

## Histograms

An important aspect of total quality is the identification and control of all the sources of variation so that processes produce essentially the same result again and again. A histogram is a tool that allows you to understand at a glance the variation that exists in a process.

Although the histogram is essentially a bar chart, it creates a “lumpy distribution curve” that can be used to help identify and eliminate the causes of process variation. Histograms are especially useful in the *measure*, *analyze* and *control* phases of the Lean Six Sigma methodology.



### ***What can it do for you?***

- A histogram will show you the central value of a characteristic produced by your process, and the shape and size of the dispersion on either side of this central value.
- The shape and size of the dispersion will help identify otherwise hidden sources of variation.
- The data used to produce a histogram can ultimately be used to determine the capability of a process to produce output that consistently falls within specification limits.

## How do you do it?

1. Decide which Critical-To-Quality (CTQ) characteristic you wish to examine. This CTQ must be measurable on a linear scale. That is, the incremental value between units of measurement must be the same. For example, a micrometer or a thermometer or a stopwatch can produce linear data. Asking your customers to rate your performance from “poor” to “excellent” on a five-point scale probably will not.
2. Measure the characteristic and record the results. If the characteristic is continually being produced—such as voltage in a line or temperature in an oven, or if there are too many items being produced to measure all of them, you will have to sample. Take care to ensure that your sampling is random.
3. Count the number of individual data points. Add the values for each of the data points and divide by the number of points. This is the mean (or average) value.
4. Determine the highest data value and the lowest data value. Subtract the lower number from the higher. This is the range.
5. The next step is determining how many “classes” or bars your histogram should have.

**To make an initial determination, you can use this table:**

<b>Number of data points</b>	<b>Number of classes</b>
under 50	5 to 7
50 to 100	6 to 10
100 to 250	7 to 12
over 250	10 to 20

6. Divide the range by the trial number of classes you selected. The resulting number will be your trial *class interval* (the horizontal graduation or width) for each bar on your chart. You may round or simplify this number to make it easier to work with, but the total number of classes should be within those shown above.

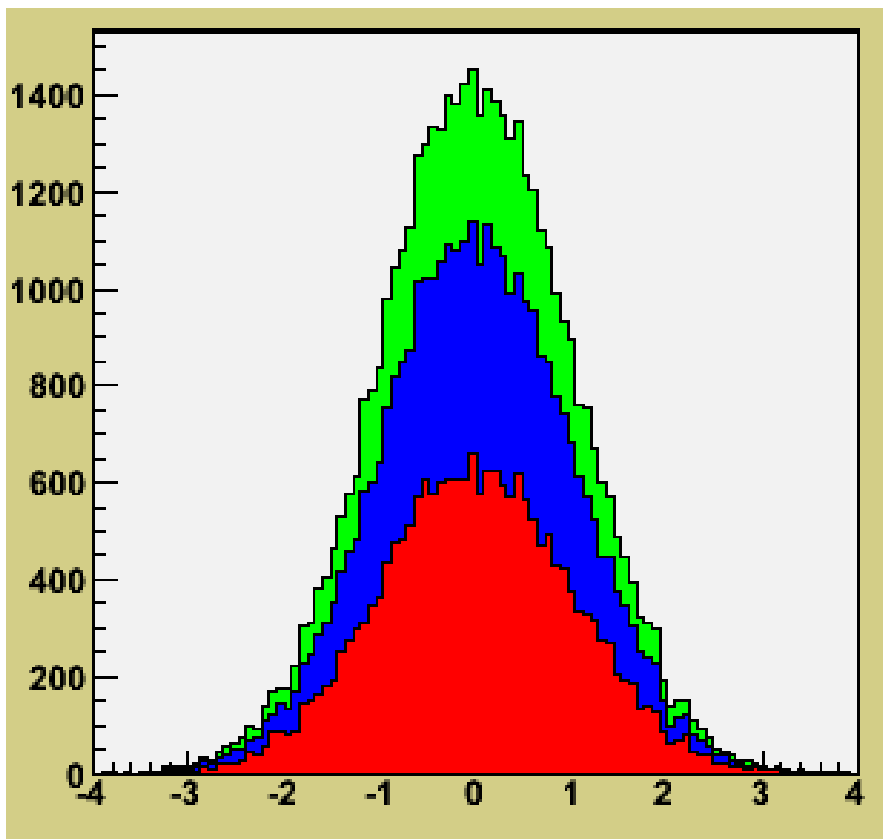
In determining the number of classes and the class interval, consider how you are measuring data. Increase or decrease the number of classes or modify the class interval until there is essentially the same number of measurement possibilities in each class.

