# Cpk vs Ppk: Who Wins?

Your supplier has sent you the process capability charts you requested. The supplier has produced some very nice charts, obviously generated with some fancy software package – and, of course, with all those accompanying statistics. You know, things like Cpk, Ppk, sigma level, ppm out of spec and so on. Very pretty charts. Looks like your supplier is really performing for you. You note one capability chart that has a Ppk = 1.14 and a Cpk = 2.07. Why are those different? Well, it doesn't matter. The Cpk is above 1.33, which is what you asked the supplier for. Time to work on something else. You just missed a very important piece of information about your supplier's performance. Know what it is?
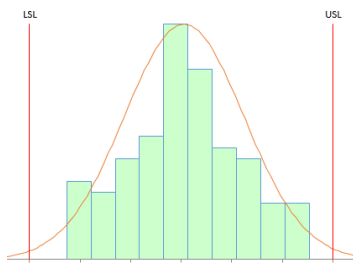
Cpk and Ppk are two commonly used measures of process capability – how well your process is meeting your customer specifications. Software today makes it easy to plug the data in and generate the results. But, far too often, we simply take the results and move forward without thinking about what they mean. In this month's SPC Knowledge Base publication we take an in-depth look at Cpk and Ppk. What are they? What are they measuring? What do the values means? Which one should you rely on? Some of the answers may well surprise you. In this publication:

Introduction

This is the third in a series of SPC Knowledge Base publications on process capability. Two months ago, we took an interactive look at process capability. We reviewed the process capability calculations, including Cp and Cpk. You could download an Excel workbook that let you visually see how changing the average and standard deviation of your process impacts your process capability. You were able to see visually how the process shifts versus your specifications. In addition, the workbook showed how Cp, Cpk, the sigma level, and the ppm out of specification changed as the average and standard deviation changed. If you are new to process capability, please take a moment to review that publication.

Last month's publication was entitled "Cpk Alone is Not Sufficient." In that publication we took a look at the why a Cpk value by itself is not sufficient to describe the process capability. We went through a process capability checklist designed to help you paint a true picture of your process capability – to increase the confidence you, your leadership, and your customers have in your process capability.

We did not mention Ppk in either publication. Time to change that in this publication.

Process Capability Review

Process capability analysis answers the question of how well your process meets specifications – either those set by your customer or your internal specifications. To calculate process capability, you need three things:
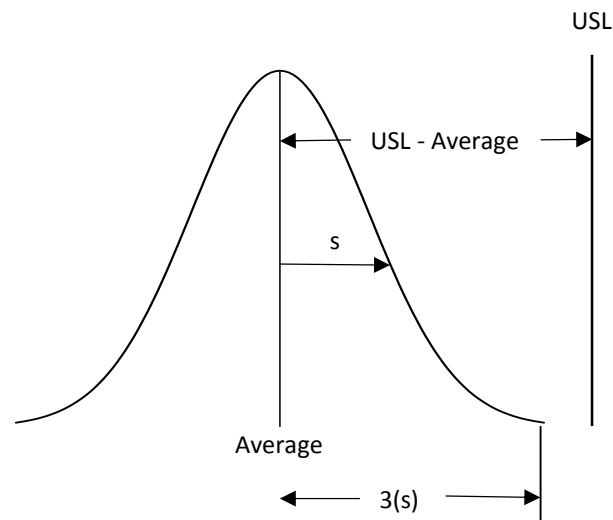
- Process average estimate
- Process standard deviation estimate
- Specification limit

This is true for both Cpk and Ppk. We will assume that our data are normally distributed.
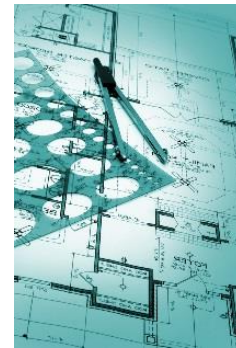
Process capability indices represent a ratio of how far a specification limit is from the average to the natural variation in the process. The natural variation in the process is taken as being 3 times the process standard deviation. Figure 1 shows the general set up for determining a process capability index based on the upper specification limit (USL) with "s" being a measure of the process variation.

