Common Tools for

Displaying and Communicating Data for

Process Improvement

Packet includes:

Source: American Society for Quality (ASQ), http://asq.org/learn-about-quality/

Box and Whisker Plot

A box and whisker plot is a graphical method of displaying variation in a set of data. In most cases, a histogram provides a sufficient display; however, a box and whisker plot can provide additional detail while allowing multiple sets of data to be displayed in the same graph. Some types are called box and whisker plots with outliers.

Box and whisker plots are very effective and easy to read. They summarize data from multiple sources and display the results in a single graph. Box and whisker plots allow for comparison of data from different categories for easier, more effective decision-making.

When to Use a Box and Whisker Plot

Use box and whisker plots when you have multiple data sets from independent sources that are related to each other in some way. Examples include test scores between schools or classrooms, data from before and after a process change, similar features on one part such as cam shaft lobes, or data from duplicate machines manufacturing the same products.

Box and Whisker Plot Procedure

A box and whisker plot is developed from five statistics.

1. Minimum value – the smallest value in the data set
2. Second quartile – the value below which the lower 25% of the data are contained
3. Median value – the middle number in a range of numbers
4. Third quartile – the value above which the upper 25% of the data are contained
5. Maximum value – the largest value in the data set

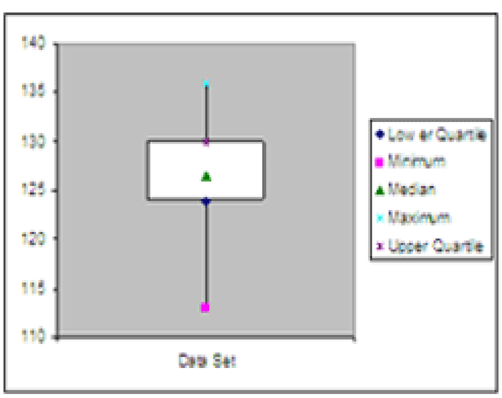
For example, given the following 20 data points, the five required statistics are displayed.

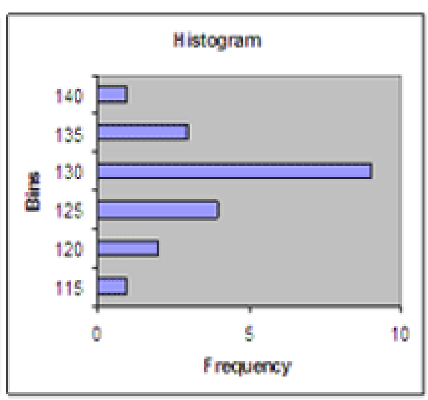
Number	Data	5 Required Statistics
1	113	Minimum: 113
2	116	
3	119	
4	121	
5	124	
	———	2nd Quartile: 124
6	124	
7	125	

8	126	
9	126	
10	126	
		Median: 126.5
11	127	
12	127	
13	128	
14	129	
15	130	
		3rd Quartile: 130
16	130	
17	131	
18	132	
19	133	
20	136	Maximum: 136

Note that for a data set with an even number of values, the median is calculated as the average of the two middle values.

Here are the data represented in a box and whisker plot format.





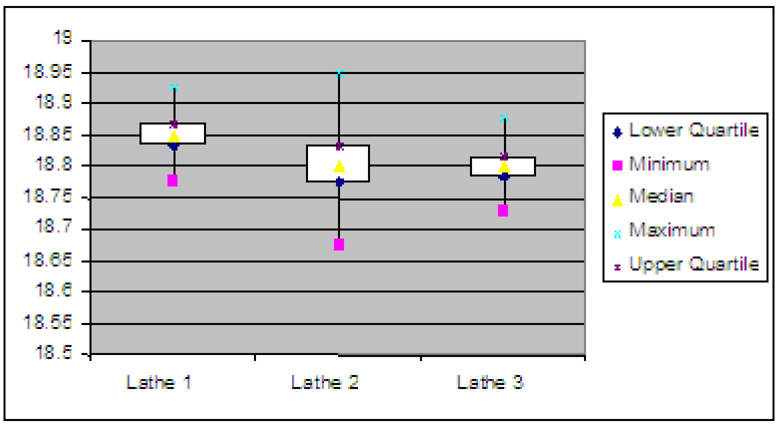
Left: The center represents the middle 50%, or 50th percentile of the data set and is derived using the lower and upper quartile values. The median value is displayed inside the "box." The maximum and minimum values are displayed with vertical lines ("whiskers") connecting the points to the center box.

