Six Sigma Tools → Multiple Means and Variances → Analysis of Means
4. Fill the Analysis of Means Input Form
    a. *Range containing the treatments with the headings:* Select both the data and the treatment names.
    b. *Treatment Data in Columns/Row:* The data can either be in columns or rows.
    c. *Title:* The title of the averages chart. Default is "Analysis of Means."
    d. *Y-Axis Label:* Default value is "Treatment Average"
    e. *X-Axis Label:* Default value is "Treatment Number."
    f. *Rounding to Use in Titles:* This is the rounding that is used in the chart titles for the averages and control limits.
5. Select OK when you have completed the input form.
6. The output for the method is described below.



[11] Example and methodology from Wheeler, Donald, <u>Advanced Topics in Statistical Process Control</u>, SPC Press, Inc, 1995.

Output

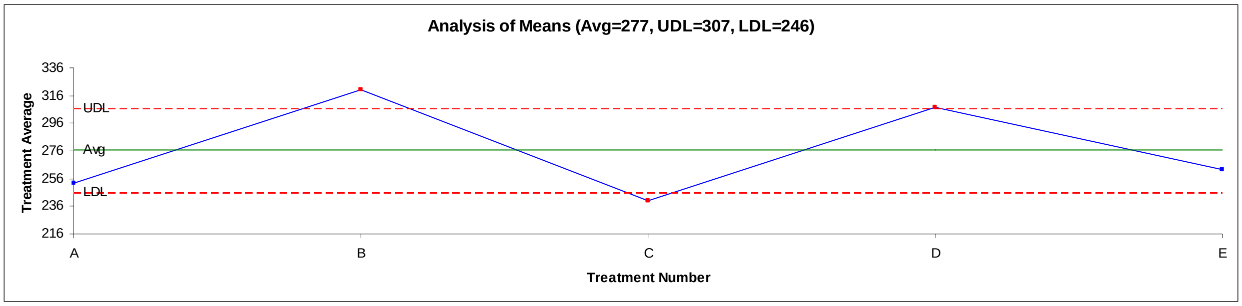
Two additional worksheets are added to the workbook. One worksheet contains the calculations. The other worksheet contains the control charts for the analysis of means. The calculations worksheet is shown below.

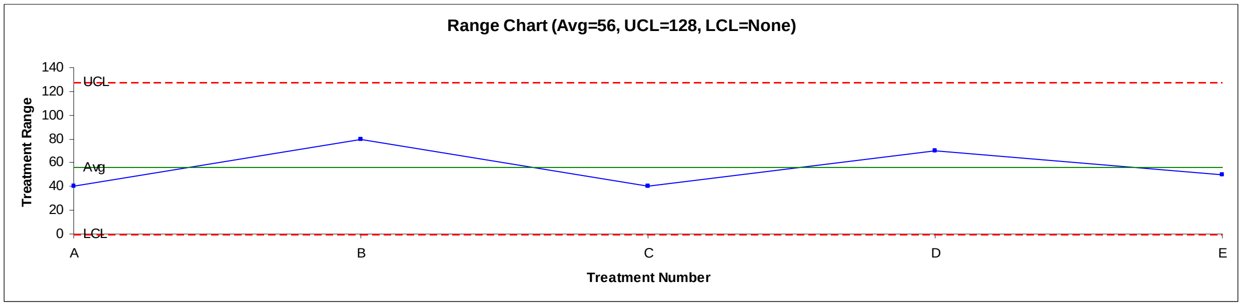
Treatment	A	B	C	D	E
	250	310	250	340	250
	260	330	230	270	240
	230	280	220	300	270
	270	360	260	320	290
Average	252.500	320.000	240.000	307.500	262.500
Maximum	270	360	260	340	290.000
Minimum	230	280	220	270	240.000
Range	40	80	40	70	50.000
Treatment Variance	291.6667	1133.333	333.3333	891.6667	491.667
Treatment DF	3	3	3	3	3
Overall Average	276.500				
Average Range	56.000				
d_2*	2.096				
Est SD(X)	26.718				
Est SD(Xbar)	13.359				
Degrees of Freedom	14				
H	2.267				
ANOM Upper Limit	306.780				
ANOM Lower Limit	246.220				

The treatment averages, maximum, minimum, range, variance, and degrees of freedom are calculated. The overall average and range are calculated. The other parameters in the table are listed below:

- d_2* = bias correction factor for using the average range to calculate the variance
- Est SD(X): the estimated standard deviation of the individual values = $\frac{\overline{R}}{d_2^*}$
- Est SD(Xbar): the estimated standard deviation of the $\overline{X}$ values = $\frac{\text{Est SD(X)}}{\sqrt{n}}$
- Degrees of freedom = the degrees of freedom associated with the average range
- H = the ANOM critical value based on α = 0.10, the number of treatments and the degrees of freedom
- ANOM Upper Limit = the upper decision limit for the analysis of means chart = Xbar + H(Est SD(Xbar)
- ANOM Lower Limit = the lower decision for the analysis of means chart = Xbar – H(Est SD(Xbar)

The second worksheet contains the analysis of means chart as shown below. Means that are in red are significantly different from the other means on the top chart. The bottom chart is a classical range control chart.





The number of observations in each treatment does not have to be equal. If they are not equal, a pooled variance estimation is used and there is no range chart in the output.

Box and Whisker Plots

A Box and Whisker plot is used to present a visual representation of how data is spread out and how much variation there is in the data. It focuses attention on the median, the quartiles, and the minimum and maximum values. The example below demonstrates how to use the Box and Whisker plot in SPC for MS Excel

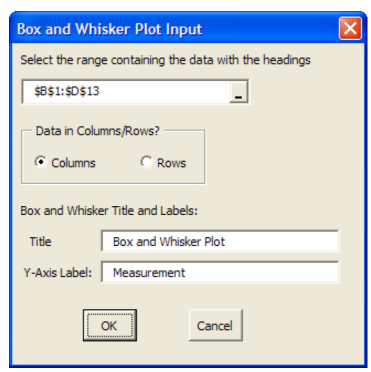
<u>Example</u>

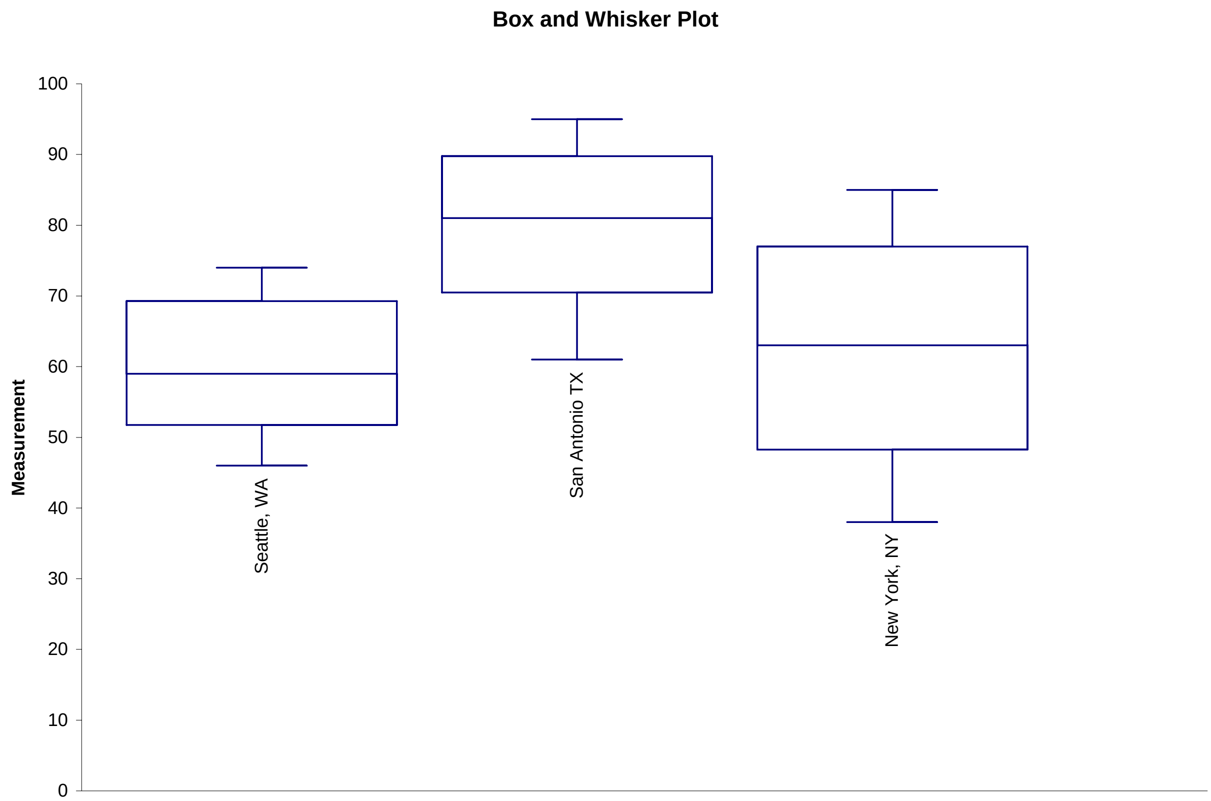
You are interested in comparing the average monthly temperature for three cities. The Box and Whisker plot can show this variation. The data you have is shown below.