Figure 1: Determining a Process Capability Index

The process capability index is then given by:

$$\text{Process capability index} = (USL - Average)/(3s)$$

To calculate process capability, you need to be able to estimate the process average and the process standard deviation. And no, it is not as easy as simply doing some calculations. Both of these statistics have to be "valid." We explore this more detail below.

Cpk and Ppk Review

Both Cpk and Ppk are the minimum of two process indices. The equations for Cpk and Ppk are shown in Table 1.

Table 1: Cpk and Ppk Equations

Cpk	Ppk
Cpk = Minimum (Cpu, Cpl)	Ppk = Minimum (Ppu, Ppl)
$$Cpu = \frac{USL - \overline{\overline{X}}}{3\sigma}$$	$$Ppu = \frac{USL - \overline{\overline{X}}}{3s}$$
$$Cpl = \frac{\overline{\overline{X}} - LSL}{3\sigma}$$	$$Ppl = \frac{\overline{\overline{X}} - LSL}{3s}$$

$\overline{\overline{X}}$ is the overall process average. In the Cpk equations, σ is used to estimate the process variation. σ is the estimated standard deviation obtained from a range control chart. In the Ppk equations, s is used to estimate the process variation. s is the calculated standard deviation using all the data.

Thus, the major difference between Cpk and Ppk is the way the process variation is estimated. So what is the difference between these two?

Within Subgroup Variation vs Overall Variation

The question of Cpk vs Ppk is really a question of within subgroup variation, σ, vs overall variation, s. Let's start with s or the calculated standard deviation, which is given by the equation below.

Within vs Overall

$$s = \sqrt{\frac{\sum (X_i - \overline{\overline{X}})^2}{N-1}}$$

X_i is an individual data point, $\overline{\overline{X}}$ is the overall average and N is the total number of data points. Look at the summation term under the square root sign: $(X_i - \overline{\overline{X}})^2$. This term is squaring how far each individual data point is from the overall average, as shown in Figure 2.

Figure 2: Standard Deviation (s)
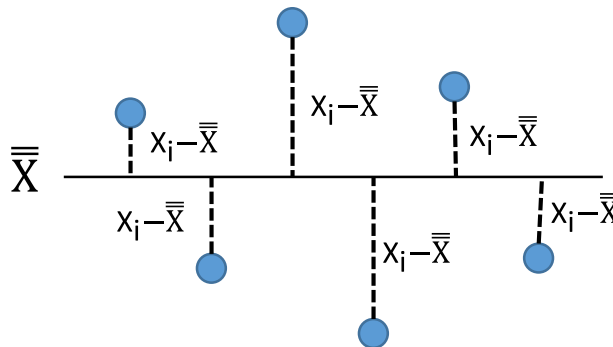
S According to the equation, you add up the squares of those deviations, divide by the total number of points minus 1 and take the square root. You can view the calculated standard deviation as the average distance each individual data point is from the overall average. Note that you use all the data in the calculation. This is why this standard deviation is sometimes called the *overall variation*. It accounts for all the variation in the data.

Now we will move to σ, which is usually referred to as the *within subgroup variation*. This estimate of the process standard deviation comes from a range control chart. For example, suppose you are using an $\overline{X}$-R control chart with a subgroup size of five. To form a subgroup, you take 5 samples. You calculate the average of those 5 samples. This is $\overline{X}$ and is plotted on the $\overline{X}$ chart. You also calculate the range of the subgroup values. The range, R, is the maximum value in the subgroup minus the minimum value in the subgroup. This is shown in Figure 3.