Right: For comparison, a histogram of the data is also shown, showing the frequency of each value in the data set.

Box and Whisker Plot Example

Suppose you wanted to compare the performance of three lathes responsible for the rough turning of a motor shaft. The design specification is 18.85 +/- 1.0 mm.

Diameter measurements from a sample of shafts taken from each roughing lathe are displayed in a box and whisker plot.



1. Lathe 1 appears to be making good parts, and is centered in the tolerance.
2. Lathe 2 appears to have excess variation, and is making shafts below the minimum diameter.
3. Lathe 3 appears to be performing comparably to Lathe 1. However, it is targeted low in the tolerance, and is making shafts below specification.

Most software packages that perform statistical analysis can create box and whisker plots.

Source

American Society for Quality (ASQ), http://asq.org/learn-about-quality/data-collection-analysis-tools/overview/box-whisker-plot.html

Check Sheet

A check sheet is a structured, prepared form for collecting and analyzing data. This is a generic tool that can be adapted for a wide variety of purposes. It typically is used to collect data on frequency of events, problems, defects, defect location, defect causes, etc.

When to Use a Check Sheet

- When data can be observed and collected repeatedly by the same person or at the same location.
- When collecting data on the frequency or patterns of events, problems, defects, defect location, defect causes, etc.
- When collecting data from a production process.

Check Sheet Procedure

1. Decide what event or problem will be observed. Develop operational definitions.
2. Decide when data will be collected and for how long.
3. Design the form. Set it up so that data can be recorded simply by making check marks or Xs or similar symbols and so that data do not have to be recopied for analysis.
4. Label all spaces on the form.
5. Test the check sheet for a short trial period to be sure it collects the appropriate data and is easy to use.

Check Sheet Example

Defect Types / Event Occurrences	Dates							TOTAL
	Sun	Mon	Tue	Wed	Thu	Fri	Sat	
Defect 1				|				**1**
Defect 2			||||	|	||	卌		**12**
Defect 3	|||			|||				**6**
Defect 4		|			||||		|	**6**
Defect 5								**0**
Defect 6		卌						**5**
Defect 7			||			卌		**7**
Defect 8							||	**2**
TOTAL	**3**	**6**	**6**	**5**	**6**	**10**	**3**	**78**

Source

American Society for Quality (ASQ), http://asq.org/learn-about-quality/data-collection-analysis-tools/overview/check-sheet.html

Control Chart

Also called: statistical process control

The control chart is a graph used to study how a process changes over time. Data are plotted in time order. A control chart always has a central line for the average, an upper line for the upper control limit and a lower line for the lower control limit. These lines are determined from historical data. By comparing current data to these lines, you can draw conclusions about whether the process variation is consistent (in control) or is unpredictable (out of control, affected by special causes of variation).

Control charts for variable data are used in pairs. The top chart monitors the average, or the centering of the distribution of data from the process. The bottom chart monitors the range, or the width of the distribution. If your data were shots in target practice, the average is where the shots are clustering, and the range is how tightly they are clustered. Control charts for attribute data are used singly.

When to Use a Control Chart

- When controlling ongoing processes by finding and correcting problems as they occur.
- When predicting the expected range of outcomes from a process.
- When determining whether a process is stable (in statistical control).
- When analyzing patterns of process variation from special causes (non-routine events) or common causes (built into the process).
- When determining whether your quality improvement project should aim to prevent specific problems or to make fundamental changes to the process.

Control Chart Basic Procedure

1. Choose the appropriate control chart for your data.
2. Determine the appropriate time period for collecting and plotting data.
3. Collect data, construct your chart and analyze the data.
4. Look for “out-of-control signals” on the control chart. When one is identified, mark it on the chart and investigate the cause. Document how you investigated, what you learned, the cause and how it was corrected.

Out-of-control signals

- A single point outside the control limits. In Figure 1, point sixteen is above the UCL (upper control limit).
- Two out of three successive points are on the same side of the centerline and farther than 2 σ from it. In Figure 1, point 4 sends that signal.
- Four out of five successive points are on the same side of the centerline and farther than 1 σ from it. In Figure 1, point 11 sends that signal.
- A run of eight in a row are on the same side of the centerline. Or 10 out of 11, 12 out of 14 or 16 out of 20. In Figure 1, point 21 is eighth in a row above the centerline.
- Obvious consistent or persistent patterns that suggest something unusual about your data and your process.