1. Enter the data into a spreadsheet as shown below.

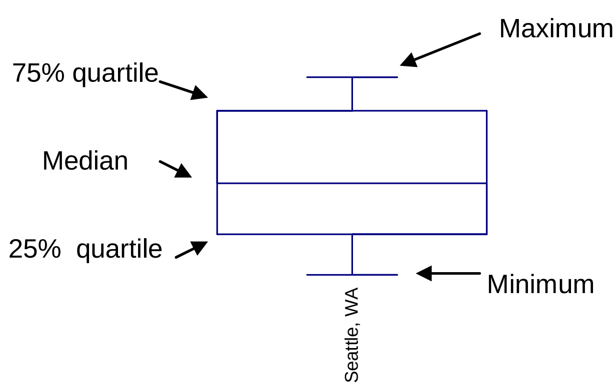
	Seattle, WA	San Antonio TX	New York, NY
Jan	46	61	38
Feb	51	66	40
Mar	54	74	50
Apr	58	80	61
May	64	85	72
June	70	92	80
July	74	95	85
Aug	74	95	84
Sept	69	89	76
Oct	60	82	65
Nov	52	72	54
Dec	46	64	43

2. Select the data and the column headings (do not include the months).
3. Select the following from the SPC Menu:
   Six Sigma Tools → Other Six Sigma Tools → Multiple Means and Variances → Box and Whisker Plots
4. Fill in the Box and Whiskers Plot Input form.
   a. *Select the range:* The default range is the range that is selected on the worksheet.
   b. *Treatment Data in Columns/Rows:* The data can be in columns or rows.
   c. *Box and Whisker Title and Labels*
   d. *Title:* The title that will appear on the chart, default is "Box and Whisker Plot."
   e. *Y and Axis Label:* The label that will appear on the y axis; default is "Measurement."
5. Select OK when you have completed the input form.
6. The Box and Whisker plot is shown on the next page.





The picture below shows the values for each part of the Box and Whisker plot. The maximum value is presented by the top line (whisker). The top part of the box is the 75% quartile; the bottom part is the 25% quartile. The minimum is represented by the bottom line (whisker).



Bartlett's Test for Equality of Variances

Bartlett's Test is used to determine if the variances from multiple treatments (or processes) are the same. These tests can be viewed as the following hypothesis test for k different treatments:

$$H_o = \alpha_1^2 + \alpha_2^2 + \alpha_3^2 + ... + \alpha_k^2$$
$$H_1\text{: the above equality does not hold for at least one } \alpha_i^2$$

There are two options for entering the data: stacked or unstacked. Stacked data has each treatment in a single column. Unstacked data has the treatment labels in one column and the results in the adjacent column. Please refer to the Stacked and Unstacked Data Entry section for more information. The example below uses stacked data. The example below shows how to use this method in SPC for MS Excel. This is the same data that was used for the Fisher LSD method.

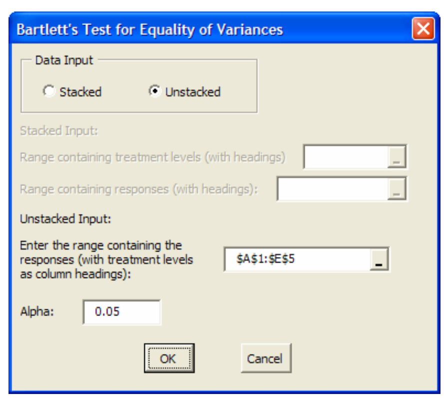
Example[12]

Five different treatments (A to E) were used to control the weight of a coating in grams. There were four samples for each treatment. We want to use Bartlett's method to determine if there are any significant differences in the treatment variances. The data is shown below.

1. Enter the data into a spreadsheet as shown below.
    a. The column headings represent the treatment labels with the responses for each treatment under the label. The number of results for each treatment does not have to be equal.

A	B	C	D	E
250	310	250	340	250
260	330	230	270	240
230	280	220	300	270
270	360	260	320	290

2. Select the data and the column headings.
3. Select the following from the SPC Menu:
    Six Sigma Tools → Other Six Sigma Tools → Multiple Means and Variances → Bartlett's Test for Equality of Variances
4. Fill in the Bartlett's Test for Equality of Variances Input form.
    a. *Data Input:* there are two options: stacked and unstacked; if the range selected on the worksheet contains more than two columns, the default is unstacked; this can be changed; the layout of the data entry is shown above.
    b. *Alpha:* the confidence level; 1- alpha is the confidence interval.
5. Select OK when you have completed the input form.
6. The output for the method is described below.



[12] Example data from Wheeler, Donald, Advanced Topics in Statistical Process Control, SPC Press, Inc, 1995.

Output

A new worksheet is added to the workbook. The top part of the worksheet contains the data in stacked format along with some statistics for each treatment including count, average, standard deviation, and variance. The value of alpha is also given. This part of the output is shown below.

	A	B	C	D	E
	250	310	250	340	250
	260	330	230	270	240
	230	280	220	300	270
	270	360	260	320	290
Count	4	4	4	4	4
Average	252.500	320.000	240.000	307.500	262.500
Std. Dev.	17.078	33.665	18.257	29.861	22.174
Variance	291.667	1133.333	333.333	891.667	491.667
Alpha	0.05				

The bottom part of the worksheet contains the information from Bartlett's test as shown below.

Bartlett's Test for Equality of Variances				
Treatment	Count	Variance	Std. Dev.	
A	4	291.667	17.078	
B	4	1133.333	33.665	
C	4	333.333	18.257	
D	4	891.667	29.861	
E	4	491.667	22.174	
Test Stat	1.889			
Chi Critical	9.488			
p Value	0.7561			
There is no evidence that the variances are different.				

The treatments are listed along with the variance and standard deviation. The following three statistics are given:

- Test Stat: this is the value of the chi-square test statistic:[13]

$$\chi_0^2 = 2.3026\frac{q}{c}$$

$$q = (N-a)\log_{10} S_p^2 - \sum_{i=1}^{a}(n_i - 1)\log_{10} S_i^2$$

$$c = 1 + \frac{1}{3(a-1)}\left(\sum_{i=1}^{a}(n_i - 1)^{-1} - (N-a)^{-1}\right)$$

$$S_p^2 = \frac{\sum_{i=1}^{a}(n_i - 1)S_i^2}{N-a}$$

  where N = total number of observations, a = number of treatments, and S_i^2 is the variance of the ith treatment.
- Chi Critical: this the critical chi-square value based on alpha and the degrees of freedom (which is the number of treatments – 1) = $\chi^2_{\alpha,a-1}$; r eject the null hypothesis if $\chi_0^2 > \chi^2_{\alpha,a-1}$
- p value: Reject the null hypothesis if p value ≤ alpha

The last entry gives the conclusion; in this example, there is no evidence that the variances are different.

---

[13] Montgomery, Douglas C., Design and Analysis of Experiments, Sixth Edition, John Wiley & Sons, Inc, 2005

Modified Levene's Test for Equality of Variances

The Modified Levene's Test is used to determine if the variances from multiple treatments (or processes) are the same. These tests can be viewed as the following hypothesis test for k different treatments:

$$H_o = \alpha_1^2 + \alpha_2^2 + \alpha_3^2 + ... + \alpha_k^2$$

$$H_1\text{: the above equality does not hold for at least one } \alpha_i^2$$

There are two options for entering the data: stacked or unstacked. Stacked data has each treatment in a single column. Unstacked data has the treatment labels in one column and the results in the adjacent column. Please refer to the Stacked and Unstacked Data Entry section for more information. The example below uses stacked data. The example below shows how to use this method in SPC for MS Excel. This is the same data that was used for the Fisher LSD method.

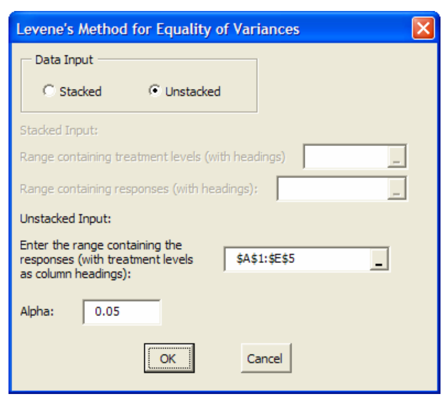
Example[14]

Five different treatments (A to E) were used to control the weight of a coating in grams. There were four samples for each treatment. We want to use Levene's method to determine if there are any significant differences in the treatment variances. The data is shown below.

1. Enter the data into a spreadsheet as shown below.
    a. The column headings represent the treatment labels with the responses for each treatment under the label. The number of results for each treatment does not have to be equal.

A	B	C	D	E
250	310	250	340	250
260	330	230	270	240
230	280	220	300	270
270	360	260	320	290

2. Select the data and the column headings.
3. Select the following from the SPC Menu:
    Six Sigma Tools → Other Six Sigma Tools → Multiple Means and Variances → Modified Levene's Test for Equality of Variances
4. Fill in the Levene's Test for Equality of Variances Input form.
    a. *Data Input:* there are two options: stacked and unstacked; if the range selected on the worksheet contains more than two columns, the default is unstacked; this can be changed; the layout of the data entry is shown above.
    b. *Alpha:* the confidence level; 1- alpha is the confidence interval.
5. Select OK when you have completed the input form.
6. The output for the method is described below.