Figure 3: Within Subgroup Variation (σ)
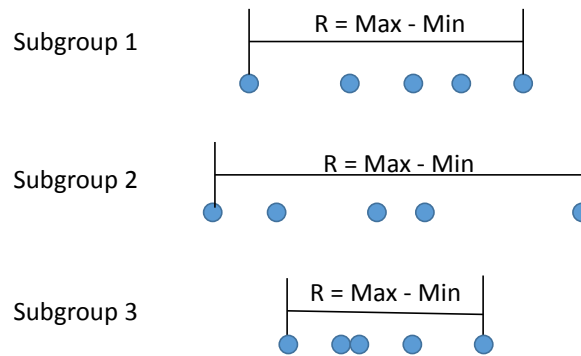
R is a measure of the variation within the subgroup. To calculate σ, you use the following equation:

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

where $\overline{R}$ is the average range and d_2 is a control chart constant that depends on subgroup size. So, σ accounts for the variation within the subgroup. It may or may not account for all the variation as we will see below.

That Little Issue of Statistical Control!

All our publications on process capability have stressed the need for the process to be in statistical control. How often is this just ignored? Last month we gave the process capability checklist developed by Dr. Don Wheeler to paint a true picture of your process capability. That checklist had five items:

1. Plot your data using a control chart to determine if the process is in statistical control (consistent and predictable)
2. For a process that is in statistical control, construct a histogram with the specifications added
3. For a process that is in statistical control, calculate the natural variation in the process data
4. For a process that is in statistical control, calculate Cp and Cpk

5. Combine these four items together and present them *all* when talking about process capability

See how often "for a process that is in statistical control" occurs? The point is that Cpk (and Ppk) have no meaning unless your process is in statistical control. And for the kicker: if your process is in statistical control, Cpk and Ppk will be very close to being equal. In fact, if you compare Cpk and Ppk values for a given process, you will find the following to be true:

- If Cpk is approximately equal to Ppk, the process is in statistical control
- If Cpk is significantly different that Ppk, the process is not in statistical control

So, when you looked at the supplier chart and noticed a big difference between Cpk and Ppk, you were given a key piece of information. Your supplier's process is not in statistical control – and you can't be sure of what you will get in the future.

In addition, if the process is not in statistical control, Cpk and Ppk have no meaning. You cannot be sure of getting similar values in the future because of the process is not consistent and predicable. We will explore this through in the following example for two processes with the same data – just in a different order.

Two Processes – Same Data, Same Ppk

We will use two processes that have the same data (the data from last month's publication). Suppose you are taking four samples per hour and forming a subgroup. You want to determine if your process is capable of meeting specifications (LSL = 65 and USL = 145). The data for the 30 subgroups for Process 1 are shown in Table 2.

Table 2: Process 1 Data

Day	X1	X2	X3	X4	Day	X1	X2	X3	X4
1	90.2	113.8	111.8	104.4	16	100.8	106	101.5	108.8
2	105.6	98.8	109.3	113.5	17	96.7	101.3	100.4	95.1
3	104	84.5	98.9	97	18	105.1	92	92.5	95
4	112.4	86.2	85.5	106.5	19	104.5	94.5	91.3	82.7
5	96.6	99.9	112.9	96.8	20	110.1	110.7	104	115.6
6	91.7	101.3	107.1	101.2	21	116.9	86.3	96.4	99.3
7	112	97.9	109	95.2	22	78.9	91.4	96.5	109.2
8	91.8	98	98.1	79.2	23	112.2	110.5	98.3	109.2
9	94.9	87.1	104.3	112.7	24	88.8	105.9	86.3	76
10	101.1	104	101.1	102.7	25	98.6	93.5	106.2	92.8
11	100.6	83.3	96.6	88.5	26	99.1	99.6	83.6	106.5
12	80.5	95	98.3	113.6	27	90.5	110	82.6	86
13	89.2	93.9	98.5	106.7	28	106.7	107.9	109.9	108.8
14	96.7	96.8	106.2	90	29	87.4	95	108.5	96.7
15	74.2	104.3	111.2	108.7	30	112.7	78.4	112.8	81.1

The data for Process 2 are the same, just in a different order. These data are shown in Table 3.