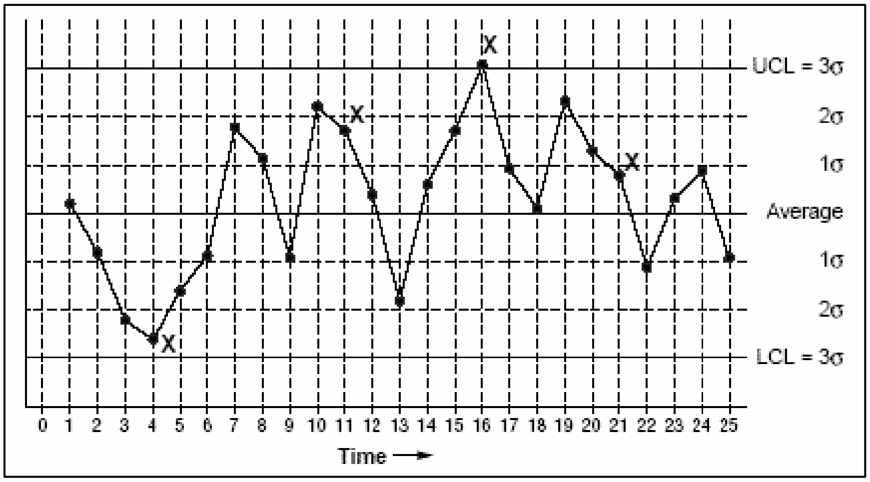


Figure 1 Control Chart: Out-of-Control Signals

5. Continue to plot data as they are generated. As each new data point is plotted, check for new out-of-control signals.
6. When you start a new control chart, the process may be out of control. If so, the control limits calculated from the first 20 points are conditional limits. When you have at least 20 sequential points from a period when the process is operating in control, recalculate control limits.

Source

American Society for Quality (ASQ), http://asq.org/learn-about-quality/data-collection-analysis-tools/overview/control-chart.html

Histogram

A histogram is the most commonly used graph to show frequency distributions. A frequency distribution shows how often each different value in a set of data occurs. It looks very much like a bar chart, but there are important differences between them.

When to Use a Histogram

- When the data are numerical.
- When you want to see the shape of the data's distribution, especially when determining whether the output of a process is distributed approximately normally.
- When analyzing whether a process can meet the customer's requirements.
- When analyzing what the output from a supplier's process looks like.
- When seeing whether a process change has occurred from one time period to another.
- When determining whether the outputs of two or more processes are different.
- When you wish to communicate the distribution of data quickly and easily to others.

Histogram Procedure

1. Collect at least 50 consecutive data points from a process.
2. Use the histogram worksheet to set up the histogram. It will help you determine the number of bars, the range of numbers that go into each bar and the labels for the bar edges. After calculating *W* in step 2 of the worksheet, use your judgment to adjust it to a convenient number. For example, you might decide to round 0.9 to an even 1.0. The value for *W* must not have more decimal places than the numbers you will be graphing.
3. Draw x- and y-axes on graph paper. Mark and label the y-axis for counting data values. Mark and label the x-axis with the *L* values from the worksheet. The spaces between these numbers will be the bars of the histogram. Do not allow for spaces between bars.
4. For each data point, mark off one count above the appropriate bar with an X or by shading that portion of the bar.

Histogram Analysis

- Before drawing any conclusions from your histogram, satisfy yourself that the process was operating normally during the time period being studied. If any unusual events affected the process during the time period of the histogram, your analysis of the histogram shape probably cannot be generalized to all time periods.

Common Histogram Shapes and What They Mean

Normal. A common pattern is the bell–shaped curve known as the "normal distribution." In a normal distribution, points are as likely to occur on one side of the average as on the other. Be aware, however, that other distributions look similar to the normal distribution. Statistical calculations must be used to prove a normal distribution.

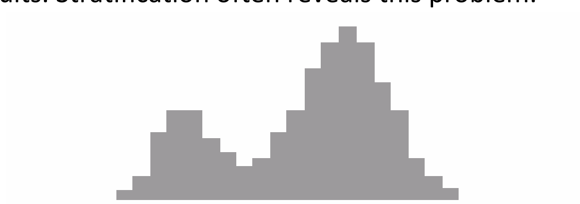
Don't let the name "normal" confuse you. The outputs of many processes—perhaps even a majority of them—do not form normal distributions, but that does not mean anything is wrong with those processes. For example, many processes have a natural limit on one side and will produce skewed distributions. This is normal — meaning typical — for those processes, even if the distribution isn't called normal.