7. Determine the class boundaries. You can do this by starting at the center of the range. If you have an odd number of classes, center the middle class approximately at the midpoint of the range, then alternately add or subtract the class interval to define the other class boundaries. If you have an even number of classes, begin the process of adding or subtracting the class interval at approximately the center of the range.

8. Tally the number of data points that fall in each of the classes. Add the frequency totals for each class. This number should equal the total number of data points. Divide the number of data points in each class by the total number of data points. This will give you the percentage of points falling in each class. Add the percentages of all the classes. The result should be approximately 100.
9. Graph the results by beginning with the lowest measurement-value class. Make the bar height correspond to the percentage of data points that fall in that class. Draw the bar for the second class to the right and touching the first bar. Again, make the height correspond to the percentage of data points in that class. Continue in this way until you have drawn in all the classes.
10. Draw a vertical dotted line through your histogram to represent the mean value of all your data points.
11. If there are specification limits for the characteristic you are studying, indicate them as vertical lines as well.
12. Title and label your histogram.

### **Now what?**

The shape that your histogram takes tells a lot about your process. Often, it will tell you to dig deeper for otherwise unseen causes of variation.

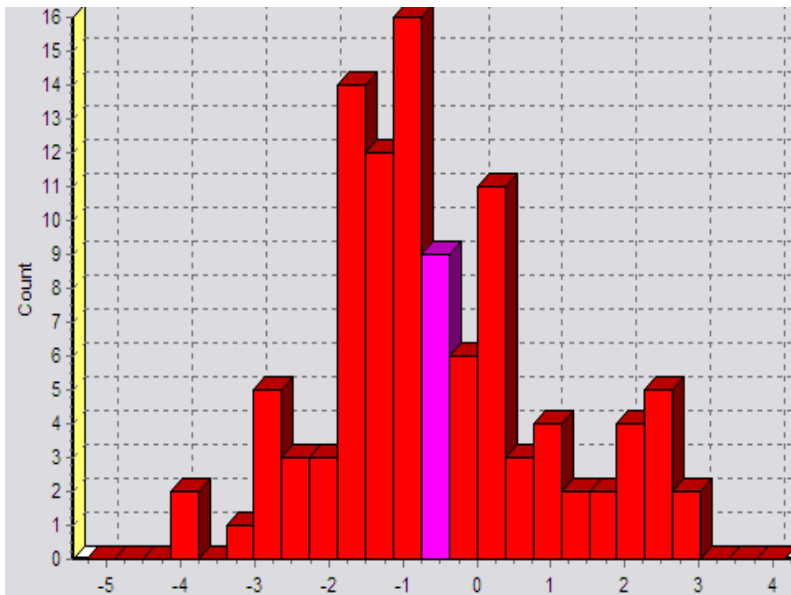


The symmetrical or bell-shaped type of histogram:

The mean value is in the middle of the range of data.

The frequency is high in the middle of the range and falls off fairly evenly to the right and left.

This shape occurs most often.



## The “comb” or multi-modal type of histogram:

Adjacent classes alternate higher and lower in frequency.

This usually indicates a data collection problem. The problem may lie in how a characteristic was measured or how values were rounded.