[14] Example data from Wheeler, Donald, Advanced Topics in Statistical Process Control, SPC Press, Inc, 1995.

Output

A new worksheet is added to the workbook. The top part of the worksheet contains the data in stacked format along with some statistics for each treatment including count, average, standard deviation, and variance. The value of alpha is also given. This part of the output is shown below.

	A	B	C	D	E
	250	310	250	340	250
	260	330	230	270	240
	230	280	220	300	270
	270	360	260	320	290
Count	4	4	4	4	4
Average	252.500	320.000	240.000	307.500	262.500
Std. Dev.	17.078	33.665	18.257	29.861	22.174
Variance	291.667	1133.333	333.333	891.667	491.667
Alpha	0.05				

The bottom part of the worksheet contains the information from Levene's Method as shown below.

Modified Levene's Test for Equality of Variances			
Treatment	Count	Variance	Std. Dev
A	4	291.667	17.078
B	4	1133.333	33.665
C	4	333.333	18.257
D	4	891.667	29.861
E	4	491.667	22.174
F Levene	0.72		
F Critical	3.06		
p value	0.5886		
There is no evidence that the variances are different.			

The treatments are listed along with the variance and standard deviation. The following three statistics are given:

- F Levene: This is the Levene statistic given by :

$$F_{Levene} = \frac{(N-a)\sum_{i=1}^{a} n_i (D_i - \overline{D})^2}{(a-1)\sum_{i=1}^{a}\sum_{j=1}^{n_i}(D_{ij} - \overline{D}_i)^2}$$

  where N = total number of observations and a = number of treatments.

- F Critical: This the critical value of F based on alpha with a-1 degrees of freedom in the numerator and N-t degrees of freedom in the denominator; reject the null hypothesis if F Levene > F Critical.

- p value: reject the null hypothesis if p value ≤ alpha

The last entry gives the conclusion; in this example, there is no evidence that the variances are different.

Cause and Effect

Correlation Coefficients

The linear correlation coefficient, R, is a measure of the association between two variables. The maximum value for R is + 1. The minimum value for R is - 1. In both these cases, all sample points fall on a straight line. As R approaches +1 or -1, the stronger the correlation between x and y. The square of this coefficient (R^2) indicates the fraction of variation in y that is associated with x. An example of how to use this test is given below.

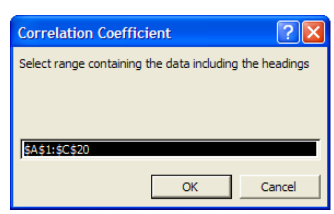
<u>Example</u>

An engineer wants to know if there is a linear correlation between the hardening temperature, the tempering temperature, and the hardness of steel part. The data that has been collected is shown below.

1. Enter the data into a spreadsheet as shown below.

Hardening Temp.	Tempering Temp	Hardness
1670	1195	33.8
1665	1235	30
1665	1185	33.8
1635	1160	31.6
1640	1105	32.6
1640	1105	32.7
1680	1170	161
1640	1210	28.7
1660	1255	31.5
1670	1320	28.8
1670	1200	34.7
1665	1235	29.6
1660	1255	31.6
1635	1160	32.8
1700	1190	29.6
1675	1262	31.3
1660	1255	31.6
1605	990	303.7
1650	1175	31.5

2. Select the data and the headings.
3. Select the following from the SPC Menu:
   Six Sigma Tools → Other Six Sigma Tools → Cause and Effect → Correlation Coefficients
4. Ensure that the range in the input box is correct.
5. Select OK.
6. The output for the method is described below.



Output

The output for the example data is shown below.

R/Prob	Hardening Temp.	Tempering Temp	Hardness
Hardening Temp.	1	0.661	-0.44
	0	0.002	0.06
Tempering Temp	0.661	1	-0.662
	0.002	0	0.002
Hardness	-0.44	-0.662	1
	0.06	0.002	0

A new worksheet is added that contains the results. The variables are listed in the first row across and down the first column. For each pair of variables, the correlation coefficient is given on one row and the p value is given on the row below it. If the p value is ≤ 0.05, it is highlighted in red.

Formulas

Pearson's Correlation Coefficient: $R = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$

where:
n = number of paired data points for the variable x and the variable y
s_x = standard deviation of x
s_y = standard deviation of y
$\bar{x}$ = average of x
$\bar{y}$ = average of y

p value: determine from a two-tailed test using the t distribution with $t = \dfrac{R\sqrt{n-2}}{\sqrt{1-R^2}}$

Failure Mode and Effects Analysis Template

A Failure Mode and Effects Analysis (FMEA) template is included in the program. To access the template from the SPC Menu:

Six Sigma Tools → Other Six Sigma Tools → Cause and Effect → Failure Mode and Effects Analysis

The template is shown below. The sheet is set up to automatically calculate the Risk Priority Number when the severity, occurrence, and detection ratings are entered. Some cells are protected. You can unprotect the worksheet since there is no password. If you want to add multiple lines in a cell, such as under the "What are the Potential Effects of the Failure Mode", you enter a line, then hit ALT + Enter.

Failure Mode and Effects Analysis

Prepared by: **Date Prepared:**

What is the Process?	What is the Potential Failure Mode?	What are the Potential Effects of the Failure Mode	Severity	What are the Possible Causes of the Failure?	Occurrence	How Do We Currently Prevent Each Listed Cause of Failure from Happening?	Detection	S	O	D	RPN

Recommended Action Steps and Status				Due Date	S	O	D	New RPN
No.	What Needs to Be Done	Responsibility	Status					

Plot Multiple Y Variables against One X Variable

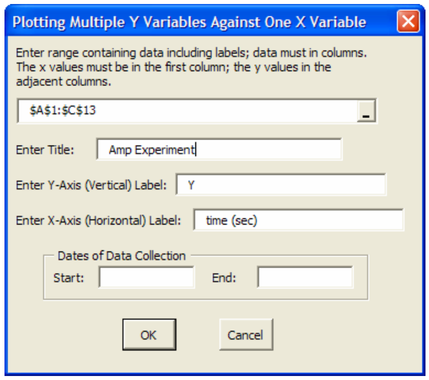
The program will plot multiple Y variables against one X variable. The data for this chart must be in columns with the X variable in the first column. The Y variables must be in adjacent columns. An example of how to create this chart is given below for plotting two Y variables against the X variable.

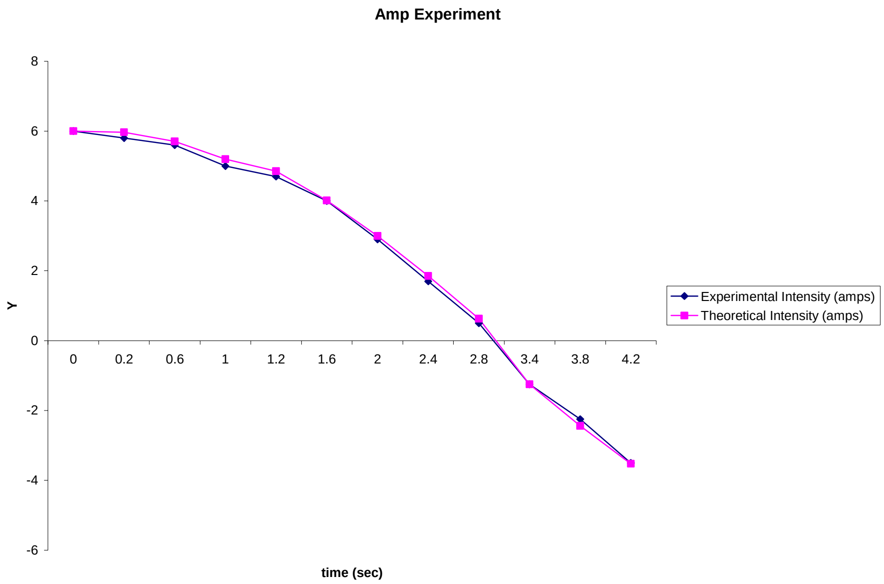
<u>Example</u>

An experiment has been to do to measure the amps from a process over time. The experimenter wants to compare the measured amps to the theoretical amp.

1. Enter the data into a spreadsheet as shown to the right.
2. Select the data and the column headings.
3. Select the following from the SPC Menu:
   Six Sigma Tools → Other Six Sigma Tools → Cause and Effect → Plot Multiple Y Variables Against One X Variable
4. Fill in the Plot Multiple Y Variables Against One X Variable Input form.
   a. *Range containing the data including the labels:* the default value is the range selected on the worksheet.
   b. *Enter Title:* this is the title of the chart
   c. *Enter Y*-Axis (Vertical) Label: enter the label for the y axis, default is Y.
   d. *Enter X*-Axis (Horizontal) Label: enter the label for the x axis; default is the column heading for the x variable
   e. *Date of Data Collection:* you may enter the dates of data collection; this is optional.
5. Select OK.
6. The chart for the example data is shown below.

time (sec)	Experimental Intensity (amps)	Theoretical Intensity (amps)
0	6	6
0.2	5.8	5.9671314
0.6	5.6	5.7063391
1	5	5.1961524
1.2	4.7	4.854102
1.6	4	4.0147836
2	2.9	3
2.4	1.7	1.854102
2.8	0.5	0.6271708
3.4	-1.25	-1.2474701
3.8	-2.25	-2.4404199
4.2	-3.5	-3.5267115





Scatter Plot Matrix

A scatter plot matrix is a two-dimensional array of scatter diagrams between every possible pair of variables in the data. This allows you to quickly scan for relationships between the variables. An example of how to do a scatter plot matrix is given below.

