Histograms and Process Capability Analysis

by Jay Arthur

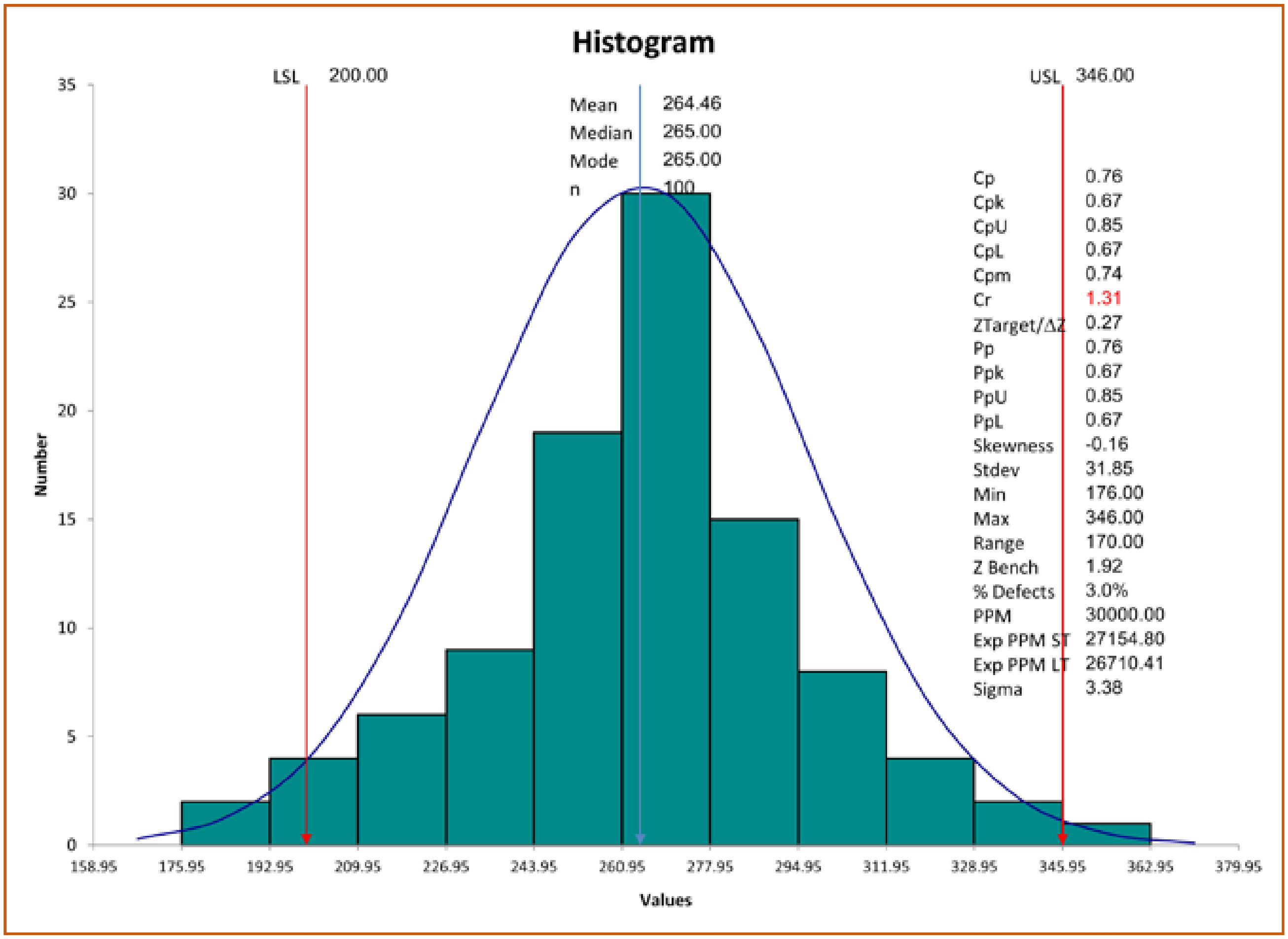
QIMacros®

Histograms and Process Capability Analysis

Table of Contents

Histograms and Process Capability Analysis

by Jay Arthur



Product and Process Variation

A The primary goals of most process improvement methodologies like Six Sigma are to reduce defects, delays and variation. The focus of this paper is on reducing variation.

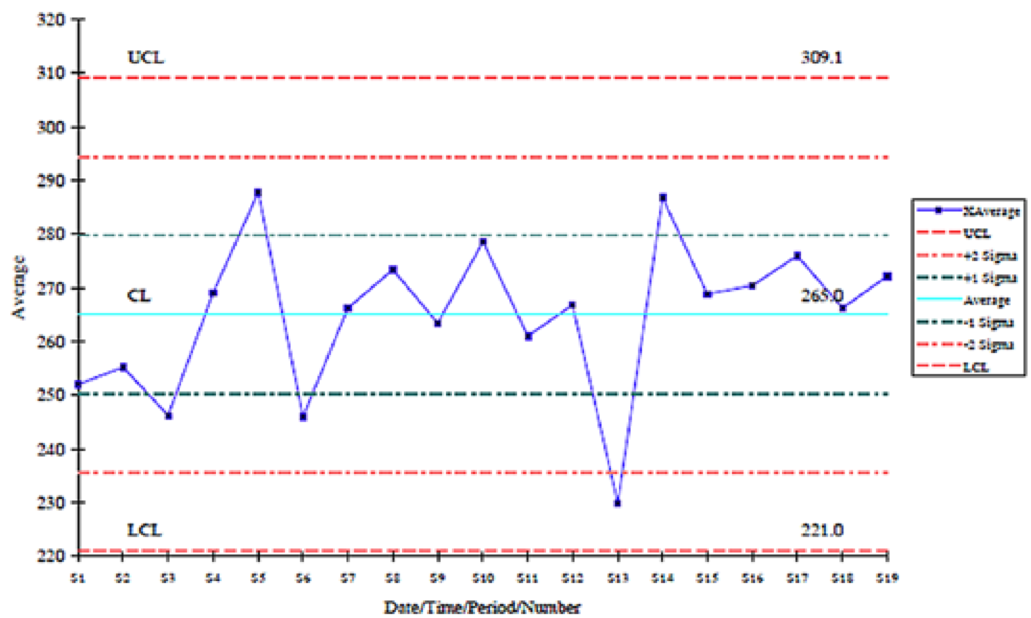
What is Variation?

Every process varies: it takes a little more or less time; a product is a little bit bigger or smaller, longer or shorter, wider or thinner, heavier or lighter, or fuller or emptier than its ideal target size, shape or whatever. The variation may be large or almost undetectable, but it's still there. The goal of Six Sigma is to reduce the amount of variation such that your product always fits well inside your customer's expectations (specifications) and hopefully centers on a *target* value for that product.

Manufacturers get into trouble when they produce products that don't fit the customer's requirements. Services get into trouble when they can't meet the customer's requirement for timeliness.

Process Capability Analysis Requires Stability

Before a process can be evaluated for capability, it must be stable. Use a control chart to determine process stability. If there are no unstable conditions (the QI Macros show unstable conditions in red) you can proceed to capability analysis.