Table 3: Process 2 Data

Day	X1	X2	X3	X4	Day	X1	X2	X3	X4
1	105.6	113.5	107.1	101.2	16	92	92.5	95	96.6
2	112.4	109.3	104	101.3	17	90.5	96.7	82.6	86
3	104.3	112.7	112	100.6	18	101.1	104	101.1	102.7
4	104.3	111.2	108.7	113.6	19	100.8	106	101.5	108.8
5	101.3	100.4	95.1	106.7	20	110.1	110.7	104	115.6
6	105.1	116.9	109.2	110	21	112.2	110.5	112.8	109.2
7	104.5	112.9	105.9	109.9	22	94.5	91.3	82.7	91.7
8	108.5	108.8	106.7	107.9	23	86.3	96.4	99.3	80.5
9	106.5	106.2	106.2	98.8	24	88.8	74.2	86.3	76
10	86.2	85.5	106.5	84.5	25	113.8	111.8	104.4	112.7
11	97.9	109	95.2	98.9	26	98.6	93.5	95	92.8
12	83.3	96.6	88.5	97	27	99.1	99.6	83.6	96.7
13	95	98.3	94.9	99.9	28	78.4	98.3	81.1	87.4
14	89.2	93.9	98.5	96.8	29	78.9	91.4	96.5	90.2
15	96.7	96.8	87.1	90	30	91.8	98	98.1	79.2

Since they are the same, the data in Tables 2 and 3 have the same average and the same standard deviation.

Average = 98.98

Standard deviation (s) = 9.89

Now draw a histogram for the data in Table 2 and a histogram for the data in Table 3. Throw in the specifications: LSL = 65 and USL = 145. If you do this, you will discover that the histograms are exactly the same – just what you expect since the data are the same. Figure 4 shows the histogram.

This process looks very good – definitely within the specifications. You are one happy person. You can calculate Ppk for Process 1 and Process 2. Since the average and standard deviation are the same, Ppk will be the same for both processes. The calculations are:

Ppk = Minimum (Ppu, Ppl)

$$\text{Ppu} = \frac{\text{USL-}\overline{\overline{X}}}{3s} = \frac{145 - 98.98}{3(9.89)} = 1.55$$

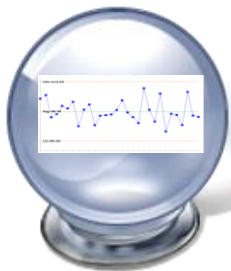$$\text{Ppl} = \frac{\overline{\overline{X}}\text{-LSL}}{3s} = \frac{98.98 - 65}{3(9.89)} = 1.14$$

Figure 4: Histogram of Process Data with Specifications Added
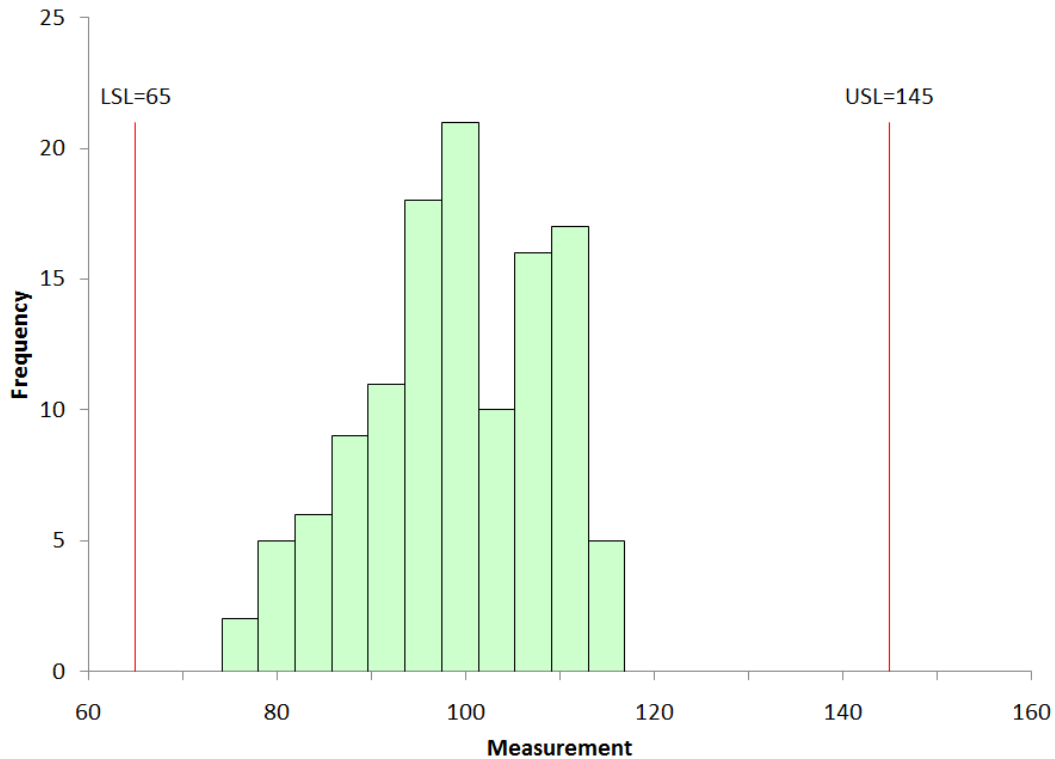
So, Ppk = 1.14 for Process 1 and Process 2.