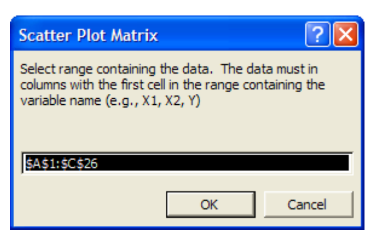
Example[15]

A soft drink bottler is trying to predict delivery times for a driver. He has collected data on the delivery time, the number of cartons delivered and the distance the driver walked. He wants to see the relationship between these three variables. We will use the scatter plot matrix to do this.

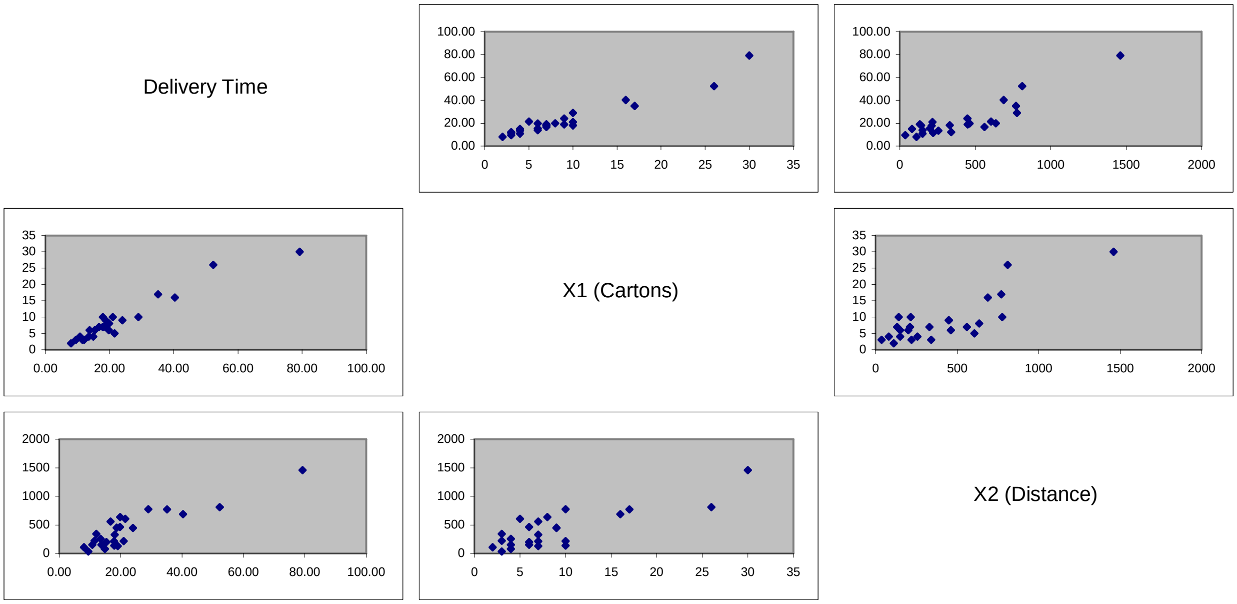
1. Enter the data into a spreadsheet as shown below.

Delivery Time	X1 (Cartons)	X2 (Distance)
16.68	7	560
11.50	3	220
12.03	3	340
14.88	4	80
13.75	6	150
18.11	7	330
8.00	2	110
17.83	7	210
79.24	30	1460
21.50	5	605
40.33	16	688
21.00	10	215
13.50	4	255
19.75	6	462
24.00	9	448
29.00	10	776
15.35	6	200
19.00	7	132
9.50	3	36
35.10	17	770
17.90	10	140
52.32	26	810
18.75	9	450
19.83	8	635
10.75	4	150

2. Highlight the data and the headings.
3. Select the following from the SPC Menu:
   Six Sigma Tools → Other Six Sigma Tools → Cause and Effect → Scatter Plot Matrix
4. Ensure the correct range is in the Scatter Plot Matrix input box.
5. Select OK.
6. The resulting scatter plot matrix for this example is shown on the next page.



[15] Montgomery, D.C., Peck, E. A., Vinning, G, G, Introduction to Linear Regression Analysis, Wiley & Sons, 2006.



Note that the diagonals contain the names of the variables. That is the y axis variable for that row of charts. For example, on the first row of charts above, the delivery time is the y axis. The first chart has X1 as the x axis and the last chart in the row has X2 as the x axis.

Miscellaneous

Normal Probability Plot with Anderson-Darling Test

This test helps determine if your data is normally distributed. Data that is normally distributed will fall close to a straight line on the normal probability plot. The Anderson-Darling test calculates a statistic that helps detects most departures from normality.

If the p value (probability) for the Anderson-Darling statistic is less than 0.05, there is statistical evidence that the data is not normality distributed. If the p value is greater than 0.20, the conclusion is that the data is normally distributed. More data might be needed for values of p between 0.05 and 0.20.

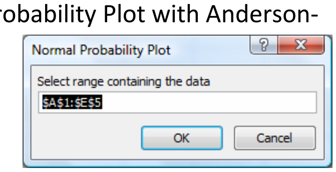
The example below shows how to do this test using SPC for MS Excel.

<u>Example</u>

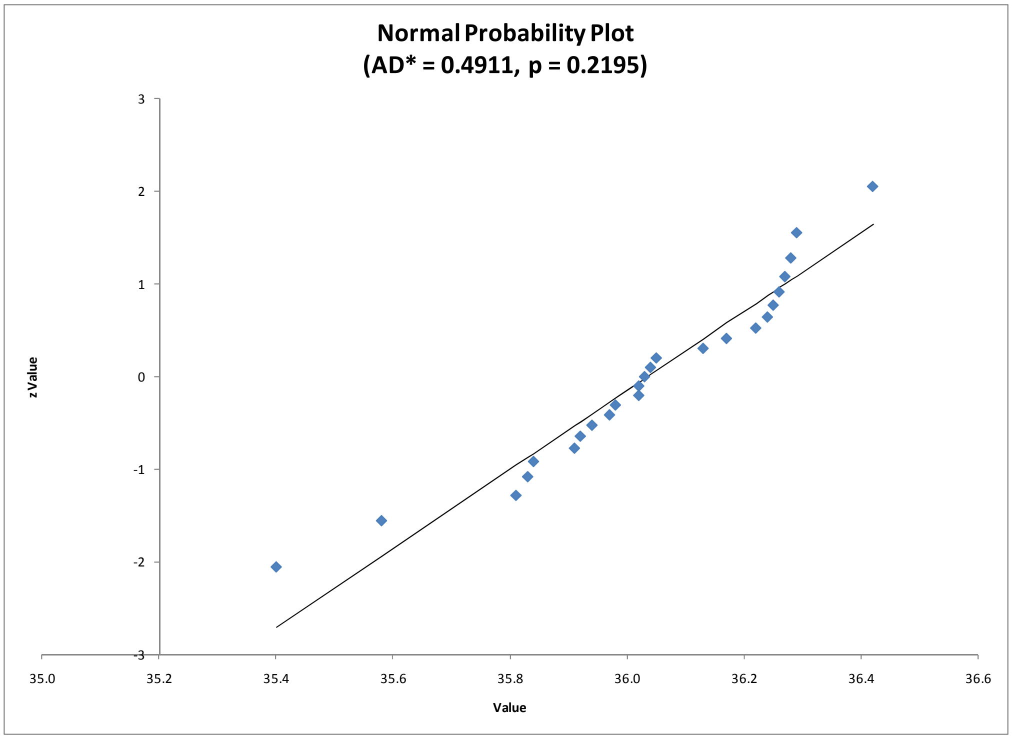
A hardness tester's calibration is checked by testing a standard block once a day. The measurement error should be normally distributed. Data for the last 25 days is given below. We want to determine if the results are normally distributed.

35.4	36.25	36.13	36.02	36.26
36.05	35.97	36.17	35.91	36.03
35.81	36.22	35.94	35.58	36.42
36.02	36.27	36.29	35.83	36.24
35.92	36.28	36.04	35.84	35.98

1. Select the data in the table.
2. Select the following from the SPC Menu:
   Six Sigma Tools → Other Six Sigma Tools → Miscellaneous → Normal Probability Plot with Anderson-Darling Test option.
3. Enter the range into the input form that is displayed (default is the range selected on the worksheet).
4. Select OK.
5. The normal probability plot on the next page is generated.



The points generally lie along the straight line on the normal probability plot. The Anderson-Darling statistic is printed along with the p value, which in this case is 0.2195. Since it is greater than 0.20, you conclude that the data comes from a normal distribution.


Normal Probability Plot
(AD* = 0.4911, p = 0.2195)
3
2
1
0
-1
-2
-3
z Value
35.0
35.2
35.4
35.6
35.8
36.0
36.2
36.4
36.6
Value

Distributions

This software has the ability to visually display 9 distributions. This is very helpful for understanding what a distribution looks like and how it shape changes as you change parameters. There are three discrete distributions and six continuous distributions.