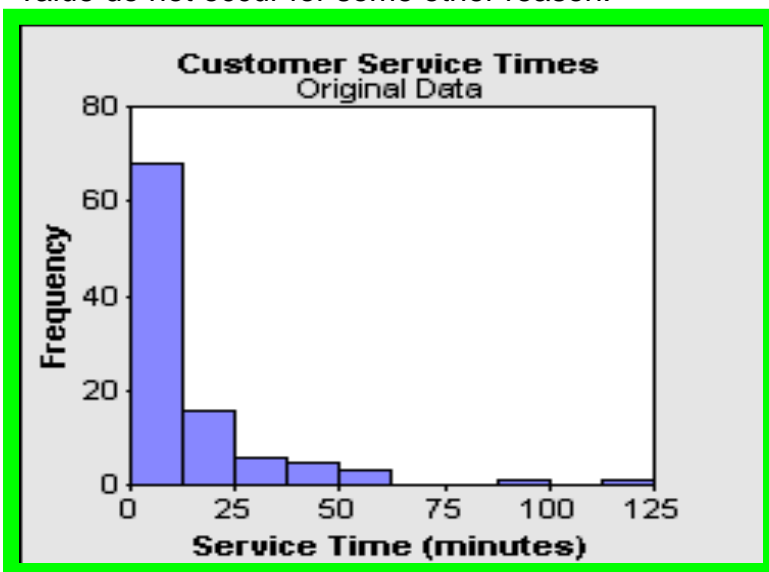
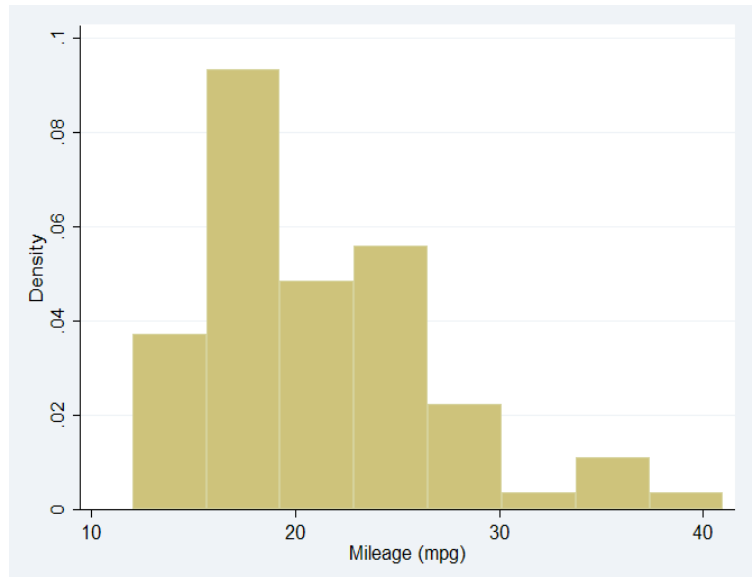
It could also indicate an error in the calculation of class boundaries.

If the distribution of frequencies is shifted noticeably to either side of the center of the range, the distribution is said to be skewed.

### When the histogram is positively skewed

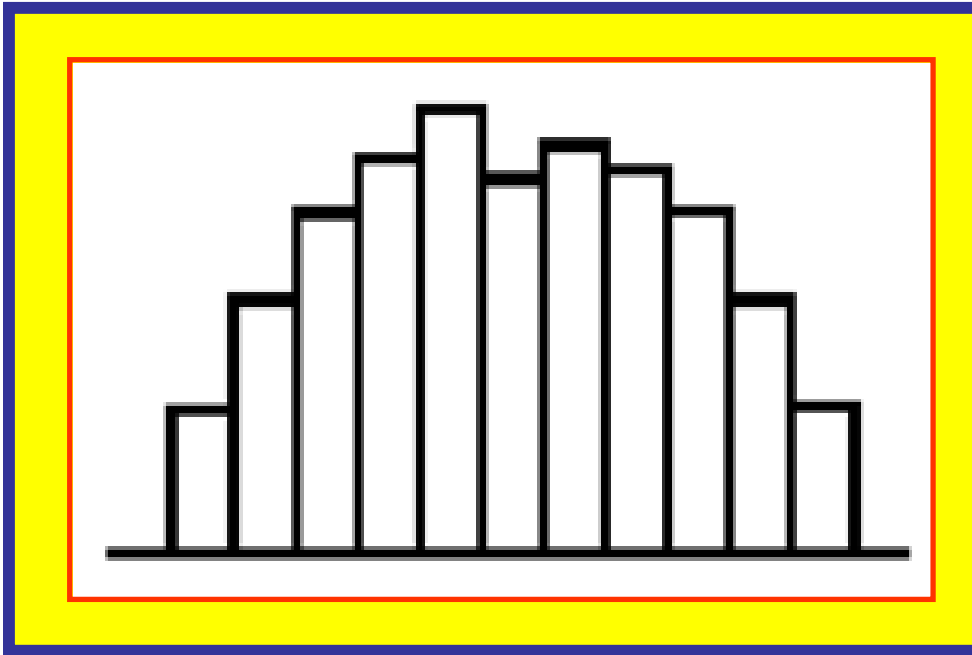
The mean value is to the left of the center of the range, and the frequency decreases abruptly to the left but gently to the right.