Specification Limits

The goal for all problems associated with variation is to *center the distribution over the ideal target value and minimize the amount of variation around that target value*. For most products, customers have a *target value* and some tolerance for parts around the target value. Your ability to produce products centered around the target value with a minimum amount of variation will determine the quality of your product.

For parts to fit together properly the bolt cannot be bigger or smaller than the nut it screws into; the cap cannot be bigger or smaller than its bottle. In many ways, this is like the goal posts in an American football game: there's a left and a right post and the kicker's job is to kick the ball between the two posts. Anything outside of the posts results in no score (or in Six Sigma terms: a non-conforming part). The left and right post might be considered to be the game's *specification limits*.

Customers specify their requirements for these targets and tolerances in one of two ways:

- Target and a *bilateral* (i.e., two-sided) tolerance (e.g., 74 + 0.05)
- Target and a *unilateral* (i.e., one-sided) tolerance (e.g., 74 + 0.05)

These are used to determine the specification limits (the goal posts).

- Upper Specification Limit (USL) = 74 +0 .05 = 74.05
- Lower Specification Limit (LSL) = 74 - 0.05 = 73.95

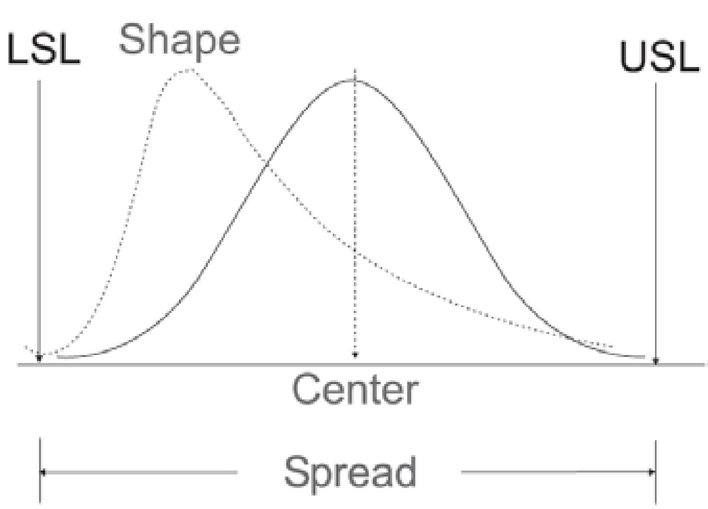
Specification limits apply to services too. I may expect to wait 3 minutes in line at the grocery store (the target.). I would be pleasantly surprised if I didn't have to wait in line at all (LSL=0 or no LSL). I would leave the store without buying anything if a line looked longer than 10 minutes (USL= 10).

Tip: Don't confuse *specification limits* (i.e., USL and LSL) with *control limits* (UCL and LCL). Customers set specification limits; control charts use *your data* to calculate control limits.

Understanding Distributions

Center, Spread and Shape

Before you can start to analyze capability, you'll want to understand the basic concept of distributions (e.g., the bell-shaped curve).



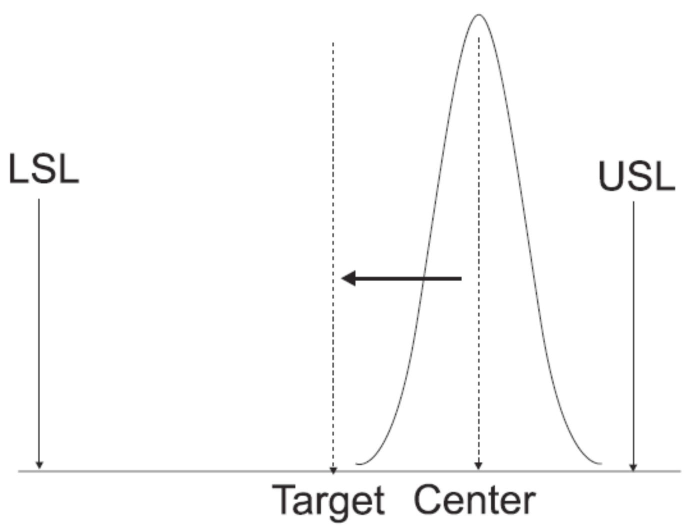
Distributions have three key characteristics: center, spread, and shape:

The *center* is usually the *average* (a.k.a. the *mean*) of all of the data points although other measures of the center can be used (e.g., *median*—center point or *mode*—most frequent data value).

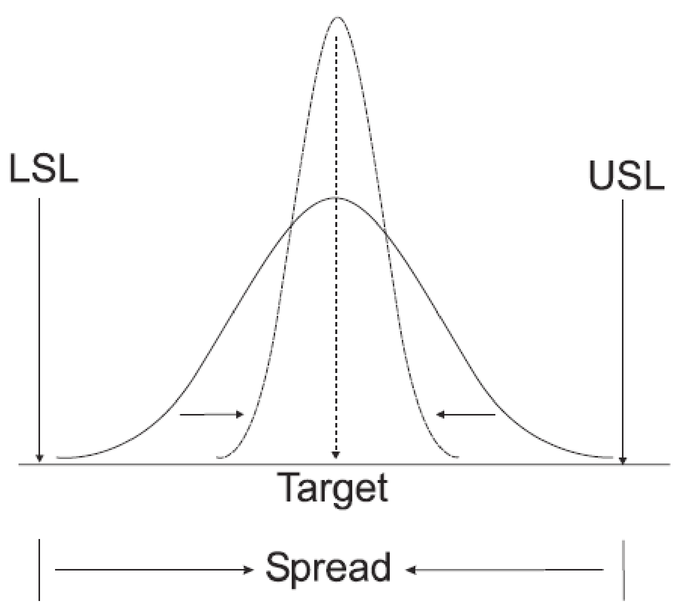
Spread is the distance between the minimum and the maximum values, and the *shape* can be bell-shaped, skewed (i.e., leaning) left or right, and so on.

There Are Two Outcomes for Any Improvement Effort

1. *Center* the distribution over the target value:



2. Reduce the *spread* of the distribution:



These two outcomes can be easily monitored using histograms which help you determine the capability of your process.

Histograms

Histograms are simply bar charts that show the *distribution* of your data using the number of times your data points fall into each of the bars on the histogram. When you add the upper and lower specification limits, it's easy to see how your data fits your customer's requirements and what improvements might be necessary..

Capability Indicies

Using the specification limits, there are four key indicators of process capability:

1. **Cp** is the capability index. It measures how well your data might fit between the upper and lower specification limits. It doesn't really care if the process is centered within the limits, only if the data *would* fit if the data was centered.

2. **Cpk** is the centering capability index. It measures how well your data is centered between the upper and lower specification limits.

   Cp and Cpk use an estimation of the standard deviation know as *sigma estimator* to calculate the spread of your data. If the variation (i.e., range or standard deviation) between samples is small, Cp and Cpk often provide better predictors of capability.

3. **Pp** is the performance index. Like Cp, it measures how well your data fits within the USL and LSL. Unlike Cp, Pp uses the actual standard deviation of your data, not the estimate.