Two Processes – Same Data, Different Cpk

But wait – what is the value for Cpk for these two processes? Are the same? No, they are not the same. Remember Cpk is based on the within subgroup variation. And although the data are the same in both processes, they are in different order – which changes the within subgroup variation. To calculate the Cpk values, you need to estimate the standard deviation (σ) from the range chart and the overall process average from the $\overline{X}$ chart. This is a very important difference between the Ppk approach and Cpk approach. Ppk simply uses calculations; Cpk uses control charts to estimate the average and the process variation. It is the way you tell the future about your process.

The control charts for Process 1 are shown below. The range chart is shown in Figure 5. The range chart is in statistical control. Because it is in control, it is consistent and predictable. You can now estimate the standard deviation using the following:

$$\sigma = \frac{\overline{R}}{d_2} = \frac{19.65}{2.059} = 9.54$$

Note that since the range chart is in statistical control, the within subgroup variation is consistent and predictable. The value for the process standard deviation is "valid." The process that generated it is consistent and predictable and will remain so as long as the process stays the same.

Figure 5: Process 1 Range Chart

Est. Sigma = 9.54
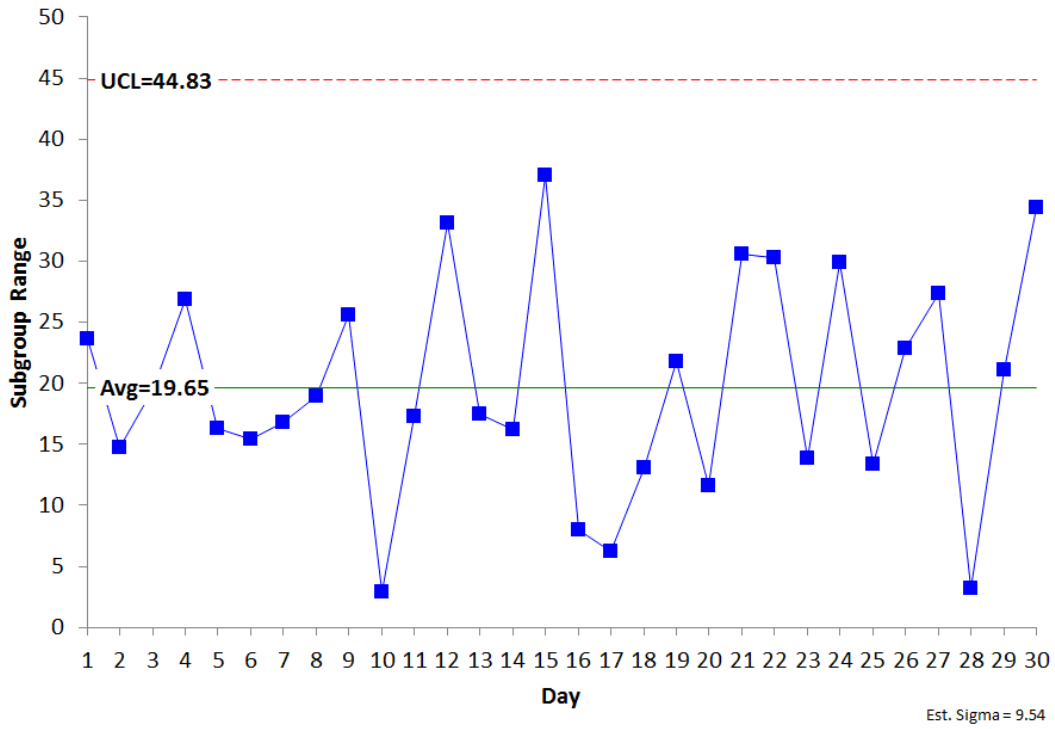
To calculate Cpk, you need an estimate of the average. That comes from the $\overline{X}$ control chart. Figure 6 shows this chart.