Normal distribution

Skewed. The skewed distribution is asymmetrical because a natural limit prevents outcomes on one side. The distribution's peak is off center toward the limit and a tail stretches away from it. For example, a distribution of analyses of a very pure product would be skewed, because the product cannot be more than 100 percent pure. Other examples of natural limits are holes that cannot be smaller than the diameter of the drill bit or call-handling times that cannot be less than zero. These distributions are called right – or left–skewed according to the direction of the tail.

Right-skewed distribution

Double-peaked or bimodal. The bimodal distribution looks like the back of a two-humped camel. The outcomes of two processes with different distributions are combined in one set of data. For example, a distribution of production data from a two-shift operation might be bimodal, if each shift produces a different distribution of results. Stratification often reveals this problem.

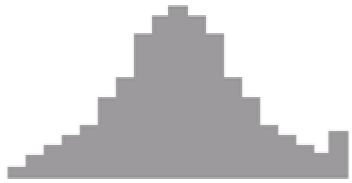


Bimodal (double-peaked) distribution

Plateau. The plateau might be called a "multimodal distribution." Several processes with normal distributions are combined. Because there are many peaks close together, the top of the distribution resembles a plateau.

Plateau distribution

Edge peak. The edge peak distribution looks like the normal distribution except that it has a large peak at one tail. Usually this is caused by faulty construction of the histogram, with data lumped together into a group labeled "greater than…"

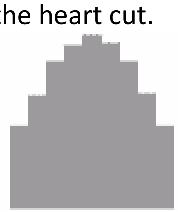


Edge peak distribution

Comb. In a comb distribution, the bars are alternately tall and short. This distribution often results from rounded-off data and/or an incorrectly constructed histogram. For example, temperature data rounded off to the nearest 0.2 degree would show a comb shape if the bar width for the histogram were 0.1 degree.

Comb distribution

Truncated or heart-cut. The truncated distribution looks like a normal distribution with the tails cut off. The supplier might be producing a normal distribution of material and then relying on inspection to separate what is within specification limits from what is out of spec. The resulting shipments to the customer from inside the specifications are the heart cut.



Truncated or heart-cut distribution

Dog food. The dog food distribution is missing something—results near the average. If a customer receives this kind of distribution, someone else is receiving a heart cut, and the customer is left with the "dog food," the odds and ends left over after the master's meal. Even though what the customer receives is within specifications, the product falls into two clusters: one near the upper specification limit and one near the lower specification limit. This variation often causes problems in the customer's process.

Dog food distribution

Source

American Society for Quality (ASQ), http://asq.org/learn-about-quality/data-collection-analysis-tools/overview/histogram.html

Pareto Chart

Also called: Pareto diagram, Pareto analysis

A Pareto chart is a bar graph. The lengths of the bars represent frequency or cost (time or money), and are arranged with longest bars on the left and the shortest to the right. In this way the chart visually depicts which situations are more significant. A Pareto chart helps a team focus on problems that offer the greatest potential for improvement.

The Pareto principle: 20% of sources cause 80% of problems.

When to Use a Pareto Chart

- When analyzing data about the frequency of problems or causes in a process.
- When there are many problems or causes and you want to focus on the most significant.
- When analyzing broad causes by looking at their specific components.
- When communicating with others about your data.