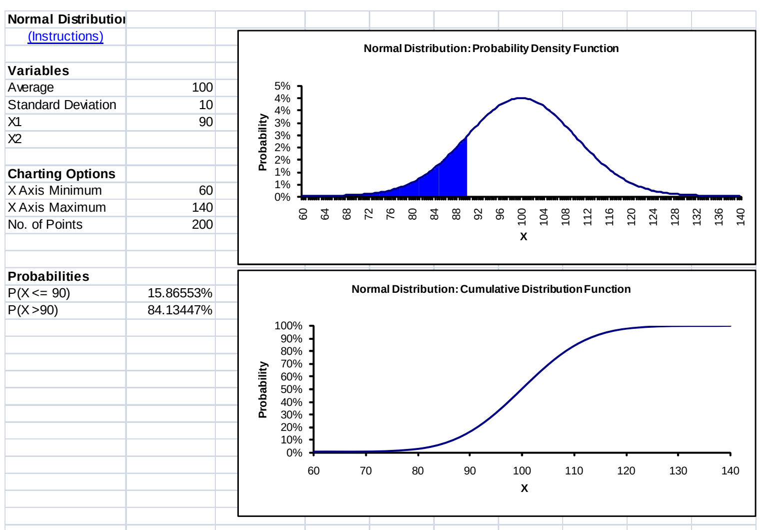
Discrete	Continuous
• Binomial • Hypergeometric • Poisson	• Beta • Exponential • Gamma • Lognormal • Normal • Weibull

The example below shows how to use this feature for the normal distribution.

Example

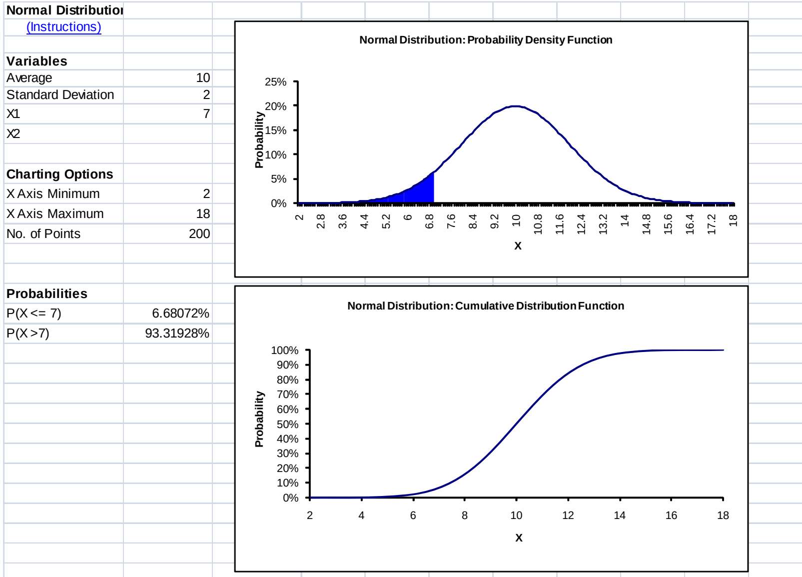
You are interested in seeing how the normal distribution changes as the average and standard deviation change.

1. Select the Six Sigma Tools icon from the SPC Menu to show the form "Other Six Sigma Tools."
2. Select the Miscellaneous page.
3. Select the Distributions Option
4. Select the Normal Distribution Option
5. A new worksheet is added to the workbook as shown below.
   a. There are more instructions on the worksheet inserted into workbook (click on instructions in cell A2)

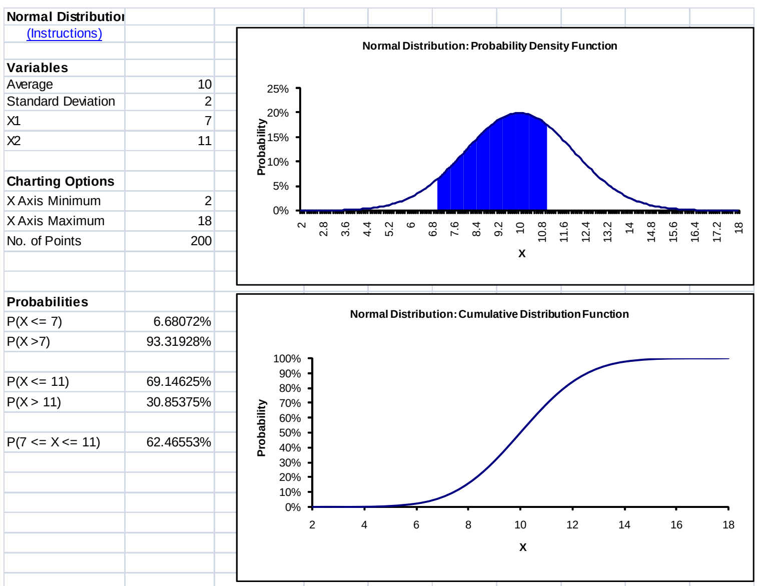


The output shows two curves: the probability density function and the cumulative distribution function. The probability density function shades the area less than or equal to X1. . You can adjust the variables for each distribution. You can enter a new average, standard deviation, and value of X1 in the worksheet. For example, you can enter these values: average = 10, standard deviation = 2 and X1 = 7.

To update the charts with the new information, select Update Charts icon from the SPC toolbar. You will see the output below.



You can also enter a value for X2. The probability density function will then shade the area between X1 and X2. The output when X2=11 is shown below.



The output for the other distributions is very similar to the normal distribution.

Side by Side Histogram

A side by side histogram is used to compare results. For example, you may have done a survey where respondents gave a rating between 0 and 100 for a product. You want to visually compare the responses by males and by females. You can use a side by side histogram is do this as is shown below.

Example

1. Enter the data into a spreadsheet. You have two options for entering the data.
    a. Option 1: The data is already summarized by groups. In this example, you have grouped the items based on ratings in sizes of 10 as shown below.

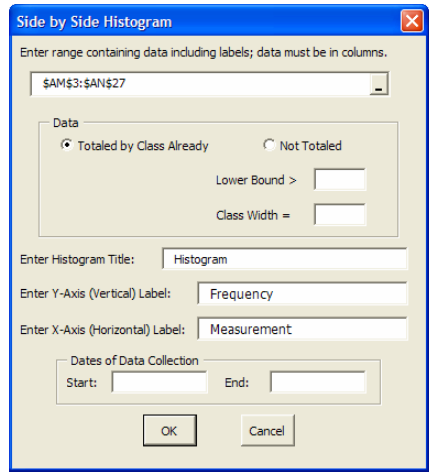
Grouping	Female	Male
0-10	1	1
11-20	3	6
21-30	1	0
31-40	0	1
41-50	2	5
51-60	4	4
61-70	3	2
71-80	3	2
81-90	3	3
91-100	4	0

    b. Option 2: The raw data is listed in columns based on how you want to view the data (in this case, male and female) as shown below.

Female	Male
41	18
99	89
78	56
96	78
99	88
58	45
69	51
85	65
54	45
79	60
65	47
7	4
52	36
23	14
15	12
63	45
41	15
78	55
12	11
87	76
96	87
13	12
55	45
85	64

2. Highlight the data in the worksheet including the column headings.
3. Select the Six Sigma Tools icon from the SPC Menu to show the form “Other Six Sigma Tools.”

4. Select the Miscellaneous page
5. Select the Side by Side Histogram option.
6. Fill out the Side by Side Histogram Input form.

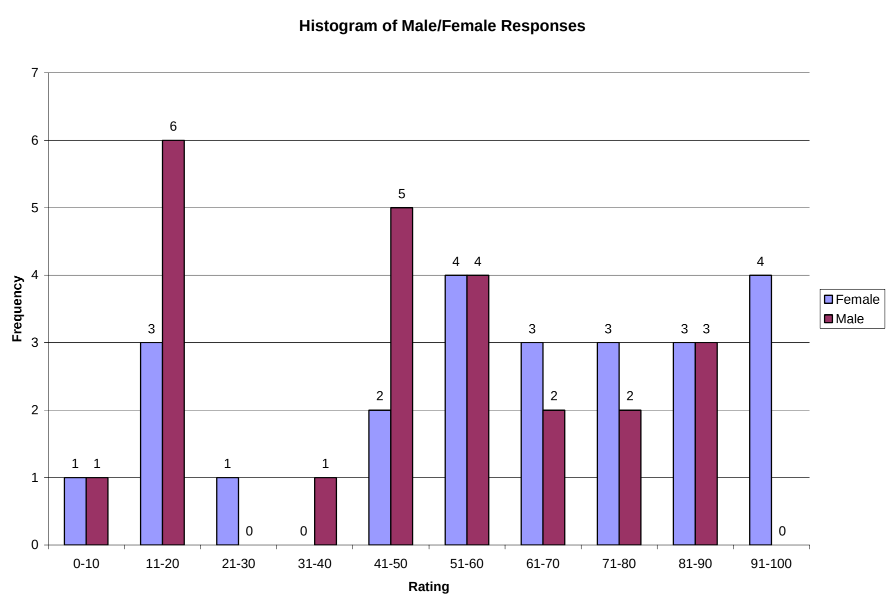


   a. *Enter range containing data including labels; data must be in columns:* Enter the worksheet range containing the data including the column headings (see examples above). The data must be in columns. The default value is the range selected on the worksheet before selecting the Side by Side Histogram option.