Figure 6: Process 1 $\overline{X}$ Chart
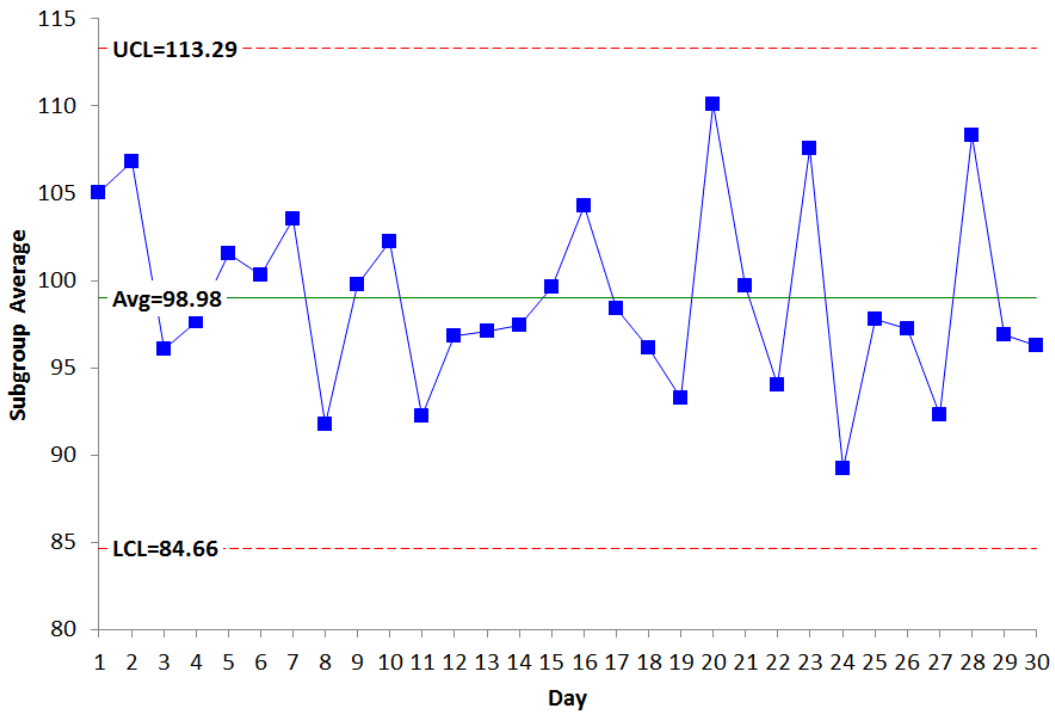
Cpk

The $\overline{\mathrm{X}}$ chart is also in statistical control. The between subgroup variation is consistent and predictable. This means that you have a good ("valid") estimate of the process average. You can now use that average, along with σ, to determine the Cpk values.