This shape normally occurs when the lower limit, the one on the Left, is controlled either by specification or because values lower than a certain value do not occur for some other reason.



If the skewness of the distribution is even more extreme, a clearly asymmetrical, precipice-type histogram is the result.

This shape frequently occurs when a 100% screening is being done for one specification limit.



If the classes in the center of the distribution have more or less the same frequency, the resulting histogram looks like a plateau.

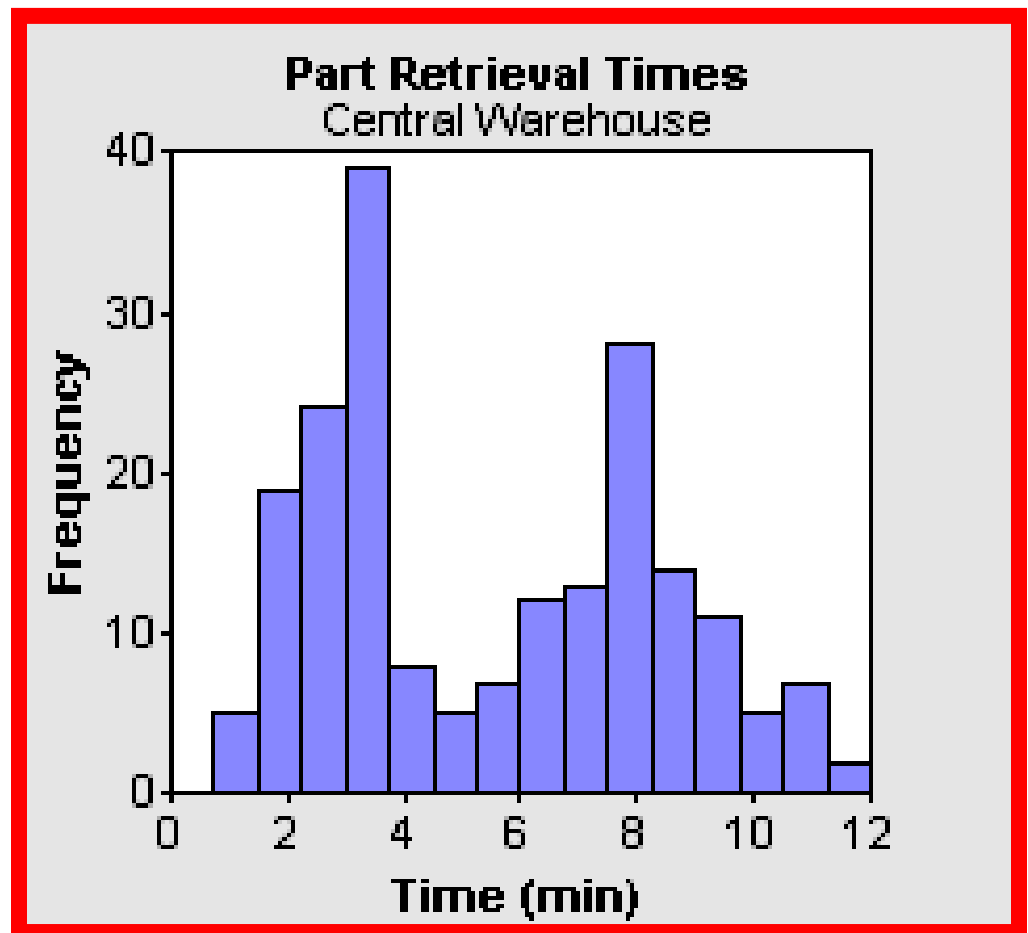
This shape occurs when there is a mixture of two distributions with different mean values blended together.