4. **Ppk** is the performance centering index. Like Cpk, it measures how well your data is centered between the USL and LSL. Unlike Cpk, Ppk uses the standard deviation to determine the spread of your data.

Defects Per Million

Because we're using a small sample to analyze process capability, it might seem difficult to calculate the estimated defects, but statistics makes it easy. Since we know the standard deviation and the specification limits, through the magic of statistics we can calculate the actual defect rate in parts per million (PPM) and using the standard deviation, estimate the defect rate in parts per million for the entire population.

Interpreting the Results to Improve Your Process

Once you have run a histogram to calculate Cp and Cpk, you can decide how to improve. If the process is off-center, adjust your work so that it becomes centered. If the capability is less than 1.33, adjust your process so that there is less variation. In manufacturing, customers require Cp=Cpk greater than 1.33 (4-Sigma). If you are producing products for the Asian market, especially Japan, they require Cp=Cpk greater than 1.66 (5-Sigma).

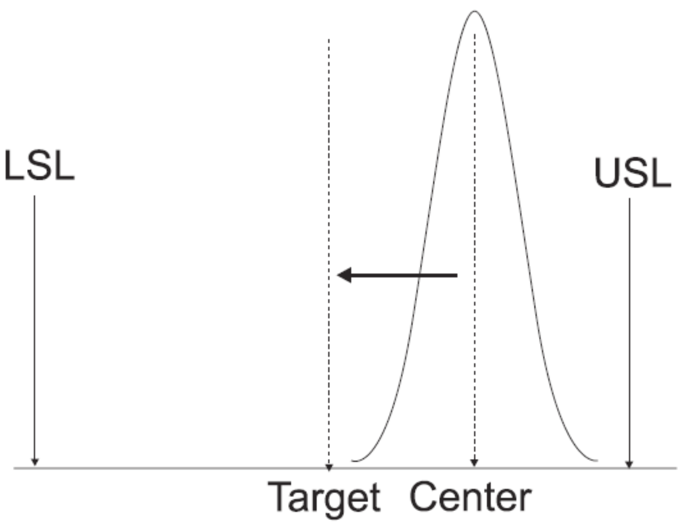
Sigma Level

From a Six Sigma perspective, Cp and Cpk directly correlate with Six Sigma targets:

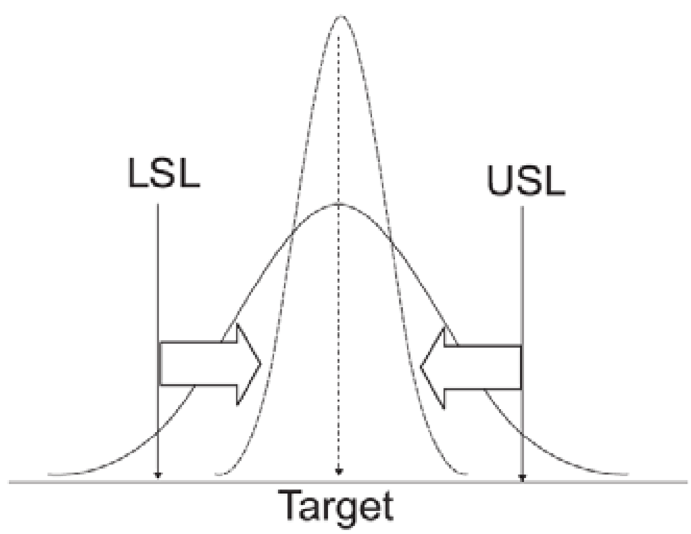
Cp and Cpk Pp and Ppk	Sigma Level
1.0	3
1.33	4
1.66	5
2.0	6

Corrective Action

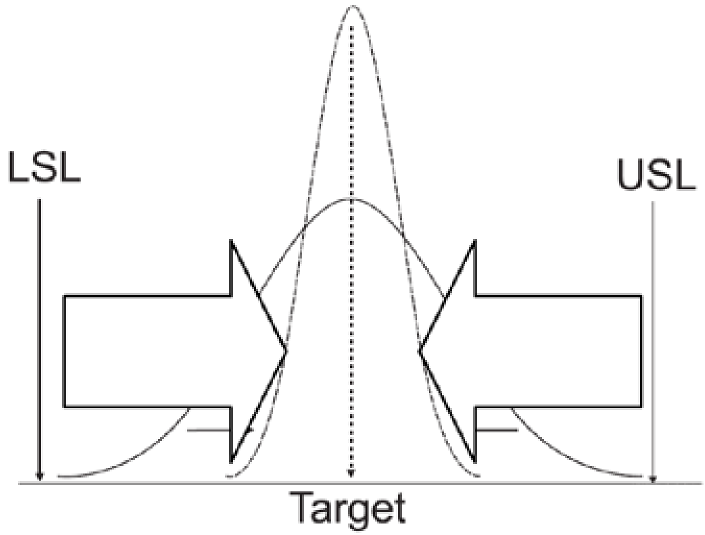
If Cp is greater than Cpk (data fits between the spec limits but is not centered) the corrective action is to center the process:



If Cp is less than Cpk (data is centered but does not fit between the spec limits.) the corrective action is to improve the process to eliminate non-conforming parts:



If Cp = Cpk and both are greater than 1.0 but less than the target value of 1.33 (4-sigma) the corrective action is to improve the process to reduce variation:



Formulas

Formulas for Cp and Cpk

$$Cp = \frac{(USL - LSL)}{6\hat{\sigma}}$$

$$CpU = \frac{(USL - \overline{X})}{3\hat{\sigma}}$$

$$CpL = \frac{(\overline{X} - LSL)}{3\hat{\sigma}}$$

$$Cpk = Min(CpU, CpL)$$

Formulas for Pp and Ppk

For Process Performance, use standard deviation (σ) of the population instead of sigma estimator:

$$Pp = \frac{(USL - LSL)}{6\sigma}$$

$$PpU = \frac{(USL - \overline{X})}{3\sigma}$$

$$PpL = \frac{(\overline{X} - LSL)}{3\sigma}$$

$$Ppk = Min(PpU, PpL)$$

Another View of the Formulas

Formulas for Cp and Pp

Cp	Pp
$\frac{\text{USL-LSL}}{(6 * \text{sigma estimator})}$	$\frac{\text{USL-LSL}}{(6 * \text{standard deviation})}$
Use when you have a sample	Use when you have the total population

Formulas for Cpk and Ppk

	Cpk		Ppk
	Minimum of CpU and CpL		Minimum of PpU and PpL
CpU	$\frac{\text{USL-Xbar}}{(3 * \text{sigma estimator})}$	PpU	$\frac{\text{USL-Xbar}}{(3 * \text{standard deviation})}$
CpL	$\frac{\text{Xbar - LSL}}{(3 * \text{sigma estimator})}$	PpL	$\frac{\text{Xbar - LSL}}{(3 * \text{standard deviation})}$
	Use when you have a sample		Use when you have the total population

Points to note:

- Xbar = the average of the data points = ΣX /n.
- Changing the spec limits will change Cp and Pp and may change Cpk and Ppk.
- Cp and Cpk use sigma estimator because they assume your data represents a sample of the population.
- Pp and Ppk use standard deviation because they assume your data represents the total population.