Pareto Chart Procedure

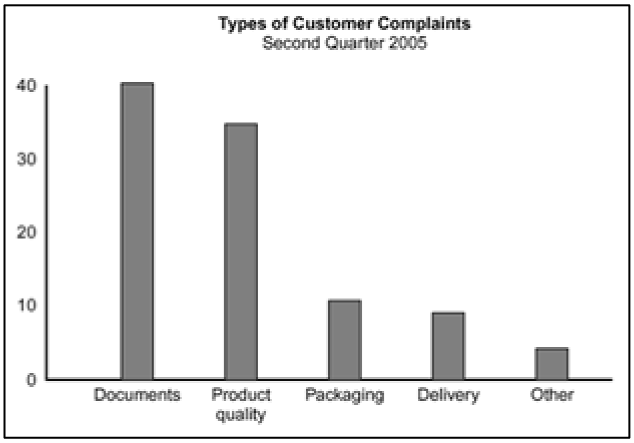
1. Decide what categories you will use to group items.
2. Decide what measurement is appropriate. Common measurements are frequency, quantity, cost and time.
3. Decide what period of time the Pareto chart will cover: One work cycle? One full day? A week?
4. Collect the data, recording the category each time. (Or assemble data that already exist.)
5. Subtotal the measurements for each category.
6. Determine the appropriate scale for the measurements you have collected. The maximum value will be the largest subtotal from step 5. (If you will do optional steps 8 and 9 below, the maximum value will be the sum of all subtotals from step 5.) Mark the scale on the left side of the chart.
7. Construct and label bars for each category. Place the tallest at the far left, then the next tallest to its right and so on. If there are many categories with small measurements, they can be grouped as “other.”

Steps 8 and 9 are optional but are useful for analysis and communication.

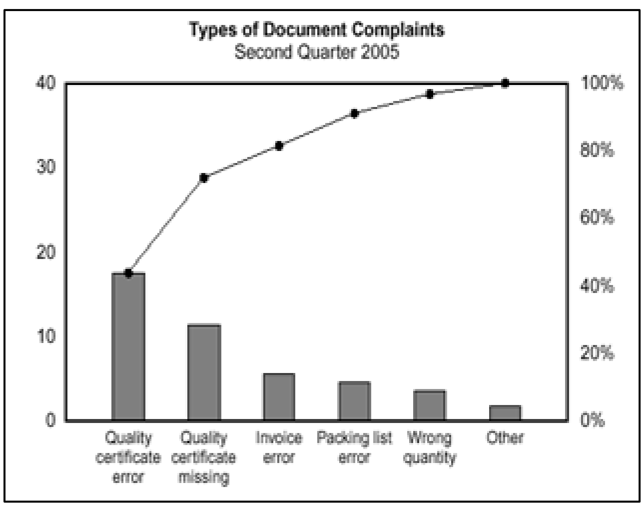
8. Calculate the percentage for each category: the subtotal for that category divided by the total for all categories. Draw a right vertical axis and label it with percentages. Be sure the two scales match: For example, the left measurement that corresponds to one-half should be exactly opposite 50% on the right scale.
9. Calculate and draw cumulative sums: Add the subtotals for the first and second categories, and place a dot above the second bar indicating that sum. To that sum add the subtotal for the third category, and place a dot above the third bar for that new sum. Continue the process for all the bars. Connect the dots, starting at the top of the first bar. The last dot should reach 100 percent on the right scale.

Pareto Chart Examples

Example #1 shows how many customer complaints were received in each of five categories.
Example #2 takes the largest category, "documents," from Example #1, breaks it down into six categories of document-related complaints, and shows cumulative values.
If all complaints cause equal distress to the customer, working on eliminating document-related complaints would have the most impact, and of those, working on quality certificates should be most fruitful.



Example #1



Example #2

Source

American Society for Quality (ASQ), http://asq.org/learn-about-quality/cause-analysis-tools/overview/pareto.html

Run Chart

A run chart is used to study collected data for trends or patterns over a specific period of time.

The run chart is a running record of a process over time:

- The vertical axis represents the process being measured
- The horizontal axis represents the units of time by which the measurements are made
- The centerline of the chart is the mean or average

A **run** is defined as one or more consecutive data points on the same side of the mean line.

When to Use a Run Chart

A run chart will help you:

- Monitor data over time to detect trends, shifts, or cycles
- Compare a measure before and after the implementation of solution to measure impact
- Focus attention on vital changes, not normal variation
- Track useful information for predicting trends

How to Create a Run Chart

1. Choose which data you will measure and track
2. Gather data: Generally, collect 20-25 data points, with which you can detect meaningful patterns over time
3. Create a graph on which you can plot your data (*y* axis, or vertical line) over time (*x* axis, or horizontal line)
4. Plot the data
5. Interpret the chart: Focus on the vital changes or meaningful trends/patterns, rather than each and every data variation; keep reading for interpretation tips

Using a Run Chart to Test for Special Causes

Test #1: The presence of too much or too little variability

Use when there are too few or too many runs.