- *Data:* There are the two options explained above.
  - o *Totaled by Class Already:* The data has already be summed by class.
  - o *Not Totaled:* The data has not been totaled. The program will perform this function based on the following two values:
    - ▪ *Lower Bound >:* This value gives the starting point for the histogram. For example, if you enter 0 here, the first class will involve values greater than 0 to the class width
    - ▪ *Class Width =:* This is the width of one class.
- *Enter Histogram Title:* Enter the title that will appear on the chart. The default title is "Histogram."
- *Enter Y*-Axis (Vertical Label): Enter the label for the y axis. The default value is "Frequency."
- *Enter X*-Axis (Horizontal Label): Enter the label for the x axis. The default value is "Measurement."
- *Date of Data Collection:* You can enter the dates of data collection. This is optional.

7. Select OK to create the side by side histogram.

The side by side histogram for this example is shown below.



Chi-Square Goodness of Fit Test

This test is used to determine if a set of data fits a particular distribution. The chi-square statistic is used to determine if the data fits a particular distribution. The example below shows how to do this test using the SPC for MS Excel software.

Example[16]

A 1959 study measured the number of passengers in a car in urban traffic at one intersection. The data is given below. Does this data represent a Poisson distribution? There are 1,011 observations.

Number of Passengers	0	1	2	3	4	5+
Observed Frequency	678	227	56	28	8	14

To use this technique, you must determine the expected frequency or number. To do this, first determine the average of the observed frequencies. This is defined as the following:

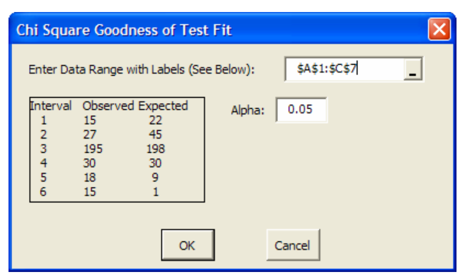
$$\lambda = [(678*0)+(227*1)+(56*2)+(28*3)+(8*4)+(14*5)]/(678+227+56+28+8+14) = 0.519$$

You can use Excel's Poisson function to find the expected values. For each number of passengers, use POISSON(x, 0.519, False) to find the expected value where x is the number of passengers. For example, for x = 0, the expected value is 602. Once this is complete, you can apply the Chi-Square Goodness of Fit test.

1. Enter the data into an Excel spreadsheet as shown below.

Number of Passengers	Observed Frequency	Expected Frequency
0	678	601.66
1	227	312.26
2	56	81.03
3	28	14.02
4	8	1.82
5	14	0.19

1. Select all the data in the table above including the headings.
2. From the SPC menu, select "Six Sigma → Miscellaneous page → Chi-Square Goodness of Fit Test
3. Fill in the Chi-Square Test for Association Input form.
    a. *Enter Data Range with Labels:* this is the range containing the table above; the default is the range selected on the worksheet.
    b. *Alpha:* this is the confidence coefficient
4. Select OK.
5. The output from the test is shown on the next page.



[16] Example from Statistics and Data Analysis, by Ajit Tamhane and Dorothy Dunlop, Prentice-Hall, 2000.

Chi Square Results for Goodness of Fit			
H_0: The data is consistent with the specified distribution.			
H_1: The data is not consistent with the specified distribution.			
Number of Passengers	**Observed Frequency**	**Expected Frequency**	**Contribution to Chi-Square**
0	678	601.66	9.68619
1	227	312.26	23.27954
2	56	81.03	7.73172
3	28	14.02	13.94011
4	8	1.82	20.98484
5	14	0.19	1003.76895
Alpha	0.05		
χ^2	1079.391		
$\chi^2_{(0.05, 5)}$	11.070		
p Value	0.0000		
The null hypothesis is rejected.			
There is evidence that the data is not consistent with the specified distribution.			

The null and alternate hypothesis is printed at the top of a new worksheet. The data is printed and the contribution to chi-square determine. The value of alpha is printed and the calculated and critical values for chi-square are calculated. The p value is determined. The conclusion is then generated based on the values of p and alpha. In this example, the null hypothesis is rejected. There is evidence that the data does not follow a Poisson distribution.

Chi-Square Test for Association

The chi-square test for association is used to determine if there is any association between two variables. It is really a hypothesis test of independence. The null hypothesis is that the two variables are not associated, i.e., independent. The alternate hypothesis is that the two variables are associated. The example below shows how to do this test using the SPC for MS Excel software.

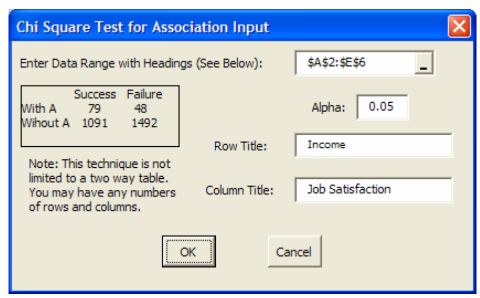
Example[17]

A survey was done to determine if job satisfaction was related to income. A total of 901 people participated in the survey. The data is shown below. We will use the chi-square test for association to determine if the two variables are associated.

6. Enter the data into an Excel spreadsheet as shown below.

	Job Satisfaction			
Income	Very Dissatisfied	Little Dissatisfied	Moderately Satisfied	Very Satisfied
<6000	20	24	80	82
6000 - $15000	22	38	104	125
15000 - 25000	13	28	81	113
>25000	7	18	54	92

7. Select the shaded area as shown in the table above.
8. Select the Six Sigma Tools icon from the SPC Menu to show the form "Other Six Sigma Tools."
9. Select the Miscellaneous page
10. Select the Chi-Square Test for Association option.
11. Fill in the Chi-Square Test for Association Input form.
    a. *Enter Data Range with Headings:* this is the range that contains the data plus the row and column headings.
    b. *Alpha:* (1- alpha) is the confidence level; default is 0.05.
    c. *Row Title:* This is the heading for the levels of the variable listed in the rows in the first column; the default is the first cell in the data range.
    d. *Column Title:* This is the heading for the levels of the variable listed in the columns in the first row of the selected range. The default is the cell above the first row in the second column (see example above).
12. Select OK to run the test.
13. The output for this example is shown on the next page and explained below.



The top part of the output contains the data with the observed and expected values as well as the contribution of each to χ^2. The row and column totals are also given.

The middle portion of the output contains the following:

- Alpha (entered)
- The calculated χ^2.
- The degrees of freedom
- The critical χ^2 value based on alpha and the degrees of freedom
- The calculated p value (will be in red if ≤ alpha)

[17] Example from Statistics and Data Analysis, by Ajit Tamhane and Dorothy Dunlop, Prentice-Hall, 2000.

-

Chi Square Results for Association						
		Job Satisfaction				
Income		Very Dissatisfied	Little Dissatisfied	Moderately Satisfied	Very Satisfied	Row Total
<6000	Observed	20	24	80	82	206
	Expected	14.18	24.69	72.93	94.20	
	Contribution to χ^2	2.393	0.019	0.684	1.579	
6000 - $15000	Observed	22	38	104	125	289
	Expected	19.89	34.64	102.32	132.15	
	Contribution to χ^2	0.225	0.326	0.028	0.387	
15000 - 25000	Observed	13	28	81	113	235
	Expected	16.17	28.17	83.20	107.46	
	Contribution to χ^2	0.622	0.001	0.058	0.286	
>25000	Observed	7	18	54	92	171
	Expected	11.77	20.50	60.54	78.19	
	Contribution to χ^2	1.931	0.304	0.707	2.438	
Column Total		62	108	319	412	901
Alpha	0.05					
χ^2	11.989					
Degrees of Freedom	9					
$\chi^2_{(0.05, 9)}$	16.919					
p Value	0.2140					
Residuals						
	Job Satisfaction					
Income	Very Dissatisfied	Little Dissatisfied	Moderately Satisfied	Very Satisfied		
<6000	5.82	-0.69	7.07	-12.20		
6000 - $15000	2.11	3.36	1.68	-7.15		
15000 - 25000	-3.17	-0.17	-2.20	5.54		
>25000	-4.77	-2.50	-6.54	13.81		
The null hypothesis is not rejected.						
There is no evidence that Income and Job Satisfaction are associated						

The bottom portion of the output contains the residuals. The residuals are the difference between the observed and the expected values. The conclusion is then given based on the values of alpha and the p value. The null hypothesis (that the variables are not associated) is rejected if the p value ≤ alpha.

Nonparametric

One Sample Sign Test

Many statistical methods require the assumption of normally distributed data. This assumption often does not hold. In addition, some sample sizes are so small that you can't verify if the distribution is normal. Nonparametric statistical methods do not make too many assumptions about the population from which the sample is drawn. The one sample sign test is one of these nonparametric statistical methods.

This method is based on collecting a number of samples from a population with unknown median, m. The median is used in place of the average because the median is a better measure of the central tendency of the data than the average for non-normal data. The example below shows how to do this test using the SPC for MS Excel software.