Formula for Sigma Estimator

Standard deviation of a population can be estimated from the average range or average standard deviation of the samples in each subgroup. These are used to calculate Cp and Cpk.

$$\hat{\sigma} = \frac{\overline{R}}{d_2}$$

For n=1-4, use R/d2 formula; for n>4 use s/c4 formula

$$\hat{\sigma} = \frac{\overline{s}}{c_4}$$

Here is another way of looking at it:

Subgroup Size	Sigma Estimator Formula	Definitions
1 to 4	$\left(\frac{Rbar}{d_2}\right)$	**Rbar** = Average of the ranges **d_2** is a constant based on the sample size
5 or more	$\left(\frac{Sbar}{c_4}\right)$	**Sbar** = Average of the standard deviations **c_4** is a constant based on the sample size

Constants for Sigma Estimator Calculation

(Source:ASTM Manual on Presentation of Data and Control Chart Analysis – Table 16, 2002)

Subgroup Size	Constant	Value	Subgroup Size	Constant	Value
1	d_2	1.128	**26**	c_4	0.9901
2	d_2	1.128	**27**	c_4	0.9905
3	d_2	1.693	**28**	c_4	0.9908
4	d_2	2.059	**29**	c_4	0.9912
5	c_4	0.94	**30**	c_4	0.9915
6	c_4	0.9515	**31**	c_4	0.9917
7	c_4	0.9594	**32**	c_4	0.992
8	c_4	0.965	**33**	c_4	0.9922
9	c_4	0.9693	**34**	c_4	0.9925
10	c_4	0.9727	**35**	c_4	0.9927
11	c_4	0.9754	**36**	c_4	0.9929
12	c_4	0.9776	**37**	c_4	0.9931
13	c_4	0.9794	**38**	c_4	0.9933
14	c_4	0.981	**39**	c_4	0.9935
15	c_4	0.9823	**40**	c_4	0.9936
16	c_4	0.9835	**41**	c_4	0.9938
17	c_4	0.9845	**42**	c_4	0.9939
18	c_4	0.9854	**43**	c_4	0.9941
19	c_4	0.9862	**44**	c_4	0.9942
20	c_4	0.9869	**45**	c_4	0.9944
21	c_4	0.9876	**46**	c_4	0.9945
22	c_4	0.9882	**47**	c_4	0.9946
23	c_4	0.9887	**48**	c_4	0.9947
24	c_4	0.9892	**49**	c_4	0.9948
25	c_4	0.9896	**50**	c_4	0.9949

Formulas for One Sided Spec Limits

LSL Only	USL Only
Cp = Cpk = CpL Pp = Ppk = PpL	Cp = Cpk = CpU Pp = Ppk = PpU

Formula for Defects in Parts Per Million

Actual	Estimated for Population
(# of non conforming)*1000000 / (# of parts)	PPMU = NORMSDIST(Z upper)*1000000 + PPML = NORMSDIST(Z lower)*1000000

Formulas for Z Scores

Z scores help estimate the non-conforming PPM. Z scores standardize +/-3* sigma estimator values into +/-3.

Zlower	(LSL - Xbar)/sigest
Zupper	(USL-Xbar)/sigest
Zbench is the Z score for the Expected PPM	normsinv(1-(Expected PPM/1,000,000))
ZT (target) = Cpk for a target value instead of the USL or LSL. If not defined, use the midpoint between the USL and LSL	(Xbar-Target)/(3*sigest)

Sample Calculation

Let's perform calculations using the following sample data from Montgomery, Intro to SPC, 4th Ed., pgs. 353-358. You can download this data as part of the QI Macros test data at: **http://www.qimacros.com/testdata/SPCManufacturing.xls.** Open the spreadsheet and click on the histogram tab.

Sample	Obs 1	Obs 2	Obs 3	Obs 4	Obs 5
S1	265	205	263	307	220
S2	268	260	234	299	215
S3	197	286	274	243	231
S4	267	281	265	214	318
S5	346	317	242	258	276
S6	300	208	187	264	271
S7	280	242	260	321	228
S8	250	299	258	267	293
S9	265	254	281	294	223
S10	260	308	265	283	277
S11	200	235	246	328	296
S12	276	264	269	235	290
S13	221	176	258	263	231
S14	334	280	265	272	283
S15	265	262	271	245	301
S16	280	274	253	287	258
S17	261	248	260	274	337
S18	250	278	254	274	275
S19	278	250	265	270	298
S20	257	210	280	269	251

- Assume the USL = 346 and the LSL = 200.
- Since there are 5 subgroups, sigma estimator will use the formula SBar/c4

If we look in the Constants for Sigma Estimator Calculation table on page 8, the constant for a subgroup of 5 is 0.94

- Other calculations for this data set are:
  - Xbar = 26,446/100=264.46
  - Standard deviation = 31.85
  - Sigma estimator = (SBar/c4) = (30.02/.94) = 31.93
- If you are trying to recalculate this manually, use the statistical functions in Excel to calculate: standard deviation, normdist and normsinv.

	Formula	Calc	Calc	Value
Cp	(USL-LSL) / 6*(SBar/c_4)	(346-200) / 6*(30.02/.94)	146 / 191.62	.762
Pp	(USL-LSL) / (6 *standard dev)	(346-200) / (6 * 31.85)	146 / 191.10	.764
CpU	(USL-Xbar) / 3 *(SBar/c_4)	(346-264.46) / (3 * (30.02/.94)	81.54 / 95.81	.851
CpL	(Xbar - LSL) / 3 * (SBar/c_4)	(264.46-200) / (3 * (30.02/.94)	64.46 / 95.81	.673
Cpk	Minimum of CpU and CpL	.851 vs .673		.673
PpU	(USL-Xbar) / (3 * standard deviation)	(346-264.46) / 3 * 31.85	81.54 / 95.55	.853
PpL	(Xbar - LSL-) / (3 * standard deviation)	(264.46-200) / 3 * 31.85	64.46 / 95.55	.675
Ppk	Minimum of PpU and PpL	.853 vs .675		.675
Actual PPM	(# of non conforming parts / (# of parts)	3 / 100	30,000 / 1,000,000	30,000
Est PPM	PPM Upper + PPM lower = NORMDIST(Zu) * 1,000,000 + NORMDIST(Zl) * 1,000,000	NORMDIST (2.553) * 1000000 + NORMDIST (2.019) * 1000000	21,768.4 + 5334.3	27,102.7

Z upper (Zu)	(USL-Ave) / (SBar/c_4)	346 - 264.46 / 31.93	2.553	
Z lower (Zl)	(LSL -Ave) / (SBar/c_4)	200 - 264.46 / 31.93	-2.019	
Z bench	normsinv(1-(Expected PPM/1,000,000))	normsinv(1-(.0267104)	normsinv(.97329)	1.93
Target	Defined by Customer or Midpoint between USL and LSL	(346 +200) / 2	546 / 2	273
Z target	ABS(Xbar-Target) / (3*sigest)	264.46 - 273 / 3*(30.02/.94)	8.54 / 95.81	.09

The final calculated amounts are:

Cp	.762
Cpk	.673
Pp	.764
Ppk	.675
PPM	30,000
Est PPM	27,102.7
Zbench	1.93
Z target	.09

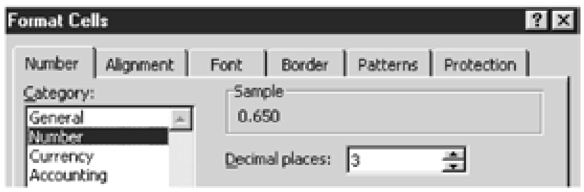
Since the standard deviation and sigma estimator were fairly close in value, the Cp and Pp and Cpk and Ppk values were very similar.