Test #2: The presence of a shift in the process

A special cause exists if a run contains too many data points (i.e., with 20 or more data points, a run of 8 or more data points is considered "too long"; with less than 20 data points, a run of 7 might also be considered "too long").

Test #3: The presence of a trend

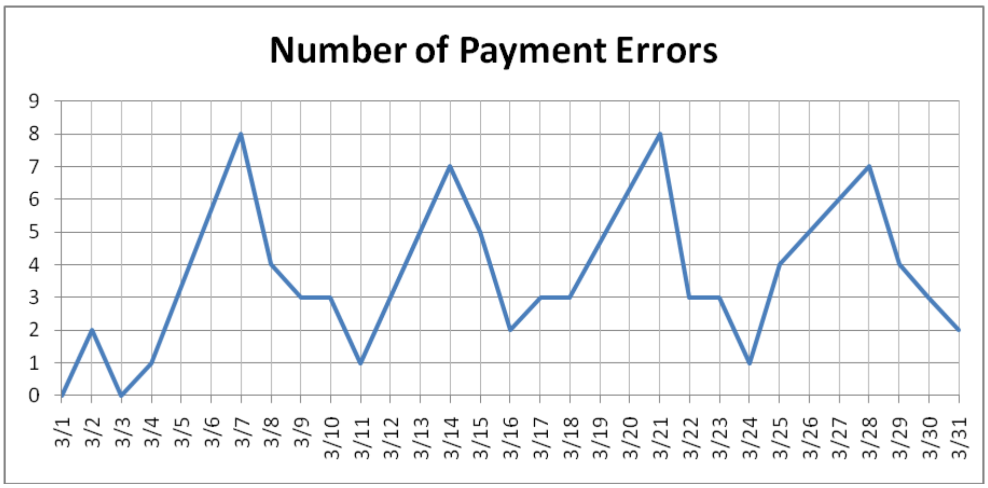
A trend is defined as an unusually long series of consecutive increases or decreases in the data, (usually at least 6 or 7).

Example

A payment processing manager is checking a recently hired associate's error rate for the month. Below is the number of errors that occurred in March.

The run chart indicates a pattern in the number of payment errors. Peaks in errors occurred on 3/7, 3/14, 3/21, and 3/28, all Mondays.

The manager further investigates the payments the new associate processed on Mondays. After identifying the root cause of the problem and initiating corrective action, her plan is to continue monitoring the error rate during the following month.



Source
American Society for Quality (ASQ), http://asq.org/service/body-of-knowledge/tools-run-chart

Scatter Diagram

Also called: scatter plot, X–Y graph

The scatter diagram graphs pairs of numerical data, with one variable on each axis, to look for a relationship between them. If the variables are correlated, the points will fall along a line or curve. The better the correlation, the tighter the points will hug the line.

When to Use a Scatter Diagram

- When you have paired numerical data.
- When your dependent variable may have multiple values for each value of your independent variable.
- When trying to determine whether the two variables are related, such as…
  - When trying to identify potential root causes of problems.
  - After brainstorming causes and effects using a fishbone diagram, to determine objectively whether a particular cause and effect are related.
  - When determining whether two effects that appear to be related both occur with the same cause.
  - When testing for autocorrelation before constructing a control chart.

Scatter Diagram Procedure

1. Collect pairs of data where a relationship is suspected.
2. Draw a graph with the independent variable on the horizontal axis and the dependent variable on the vertical axis. For each pair of data, put a dot or a symbol where the x-axis value intersects the y-axis value. (If two dots fall together, put them side by side, touching, so that you can see both.)
3. Look at the pattern of points to see if a relationship is obvious. If the data clearly form a line or a curve, you may stop. The variables are correlated. You may wish to use regression or correlation analysis now. Otherwise, complete steps 4 through 7.
4. Divide points on the graph into four quadrants. If there are X points on the graph,
   - Count X/2 points from top to bottom and draw a horizontal line.
   - Count X/2 points from left to right and draw a vertical line.
   - If number of points is odd, draw the line through the middle point.
5. Count the points in each quadrant. Do not count points on a line.
6. Add the diagonally opposite quadrants. Find the smaller sum and the total of points in all quadrants.

   A = points in upper left + points in lower right

   B = points in upper right + points in lower left

   Q = the smaller of A and B

   N = A + B
7. Look up the limit for N on the trend test table.
   - If Q is less than the limit, the two variables are related.
   - If Q is greater than or equal to the limit, the pattern could have occurred from random chance.