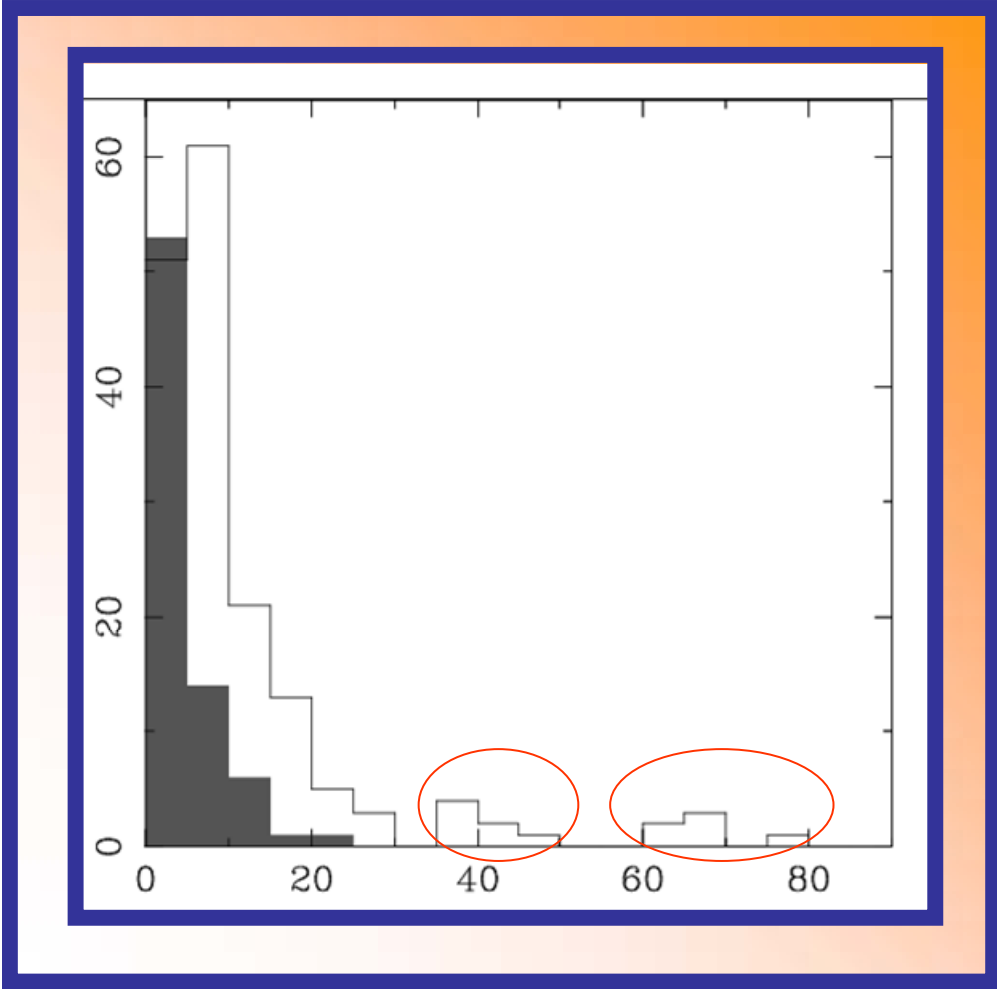
Look for ways to stratify the data to separate the two distributions. You can then produce two separate histograms to more accurately depict what is going on in the process.

If two distributions with widely different means are combined in one data set, the plateau splits to become twin peaks.

The two separate distributions become much more evident than with the plateau.

Examining the data to identify the two different distributions will help you understand how variation is entering the process.