Running a Histogram With QI Macros

If you don't have the QI Macros software you can download a free 30 day trial copy at **https://www.qimacros.com/trial/30-day**

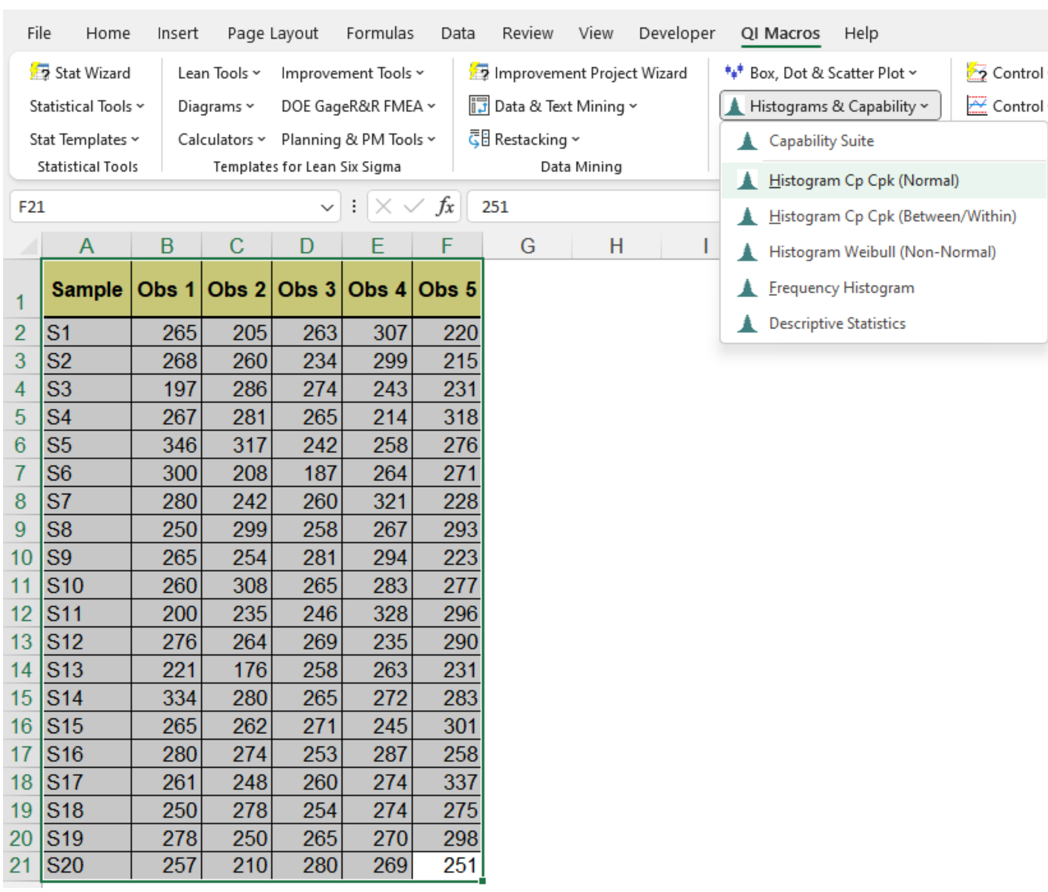
If you have the QI Macros, you can access the test data used in this example on your computer at c:/program files/qimacros/testdata/histogram.xls. This data is the same data used in the previous formula examples so that you can compare the results.

1. The number of decimals in your raw data will drive the number of decimals shown in the results. Make sure your data is formatted to the precision you want. From Format-Cells, selec Number and specify the number of decimal places you want.



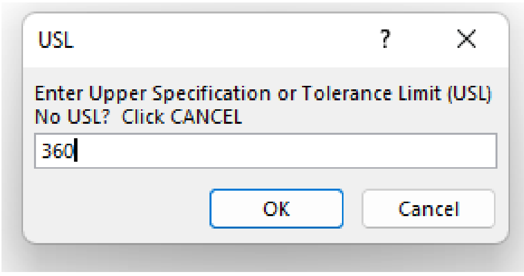
2. Highlight the data to be graphed (minimum of 20 data points recommended). Click on the top data cell and drag the mouse down to include just the data cells, then select either the histogram or frequency histogram from the QI Macros menu.

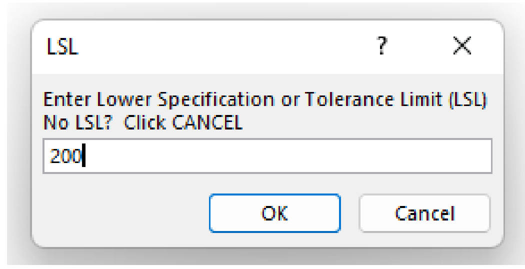
   Tip: Do NOT sort your data before running a histogram or the calculations will be incorrect.



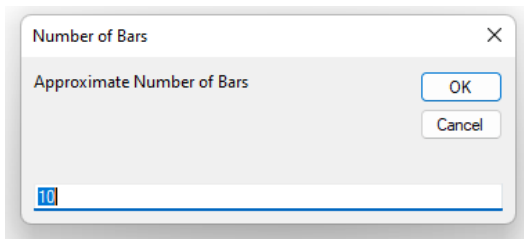
Sample	Obs 1	Obs 2	Obs 3	Obs 4	Obs 5
S1	265	205	263	307	220
S2	268	260	234	299	215
S3	197	286	274	243	231
S4	267	281	265	214	318
S5	346	317	242	258	276
S6	300	208	187	264	271
S7	280	242	260	321	228
S8	250	299	258	267	293
S9	265	254	281	294	223
S10	260	308	265	283	277
S11	200	235	246	328	296
S12	276	264	269	235	290
S13	221	176	258	263	231
S14	334	280	265	272	283
S15	265	262	271	245	301
S16	280	274	253	287	258
S17	261	248	260	274	337
S18	250	278	254	274	275
S19	278	250	265	270	298
S20	257	210	280	269	251

3. The histogram macro will prompt you for upper and lower spec limits. Use the default or input your spec limits. For one-sided histograms, click Cancel if there is no USL or LSL. In this example, we input 346 for the USL and 200 for the LSL.