Table 5.18 Trend test table.

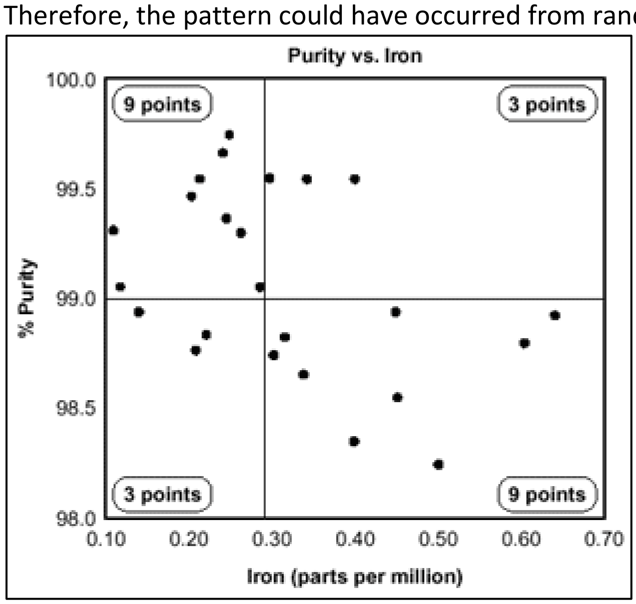
N	Limit	N	Limit
1–8	0	51–53	18
9–11	1	54–55	19
12–14	2	56–57	20
15–16	3	58–60	21
17–19	4	61–62	22
20–22	5	63–64	23
23–24	6	65–66	24
25–27	7	67–69	25
28–29	8	70–71	26
30–32	9	72–73	27
33–34	10	74–76	28
35–36	11	77–78	29
37–39	12	79–80	30
40–41	13	81–82	31
42–43	14	83–85	32
44–46	15	86–87	33
47–48	16	88–89	34
49–50	17	90	35

Scatter Diagram Example

The ZZ-400 manufacturing team suspects a relationship between product purity (percent purity) and the amount of iron (measured in parts per million or ppm). Purity and iron are plotted against each other as a scatter diagram, as shown in the figure below. There are 24 data points. Median lines are drawn so that 12 points fall on each side for both percent purity and ppm iron.

To test for a relationship, they calculate:
A = points in upper left + points in lower right = 9 + 9 = 18
B = points in upper right + points in lower left = 3 + 3 = 6
Q = the smaller of A and B = the smaller of 18 and 6 = 6
N = A + B = 18 + 6 = 24

Then they look up the limit for N on the trend test table. For N = 24, the limit is 6. Q is equal to the limit. Therefore, the pattern could have occurred from random chance, and no relationship is demonstrated.



Scatter Diagram Example

Scatter Diagram Considerations

- Here are some examples of situations in which might you use a scatter diagram:
  - Variable A is the temperature of a reaction after 15 minutes. Variable B measures the color of the product. You suspect higher temperature makes the product darker. Plot temperature and color on a scatter diagram.
  - Variable A is the number of employees trained on new software, and variable B is the number of calls to the computer help line. You suspect that more training reduces the number of calls. Plot number of people trained versus number of calls.
  - To test for autocorrelation of a measurement being monitored on a control chart, plot this pair of variables: Variable A is the measurement at a given time. Variable B is the same measurement, but at the previous time. If the scatter diagram shows correlation, do another diagram where variable B is the measurement two times previously. Keep increasing the separation between the two times until the scatter diagram shows no correlation.
- Even if the scatter diagram shows a relationship, do not assume that one variable caused the other. Both may be influenced by a third variable.
- When the data are plotted, the more the diagram resembles a straight line, the stronger the relationship.
- If a line is not clear, statistics (N and Q) determine whether there is reasonable certainty that a relationship exists. If the statistics say that no relationship exists, the pattern could have occurred by random chance.
- If the scatter diagram shows no relationship between the variables, consider whether the data might be stratified.
- If the diagram shows no relationship, consider whether the independent (x-axis) variable has been varied widely. Sometimes a relationship is not apparent because the data don't cover a wide enough range.
- Think creatively about how to use scatter diagrams to discover a root cause.
- Drawing a scatter diagram is the first step in looking for a relationship between variables.

Source

American Society for Quality (ASQ), http://asq.org/learn-about-quality/cause-analysis-tools/overview/scatter.html