Example[18]

A thermostat is used in an electrical device to check to the accuracy of the design setting of 200 °F. Ten thermostats were tested to determine their actual settings. The results are given below.

202.2, 203.4, 200.5, 202.5, 206.3, 198.0, 203.7, 200.8, 201.3, 199.0

We want to use these results to see if the median setting is significantly different from the design setting of 200.

1. Enter the results into an Excel spreadsheet as shown and highlight the data (not the headings).

Settings
202.2
203.4
200.5
202.5
206.3
198.0
203.7
200.8
201.3
199.0

2. Fill in the One Sample Sign Test form.
    a. *Range Containing Data (no heading):* this is the range containing the data; default is the range selected on the worksheet
    b. *Variable:* name of the variable; default is the cell above the data selected on the worksheet
    c. *Alpha:* (1- alpha) is the confidence level; default is 0.05.
    d. *Specified Median:* enter the specified median; 0 is default; 200 entered for this example
    e. *Type:* type of test: two sided, lower one-sided, or upper one-sided; default is two-sided.
    f. *Include confidence interval:* check if you want the confidence interval displayed; default is not to display the confidence interval.

[18] Example from Statistics and Data Analysis, by Ajit Tamhane and Dorothy Dunlop, Prentice-Hall, 2000.

g. *Output option:* select the first cell of an output range on an existing worksheet and select a cell (a check is made to ensure that there is nothing in the cells that will be written over) or select new worksheet (the default).

3. Select OK once your have completed your entries.
4. The output is shown below with an explanation of the terms to the right.

Sign Test for Settings		Explanation (not part of output)
H_0: m = m_0		Null hypothesis (m is median)
H_1: m <> m_0		Alternate hypothesis (two sided in this case)
Median (m)	201.75	Calculated median from the data
Alpha	0.05	Value of alpha entered by user
Specified Median (m_0)	200	Specified median entered by user
Number Below m_0	2	Number of values below the median
Number = m_0	0	Number of values equal to the median
Number Above m_0	8	Number of values above the median
Sample Size (Less = m_0)	10	Sample size not including values equal to specified median
p Value	0.1094	Calculated p value (will be red if ≤alpha)
The null hypothesis is not rejected.		Conclusion based on value of p and alpha.
There is no evidence that the median does not equal 200.		Reject the null if p value ≤ alpha
97.85% Confidence Interval		
Lower	199	Calculated confidence intervals
Upper	203.7	
89.06% Confidence Interval		
Lower	200.5	
Upper	203.4	

Notes:

- Any values equal to the median are not included; this reduces the effective sample size but should not be an issue as long as there are not too many values equal to the median.
- The software uses the binomial distribution for the calculations and does not use a normal approximation as the sample size becomes large.
- Exact confidence intervals cannot usually be calculated because of the discreteness of the binomial distribution. Usually two confidence intervals will be given: the one above (1-alpha) % and one below it.

One Sample Wilcoxon Signed Rank Test

Many statistical methods require the assumption of normally distributed data. This assumption often does not hold. In addition, some sample sizes are so small that you can't verify if the distribution is normal. Nonparametric statistical methods do not make too many assumptions about the population from which the sample is drawn. The one sample Wilcoxon signed rank test is one of these nonparametric statistical methods.

This method is based on collecting a number of samples from a population with unknown median, m. The median is used in place of the average because the median is a better measure of the central tendency of the data than the average for non-normal data. The example below shows how to do this test using the SPC for MS Excel software.

Example[19]

A thermostat is used in an electrical device to check to the accuracy of the design setting of 200 °F. Ten thermostats were tested to determine their actual settings. The results are given below.

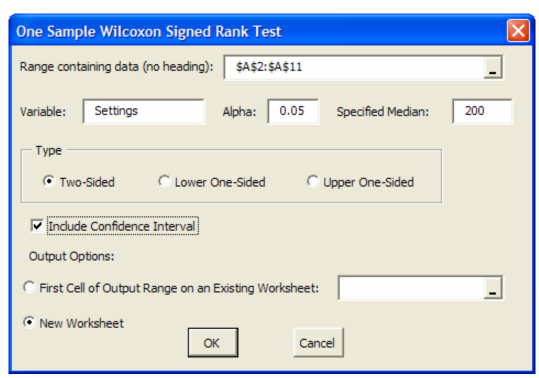
202.2, 203.4, 200.5, 202.5, 206.3, 198.0, 203.7, 200.8, 201.3, 199.0

We want to use these results to see if the median setting is significantly different from the design setting of 200.

1. Enter the results into an Excel spreadsheet as shown and highlight the data (not the headings).

Settings
202.2
203.4
200.5
202.5
206.3
198.0
203.7
200.8
201.3
199.0

2. Select the Six Sigma Tools icon from the SPC Menu to show the form "Other Six Sigma Tools."
3. Select the Nonparametric page (the last one).
4. Select the One Sample Signed Wilcoxon Rank Test option.
5. Fill in the One Sample Sign Test form.
    a. *Range Containing Data (no heading):* this is the range containing the data; default is the range selected on the worksheet
    b. *Variable:* name of the variable; default is the cell above the data selected on the worksheet
    c. *Alpha:* (1- alpha) is the confidence level; default is 0.05.
    d. *Specified Median:* enter the specified median; 0 is default; 200 entered for this example



[19] Example from Statistics and Data Analysis, by Ajit Tamhane and Dorothy Dunlop, Prentice-Hall, 2000.

e. *Type:* type of test: two sided, lower one-sided, or upper one-sided; default is two-sided.
f. *Include confidence interval:* check if you want the confidence interval displayed; default is not to display the confidence interval.
g. *Output option:* select the first cell of an output range on an existing worksheet and select a cell (a check is made to ensure that there is nothing in the cells that will be written over) or select new worksheet (the default).

6. Select OK once your have completed your entries.
7. The output is shown below with an explanation of the terms to the right.

Wilcoxon Signed Rank Test		Explanation (not part of output)
H_0: m = m_0		Null hypothesis (m is median)
H_1: m <> m_0		Alternate hypothesis (two sided in this case)
Median	201.75	Calculated median of the data
Specified Median	200	Specified median entered by user
Sample Case	Small	Calculations; if sample size < 20, the null distribution of the Wilcoxon Signed Rank Statistic is used (small case); if sample size ≥ 20, the normal distribution is used (large case)
Alpha	0.05	Value of alpha entered by user
Sample Size	10	Sample size not including values equal to specified median
W	39	W+ - W-
W+	47	Sum of ranks of the positive values
W-	8	Sum of ranks of the negative values
z Value	N/A	Calculated z value for large sample case; N/A for small sample case
p value	0.0488	Calculated p value (will be red if ≤ alpha)
The null hypothesis is rejected.		Conclusion based on the values of p and alpha
There is evidence that the median does not equal 200.		Reject the null if p value ≤ alpha
94.72% Confidence Interval		
Lower	200.1	Calculated confidence interval
Upper	203.55	

Notes:

- Any values equal to the median are not included; this reduces the effective sample size but should not be an issue as long as there are not too many values equal to the median.
- The software uses the null distribution of the Wilcoxon rank signed statistic when the sample size is less than 20 and uses a normal approximation when the sample size is greater than or equal to 20.
- Exact confidence intervals cannot usually be calculated. The confidence interval given is two-sided and based on a normal distribution approximation with a continuity correction.
- The calculated median is the median of the Walsh averages

Two Sample Mann-Whitney Test

Many statistical methods require the assumption of normally distributed data. This assumption often does not hold. In addition, some sample sizes are so small that you can't verify if the distribution is normal. Nonparametric statistical methods do not make too many assumptions about the population from which the sample is drawn. The two sample Mann-Whitney test is one of these nonparametric statistical methods.

The Mann-Whitney test is used when you want to determine if the results from one population or larger than the results from another population. For example, you might want to do if a group of patients treated with a new vaccine did not get ill as often as a group of untreated patients. The example below shows how to do this test using the SPC for MS Excel software. The median is once again used in this test since it is a better measure of the central tendency of the data than the average for non-normal data.

Example[20]

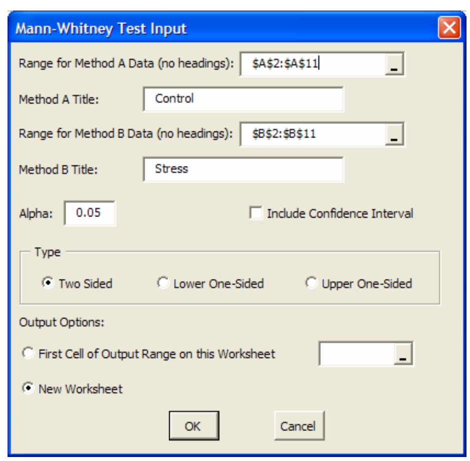
The data below is the failure time for 18 capacitors. Eight had been tested under normal operating conditions; ten had been tested under thermally stress conditions.

Normal: 5.2, 8.5 9.8, 12.3, 17.1, 17.9, 23.7, 29.8
Stress: 1.1, 2.3, 3.2, 6.3, 7, 7.2, 9.1, 15.2, 18.3, 21.1

We want to use these results to see if there is a difference between the two groups.