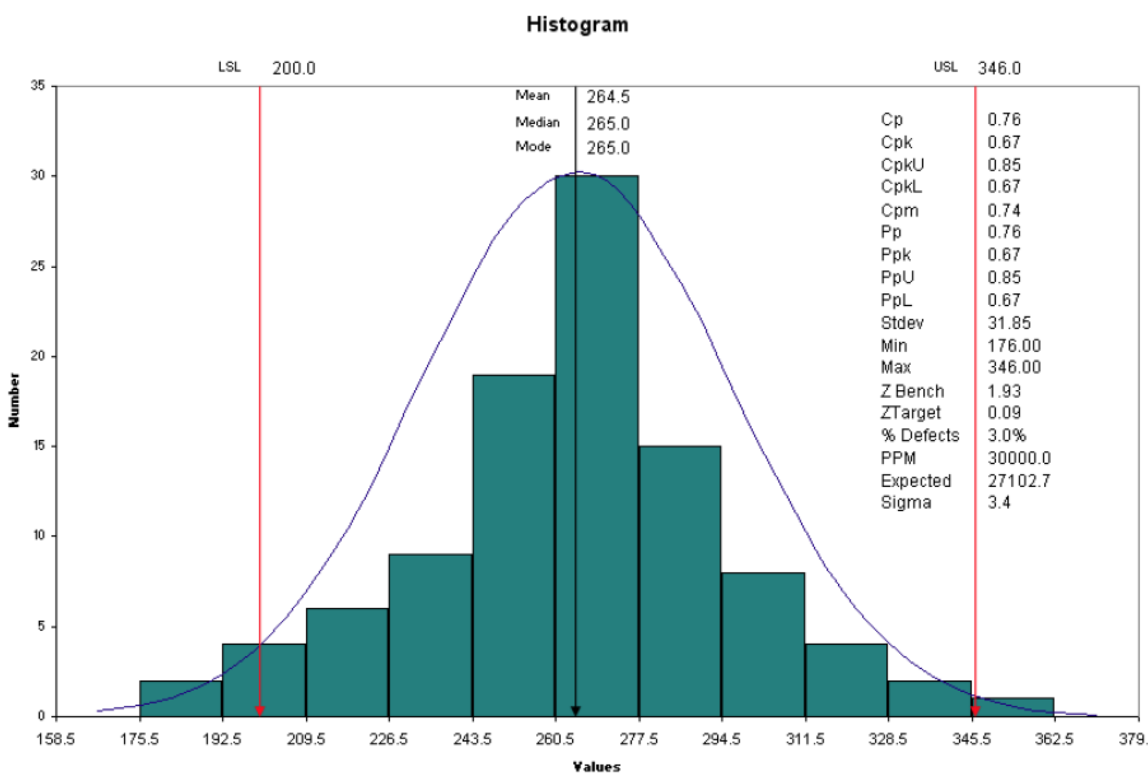




4. Next the macro will ask you for the approximate number of bars you would like to see on the histogram. Click OK to use the default or input a number. Note: this is only an approximation, the actual number of bars may be different.



5. Then, the macro will draw the graph for you. Notice that the capability indices equal what was manually calculated above.



6. To move the USL or LSL arrows or text boxes:
   - Arrows: Click on each arrow and drag it to the appropriate position. To extend an arrow, click on it, then click on the handle at either end and extend the arrow.
   - Text: Click on each text box and drag it to sit on top or beside its corresponding arrow.
7. From the F ile Menu, select Save to save the graph.
8. To revise the process capability calculations, switch to the data sheet. You can change the USL and/or LSL by typing over the current values in the cells provided. Excel will recalculate Cp, Cpk and Pp, Ppk and update the contents of the text boxes on the chart.

Maximum Value	346.0	278.0	32	32.0
Range	170.0	295.0	13	25.5
LSL	200.0	312.0	8	15.3
USL	346.0	329	**You can change these fields**	
Number of Bars	10.0	346		
Number of Classes	10.0	363.0	0	0.6
Class Width	17.0	380.0	0	0.1
Beginning Point	159.0	397.0	0	0.0

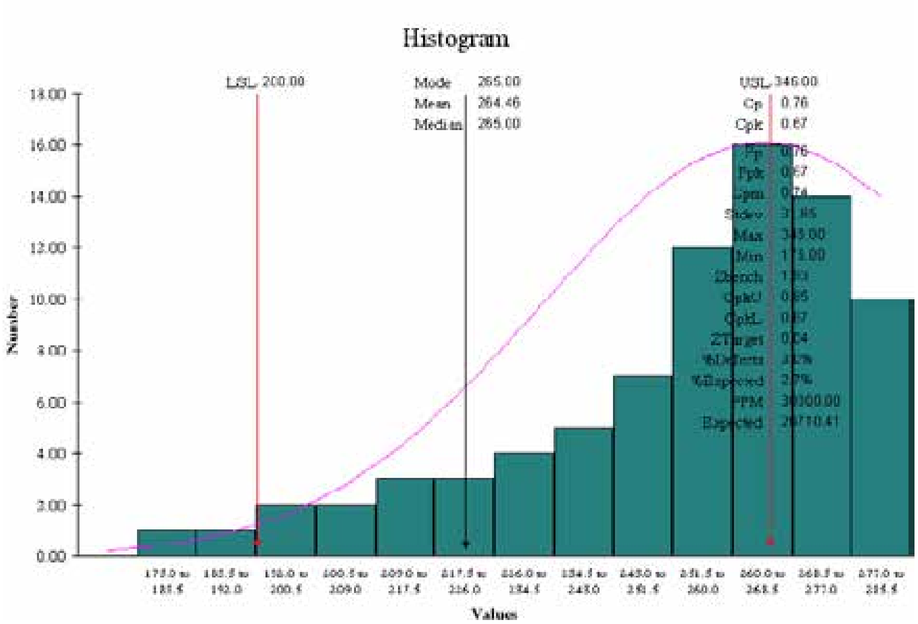
9. You can also change the number of bars used on the chart by changing either the number of bars field OR the class width field in the data sheet.

   Note: if you want an exact number of bars, use the class width field. Change the class width in small increments until you get the number of bars desired.