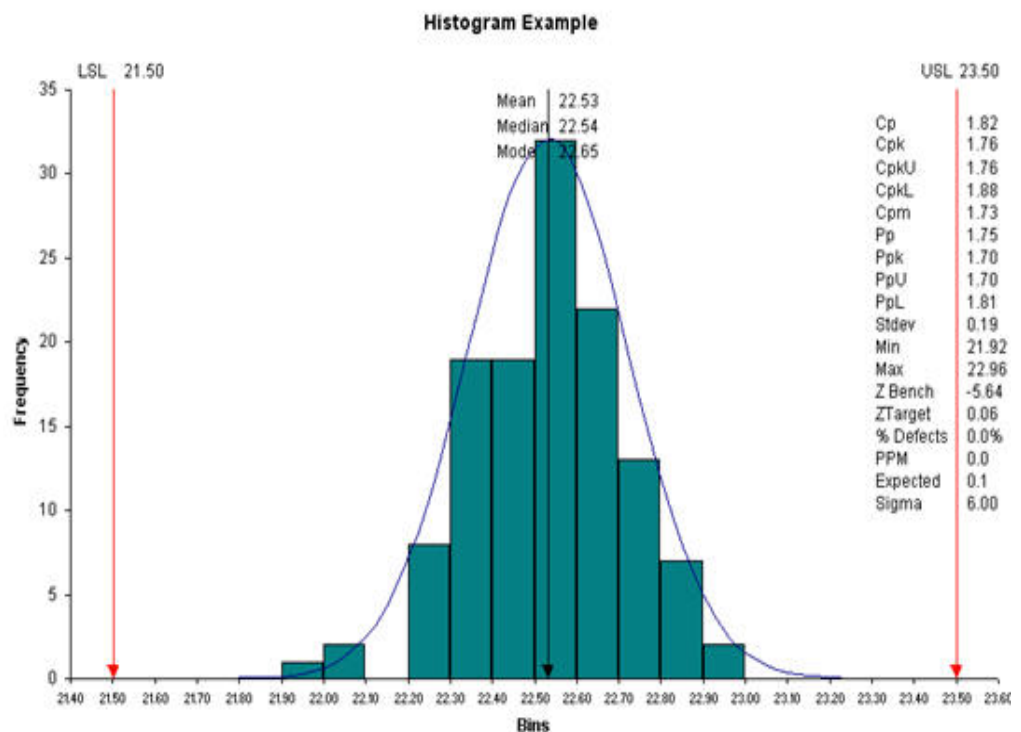


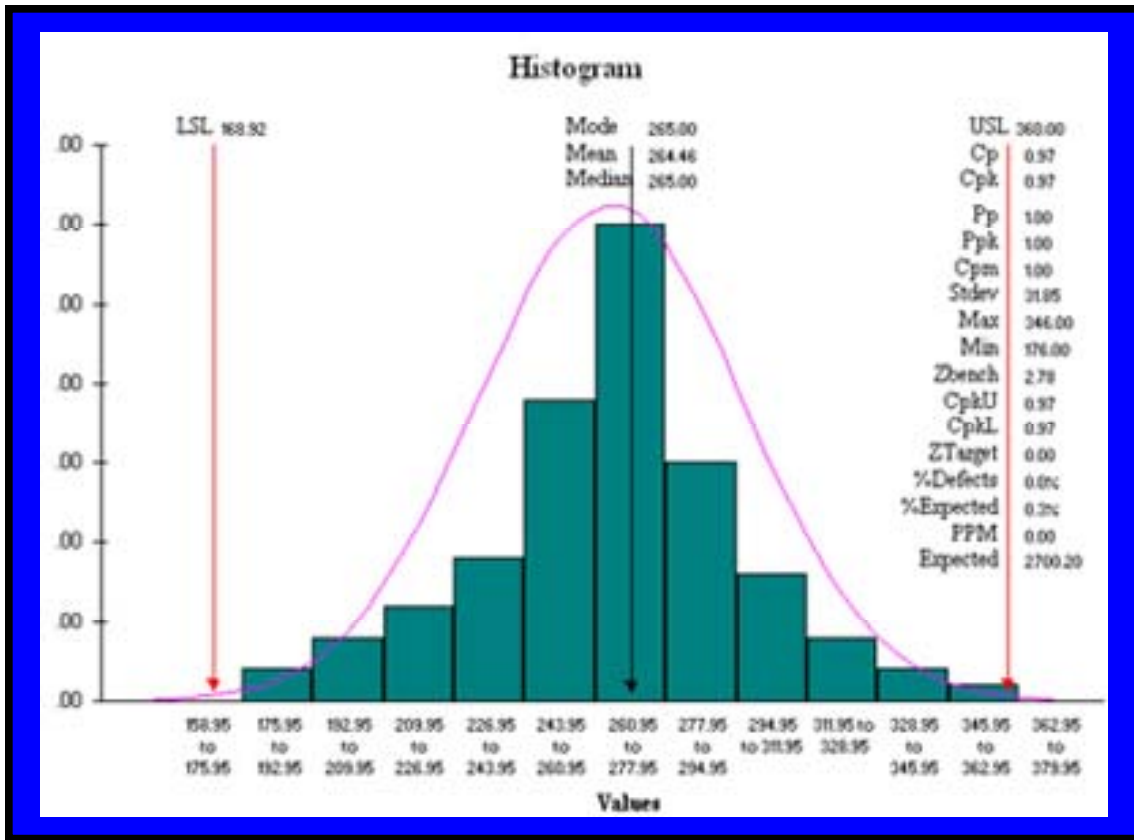
If there is a small, essentially disconnected peak along with a normal, symmetrical peak, this is called an isolated-peak histogram.