1. Enter the results into an Excel spreadsheet as shown and highlight the data (not the headings).

Normal	Stress
5.2	1.1
8.5	2.3
9.8	3.2
12.3	6.3
17.1	7
17.9	7.2
23.7	9.1
29.8	15.2
	18.3
	21.1



2. Select the Six Sigma Tools icon from the SPC Menu to show the form "Other Six Sigma Tools."
3. Select the Nonparametric page (the last one).
4. Select the Mann-Whitney Test for Two Samples option.
5. Fill in the One Sample Sign Test form.
    a. *Range for Method A Data (no headings):* this is the range containing the data for the first method; default is the first column range selected on the worksheet
    b. *Method A Title:* name of the first variable; default is the cell above the first column selected on the worksheet
    c. *Range for Method B Data (no headings):* this is the

[20] Example from Statistics and Data Analysis, by Ajit Tamhane and Dorothy Dunlop, Prentice-Hall, 2000.

range containing the data for the first method; default is the second column range selected on the worksheet

d. *Method B Title:* name of the first variable; default is the cell above the second column selected on the worksheet
e. *Alpha:* (1- alpha) is the confidence level; default is 0.05.
f. *Include confidence interval:* check if you want the confidence interval displayed; default is not to display the confidence interval.
g. *Type:* type of test: two sided, lower one-sided, or upper one-sided; default is two-sided.
h. *Output option:* select the first cell of an output range on an existing worksheet and select a cell (a check is made to ensure that there is nothing in the cells that will be written over) or select new worksheet (the default).

6. Select OK once your have completed your entries.
7. The output is shown below with an explanation of the terms to the right. The option for a confidence interval is included.

Two Sample Mann-Whitney Test			Explanation (not included in the output)
H_0: $\theta_1 = \theta_2$			Null hypothesis (medians are equal)
H_1: $\theta_1 <> \theta_2$			Alternate hypothesis (two-sided in this example)
	Control	Stress	Names to two methods
Median	14.7	7.1	Calculated medians of the two methods
Sample Size	8	10	Sample size of the two methods
Median of Differences	6.1		Median of the differences
No. of + Differences	59		Number of + differences for each pair of numbers
No of 0 Differences	0		Number where the pair of numbers are equal
No. of - Differences	21		Number of – differences for each pair of numbers
Alpha	0.05		Value of alpha entered by user
z Value	1.64		Calculated z value
p value	0.1002		Calculated p value (will be red if ≤ alpha)
The null hypothesis is not rejected.			Conclusion based on the values of p and alpha
There is no evidence that θ_1 does not equal θ_2.			Reject if p value ≤ alpha
94.94 Confidence Interval			
Lower	-1.2		Calculated confidence interval
Upper	14.7		

Notes:

- The software uses a normal approximation for this test.
- In the case of ties, 0.5 is added to both the + differences and the – differences.
- Exact confidence intervals cannot usually be calculated. The confidence interval given is two-sided and based on a normal distribution approximation with a continuity correction.

Kruskal-Wallis Test for Multiple Samples

Many statistical methods require the assumption of normally distributed data. This assumption often does not hold. In addition, some sample sizes are so small that you can't verify if the distribution is normal. Nonparametric statistical methods do not make too many assumptions about the population from which the sample is drawn. The Kruskal-Wallis test is one of these nonparametric statistical methods.

The Kruskal-Wallis test determines if there is a difference between the medians of different samples. The median is used in this test since it is a better measure of the central tendency of the data than the average for non-normal data. In addition, pairwise comparisons are given to determine which medians are significantly different. The example below shows how to do this test using the SPC for MS Excel software.

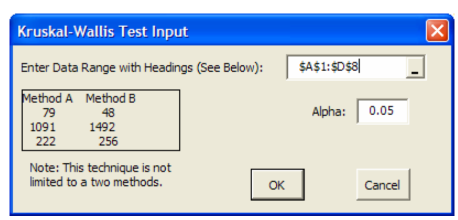
Example[21]

An experiment was done to compare four different methods of teaching the concept of percentages to sixth graders. 28 classes were randomly assigned to the four methods. A 45 item test was given to all classes. The average test scores are given below. We want to use the Kruskal-Wallis test to determine if there are significant differences in the four methods.

1. Enter the results into an Excel spreadsheet as shown below and select the data and headings.

Case Method	Formula Method	Equation Method	Unitary Analysis Method
14.59	20.27	27.82	33.16
23.44	26.84	24.92	26.93
25.43	14.71	28.68	30.43
18.15	22.34	23.32	36.43
20.82	19.49	32.85	37.04
14.06	24.92	33.9	29.76
14.26	20.2	23.42	33.88

2. Select the Six Sigma Tools icon from the SPC Menu to show the form "Other Six Sigma Tools."
3. Select the Nonparametric page (the last one).
4. Select the Kruskal-Wallis Test for Multiple Sample option.
5. Complete the Kruskal-Wallis Test Input form.
    a. *Enter Data Range with Labels:* enter the range containing the data and the column labels; default is the range selected on the worksheet.
    b. *Alpha:* (1- alpha) is the confidence level; default is 0.05.



6. Select OK once your have completed your entries.
7. The output is shown below with an explanation of the terms to the right.

---

[21] Example from Statistics and Data Analysis, by Ajit Tamhane and Dorothy Dunlop, Prentice-Hall, 2000.

Kruskal-Wallis Test with Pairwise Comparison						Explanation (not included in the output)
	Case Method	Formula Method	Equation Method	Unitary Analysis Method		Data is reprinted
	14.59	20.27	27.82	33.16		
	23.44	26.84	24.92	26.93		
	25.43	14.71	28.68	30.43		
	18.15	22.34	23.32	36.43		
	20.82	19.49	32.85	37.04		
	14.06	24.92	33.9	29.76		
	14.26	20.2	23.42	33.88		
Median	**18.15**	**20.27**	**27.82**	**33.16**		Calculated median for each method
Rank Sum	49	66.5	125.5	16 5		Calculated sum of the ranks
Count	7	7	7	7		Number of results in each method
Average Rank	7.000	9.500	17.929	23.571		Calculated average rank
Alpha	0.05					Alpha entered by the user
kw Statistic	18.134					Calculated Kruskal-Wallis statistics
p value	0.0004					Calculated p value (will be red if ≤ alpha)
Conclusion						Conclusion
There is evidence that there are differences in the methods.						If p value ≤ alpha, then there are differences
Comparison			**Pairwise Diff.**	**Critical Value**	**Diff?**	Pairwise comparison using the studentized range distribution
Case Method-Formula Method			2.500	11.299	No	If significant difference, Yes.
Case Method-Equation Method			10.929	11.299	No	
Case Method-Unitary Analysis Method			16.571	11.299	Yes	
Formula Method-Equation Method			8.429	11.299	No	
Formula Method-Unitary Analysis Method			14.071	11.299	Yes	
Equation Method-Unitary Analysis Method			5.643	11.299	No	

Notes:

- Uses the chi-square distribution to determine the p value.

Select Cells

You can save some time with some of the tools in this program by selecting the appropriate cells prior to selecting the icon on the SPC toolbar. For example, if you are making a p control chart, it is beneficial to select the subgroup identifiers before selecting the attribute control chart icon from the SPC toolbar. This can be a little cumbersome if you have lots of subgroups. The Select Cells (SC) on the SPC toolbar helps make this easier.

Selecting One Cell Only

If you have the cursor in one cell only (as shown by the shaded cell in the table to the right) and you select the SC icon on the SPC toolbar, the program will select the cells directly below the shaded cell (including the shaded cell) as shown the table below.

Date	Number of Telemarketing Calls (n)	Number that Result in an Order (np)
2/1/2003	40	5
2/2/2003	63	10
2/3/2003	47	12
2/4/2003	52	7
2/5/2003	34	3
2/6/2003	59	21

Date	Number of Telemarketing Calls (n)	Number that Result in an Order (np)
2/1/2003	40	5
2/2/2003	63	10
2/3/2003	47	12
2/4/2003	52	7
2/5/2003	34	3
2/6/2003	59	21

Selecting More Than One Cell

If you select more than one cell (as shown to the right), all adjacent cells (beginning with the first row and first column) will be selected.

Date	Number of Telemarketing Calls (n)	Number that Result in an Order (np)
2/1/2003	40	5
2/2/2003	63	10
2/3/2003	47	12
2/4/2003	52	7
2/5/2003	34	3
2/6/2003	59	21

Date	Number of Telemarketing Calls (n)	Number that Result in an Order (np)
2/1/2003	40	5
2/2/2003	63	10
2/3/2003	47	12
2/4/2003	52	7
2/5/2003	34	3
2/6/2003	59	21

Transfer Charts to PowerPoint or Word

You can transfer one or more charts to PowerPoint by selecting the PowerPoint icon on the SPC menu or to Word by selecting the Word icon on the SPC menu.

PowerPoint

To transfer to PowerPoint, a presentation will be opened if there is not one opened. The program adds a slide to the presentation and copies the chart over. You can transfer multiple charts at once by selecting multiple worksheets in Excel before running this option.

Word

A Word document will be opened if there is not a Word document open. The chart is placed in the open document wherever the cursor is. You can transfer multiple charts at once by selecting multiple worksheets in Excel before running this option.