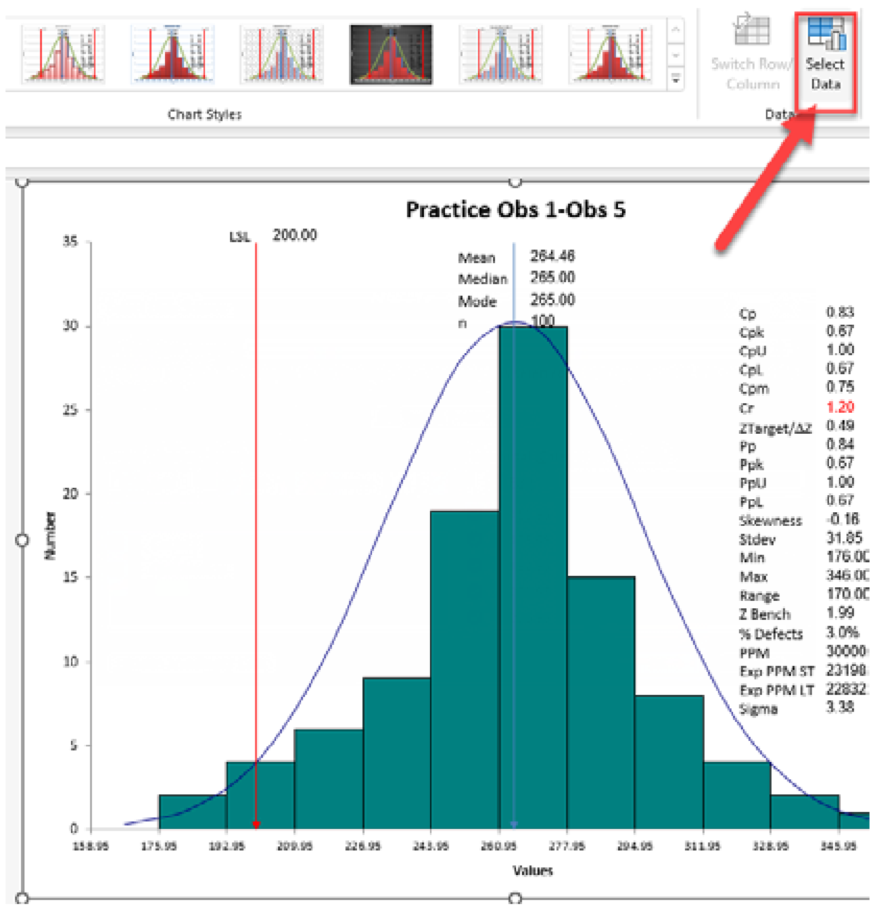
10. When you change the number of bars, the ranges and counts of data points for those ranges will change in the data sheet. You may even need to copy and paste formulas down to the new cells to include them all.

	A	G	H	I	J	K	L	M	N
1	Data1	Cp	1.0	Decimal Points	1.0	Range	Count	Distributior	Count2
14	221	ZTarget	0.0	Number of Classes	20.0	262.5	13	31.5	
15	334	% Defects	0.0%	Class Width	8.5	271.0	19	32.0	
16	265	PPM	0.0	Beginning Point	160.5	279.5	11	30.3	
17	280	Expected	2813.9	Stdev Est	31.9	288.0	10	0.0	
18	261	Sigma	4.3	d2/c4	0.9	296.5	4	0.0	
19	250			Target	264.5	305.0	5	0.0	
20	278		Observed	Expected	Z	313.5	2	0.0	
21	257	PPM<LSL	0.0	1412.7	-3.0	322.0	3	0.0	
22		PPM>USL	0.0	1401.2	3.0	330.5	1	0.0	
23		PPM	0.0	2813.9		339.0	2	0.0	
24		% Defects	0.0	0.0		347.5	1	0.0	
25				**Counts for New Bars**		356.0	0	0.0	
26						364.5	0	0.0	
27						373.0	0	0.0	

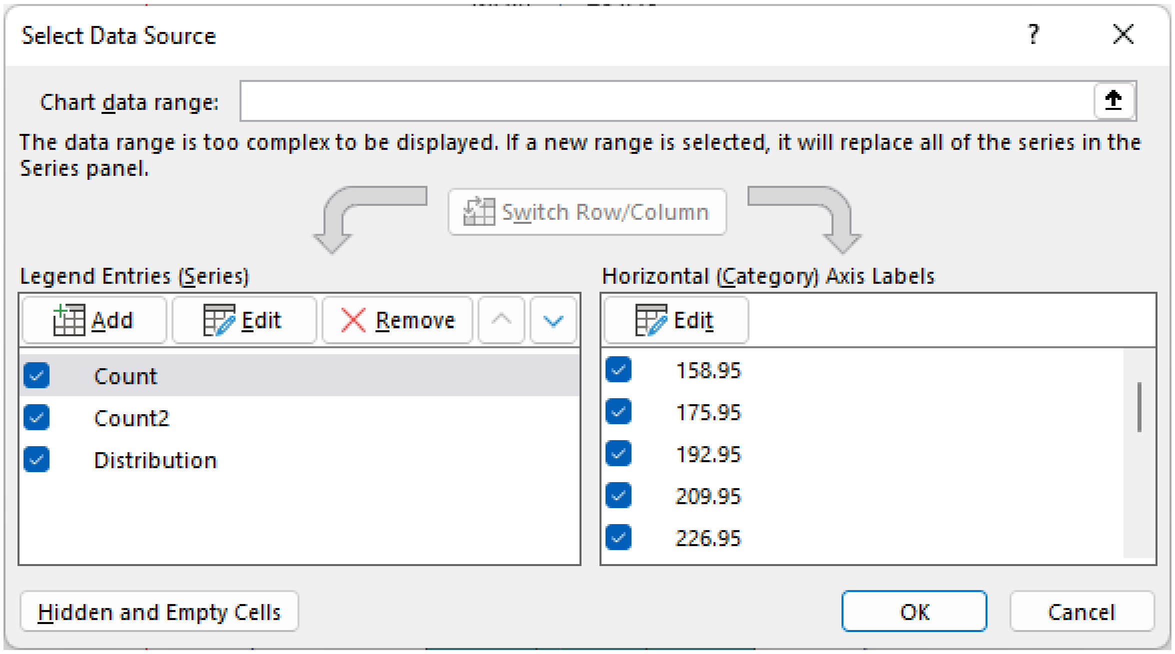
11. Next go to the Histogram chart and see if it is still centered on the page. In this example, we changed the number of bars from 10 to 20. After the change the chart looks like this:



12. Next, we need to update the Source Data the chart is graphing. Click the chart, then click Chart Design from the Excel menu, and select Select Data. Or, right-click on the chart, then click Select Data:



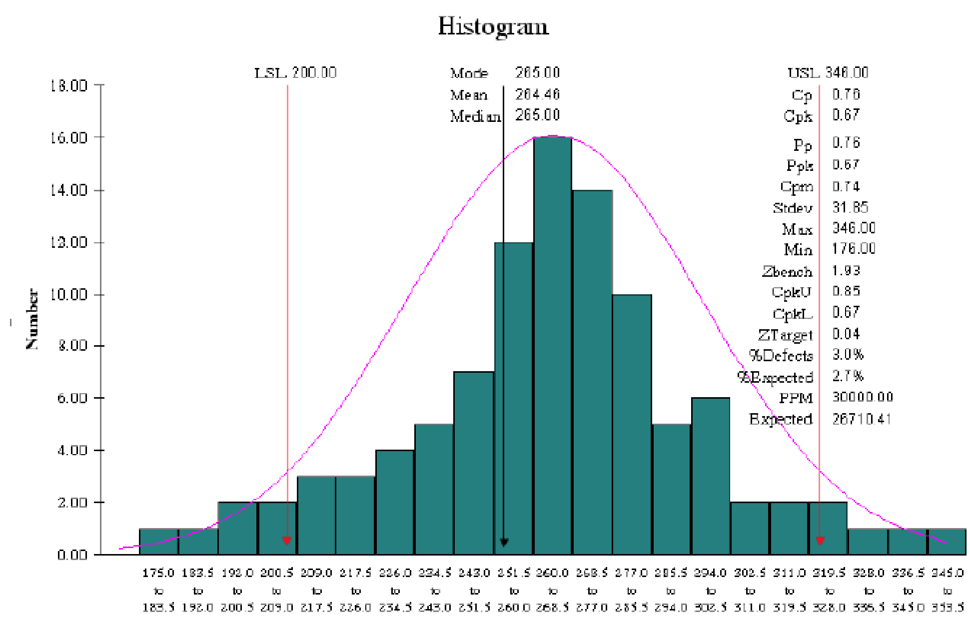
13. Excel will open a window with the source data definitions in the data sheet.