It occurs when there is a small amount of data from a different distribution included in the data set. This could also represent a short-term process abnormality, a measurement error or a data collection problem.

If specification limits are involved in your process, the histogram is an especially valuable indicator for corrective action.

The histogram shows that the process is centered between the limits with a good margin on either side. Maintaining the process is all that is needed.

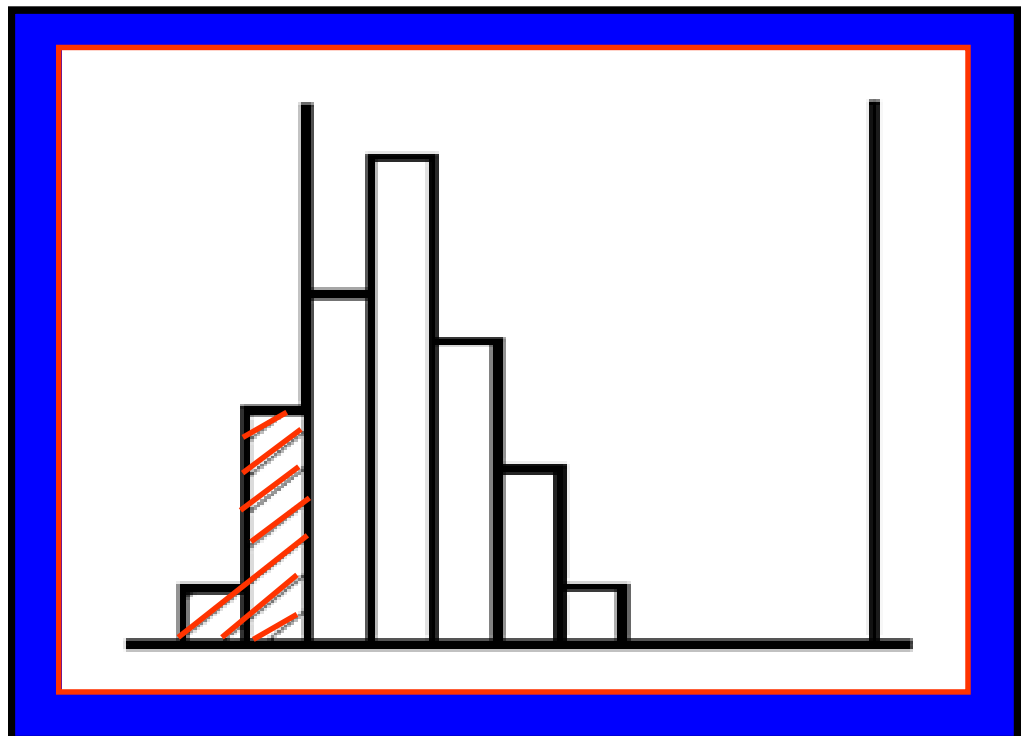


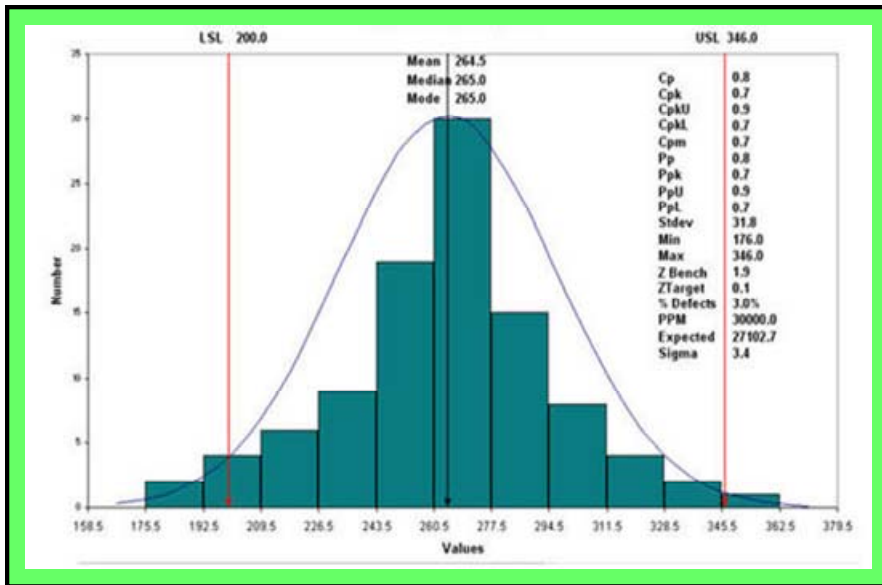


When the process is centered but with no margin, it is a good idea to work at reducing the variation in the process since even a slight shift in the process center will produce defective material.

A process that would have produced material within specification limits if it were centered is shifted to the left.

Action must be taken to bring the mean closer to the center of the specification limits.





A histogram that shows a process that has too much variation to meet specifications no matter how it is centered.

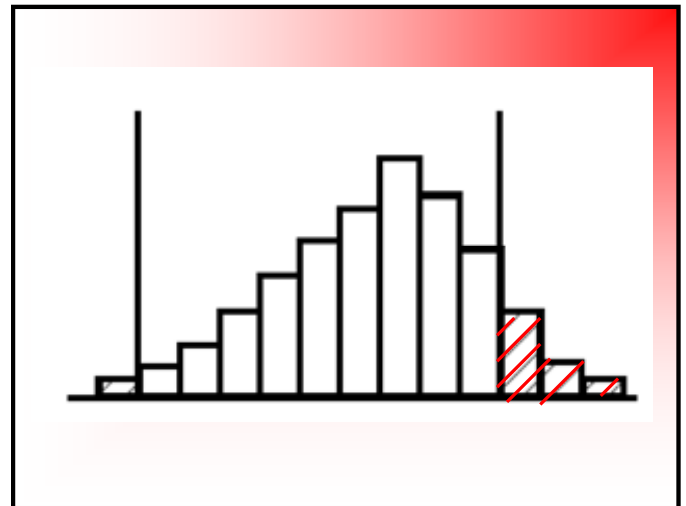
Action must be taken to reduce variation in this process.

A process that is both shifted, in this case to the right, and has too much variation.

Action is necessary to both center the process and reduce variation.

### Conclusion:

A histogram is a picture of the statistical variation in your process. Not only can histograms help you know which processes need improvement, they can also help you track that improvement.



Steven Bonacorsi is a Certified Lean Six Sigma Senior Master Black Belt instructor and coach. Steven Bonacorsi has trained hundreds of Master Black Belts, Black Belts, Green Belts, and Project Sponsors and Executive Leaders in Lean Six Sigma DMAIC and Design for Lean Six Sigma process improvement methodologies. He has led some of the largest deployments in the world and has saved hundreds of millions with his project teams results.

The AIT Group, Inc.  
 Steven Bonacorsi, Vice President, MBA, MS-CIS, PMP  
 Lean Six Sigma Master Black Belt  
 3135 South Price Road, Suite 115  
 Chandler, AZ 85248-3549  
 Phone: +(1) 888.826.2484  
[sbonacorsi@comcast.net](mailto:sbonacorsi@comcast.net)  
<http://www.linkedin.com/in/stevenbonacorsi>  
<http://www.theaitgroup.com>