Cpk = Minimum (Cpu, Cpl)

$$\mathrm{Cpu} = \frac{\mathrm{USL}\text{-}\overline{\overline{\mathrm{X}}}}{3\sigma} = \frac{145 - 98.98}{3(9.54)} = 1.61$$

$$\mathrm{Cpl} = \frac{\overline{\overline{\mathrm{X}}}\text{-LSL}}{3\sigma} = \frac{98.98 - 65}{3(9.54)} = 1.19$$

Now compare the results for Ppk and Cpk for Process 1:

Process 1: Ppk	Process 1: Cpk
s = 9.89	σ = 9.54
Ppu = 1.55	Cpu = 1.61
Ppl = 1.14	Cpl = 1.19

Note how close the results are. This will always be the case when the process is in statistical control. This is because of the following:

> *When a process is in statistical control, the within subgroup variation is a good estimate of the overall process variation, i.e., σ = s.*

Cpk is essentially the same as Ppk in this case. They are giving you the same information.

Let's move on to Process 2. The range chart for Process 2 is shown in Figure 7. The range chart is in statistical control – the within subgroup variation is consistent and predictable. You can now estimate the standard deviation using the following:

$$\sigma = \frac{\overline{\mathrm{R}}}{\mathrm{d}_2} = \frac{11.25}{2.059} = 5.46$$

Again you need an estimate of the average to determine Cpk. This comes from the $\overline{\mathrm{X}}$ chart, which is shown in Figure 8

The $\overline{\mathrm{X}}$ chart is not in statistical control – the between subgroup variation is not consistent and predictable. There are points beyond the control limits, runs above the averages – all sorts of problems with the stability of this process.

Out of Control

This means that you do not have a good estimate of the process average. It is moving around. What will the next subgroup average be? You have no idea where it will be. The process is not consistent and predictable. You can't really calculate the Cpk value.

Figure 7: Process 2 Range Chart

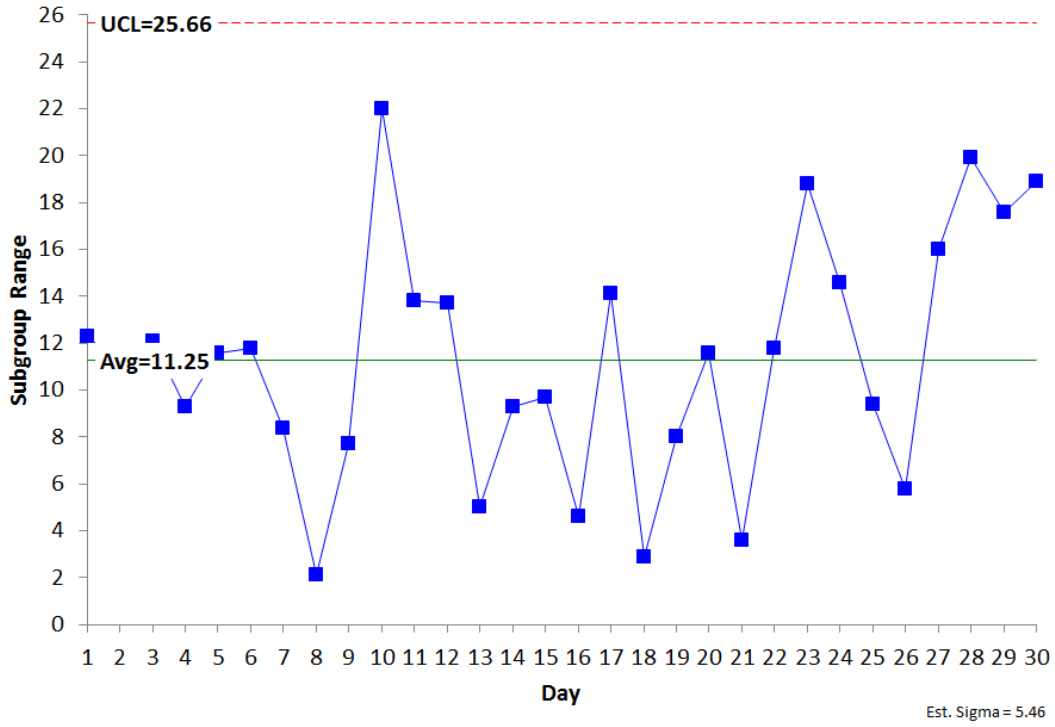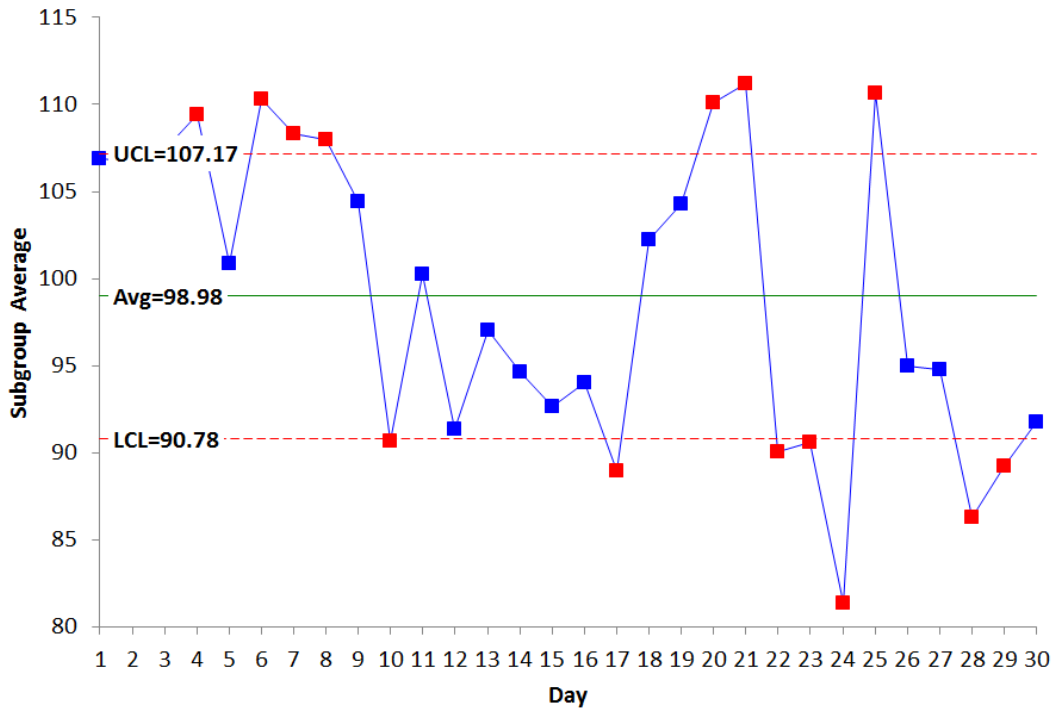Est. Sigma = 5.46