14. In step 10 we noted that our new count data goes down to row 24. Therefore, we need to change all references to row 16 to row 24 (change K$16 to K$24, etc).

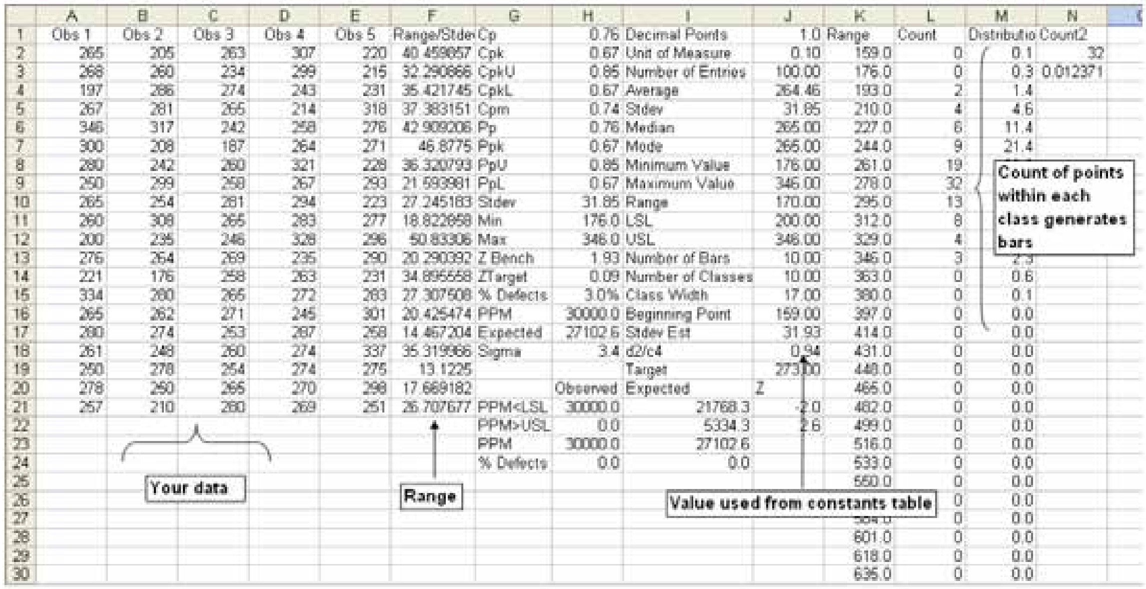
15. Change all of the values for the Count series (click Count, then select Edit) then click on Count 2 and Distribution and change the values for those as well.

16. The updated chart should look like this:



The QI Macros Calculations

The QI Macros use the formulas in this whitepaper. For a better understanding of how the calculations are done in the QI Macros and where they can be found, take a look at the data sheet created when a macro is run.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Obs 1	Obs 2	Obs 3	Obs 4	Obs 5	Range/Stde	Cp	0.76	Decimal Points	1.0	Range	Count	Distributio	Count2
2	265	205	263	307	220	40.459857	Cpk	0.67	Unit of Measure	0.10	159.0	0	0.1	32
3	268	260	234	299	215	32.290866	CpkU	0.85	Number of Entries	100.00	176.0	0	0.3	0.012371
4	197	286	274	243	231	35.421745	CpkL	0.67	Average	264.46	193.0	2	1.4	
5	267	281	265	214	318	37.383151	Cpm	0.74	Stdev	31.85	210.0	4	4.6	
6	346	317	242	258	276	42.909206	Pp	0.76	Median	265.00	227.0	6	11.4	
7	300	208	187	264	271	46.8775	Ppk	0.67	Mode	265.00	244.0	9	21.4	
8	280	242	260	321	228	36.320793	PpU	0.85	Minimum Value	176.00	261.0	19	[illegible]	
9	250	299	258	267	293	21.593981	PpL	0.67	Maximum Value	346.00	278.0	32	[illegible]	
10	265	254	281	294	223	27.245183	Stdev	31.85	Range	170.00	295.0	13	[illegible]	
11	260	308	265	283	277	18.822858	Min	176.0	LSL	200.00	312.0	8	[illegible]	
12	200	235	246	328	296	50.83306	Max	346.0	USL	346.00	329.0	4	[illegible]	
13	276	264	269	235	290	20.290392	Z Bench	1.93	Number of Bars	10.00	346.0	3	[illegible]	
14	221	176	258	263	231	34.895558	ZTarget	0.09	Number of Classes	10.00	363.0	0	0.6	
15	334	280	265	272	283	27.307508	% Defects	3.0%	Class Width	17.00	380.0	0	0.1	
16	265	262	271	245	301	20.425474	PPM	30000.0	Beginning Point	159.00	397.0	0	0.0	
17	280	274	253	287	258	14.467204	Expected	27102.6	Stdev Est	31.93	414.0	0	0.0	
18	261	248	260	274	337	35.319986	Sigma	3.4	d2/c4	0.94	431.0	0	0.0	
19	250	278	254	274	275	13.1225			Target	273.00	448.0	0	0.0	
20	278	250	265	270	298	17.669182		Observed	Expected	Z	465.0	0	0.0	
21	257	210	280	269	251	26.707677	PPM<LSL	30000.0	21768.3	-2.0	482.0	0	0.0	
22							PPM>USL	0.0	5334.3	2.6	499.0	0	0.0	
23							PPM	30000.0	27102.6		516.0	0	0.0	
24							% Defects	0.0	0.0		533.0	0	0.0	
25											550.0	0	0.0	
26											[illegible]	0	0.0	
27											[illegible]	0	0.0	
28											601.0	0	0.0	
29											618.0	0	0.0	
30											635.0	0	0.0	

About Jay Arthur

Jay Arthur, the KnowWare Man, solves problems of delay, defects and deviation - the three silent killers of productivity and profitability. He teaches people how to eliminate delay, defects and deviation in one day using Excel and the Magnificent Seven Tools of Lean Six Sigma. Jay is the shortcut to results with Lean Six Sigma.

Jay is first and foremost a Money Belt; he knows how to use data to fix broken processes to save time, save money and save lives. Jay has 25 years of experience helping companies save millions of dollars.

Jay is a frequent speaker at Lean Six Sigma conferences and is the author of many popular Lean Six Sigma books published by McGraw Hill including **Lean Six Sigma Demystified** and **Lean Six Sigma for Hospitals**. He is also the developer of **QI Macros Software for Excel.**