Figure 8: Process 2 $\bar{\text{X}}$ Chart

Many times folks just simply ignore this fact and move full steam ahead with calculating Cpk. After all, the calculated average is 98.98. The Cpk calculations are as follows:

$$Cpk = \text{Minimum (Cpu, Cpl)}$$

$$Cpu = \frac{USL\text{-}\overline{\overline{X}}}{3\sigma} = \frac{145 - 98.98}{3(5.46)} = 2.81$$

$$Cpl = \frac{\overline{\overline{X}}\text{-}LSL}{3\sigma} = \frac{98.98 - 65}{3(5.46)} = 2.07$$

So, Cpk for Process 2 is 2.07. Now compare the results for Process 1 and Process 2.

Processes 1 and 2: Ppk	Process 1: Cpk	Process 2: Cpk
s = 9.89	σ = 9.54	σ = 5.46
Ppu = 1.55	Cpu = 1.61	Cpu = 2.81
Ppl = 1.14	Cpl = 1.19	Cpl = 2.07

The values of Cpu and Cpl for Process 2 are significantly different than the Ppu and Ppl values. And it all boils down to the issue of statistical control. When there are significant differences in σ vs s, Cpu vs Cpl and Cpl vs Ppl, it is a very strong indication that the process is not in statistical control.

Summary: So, Who Wins: Cpk or Ppk?

The reality is that Cpk is a better estimate of the potential of your process. It represents the best your process can do and that is when the within subgroup variation is essentially the same as the between subgroup variation. This is what it means to be in statistical control. And if the process is in statistical control, Cpk is essentially the same as Ppk. So, you really don't need Ppk in this case.

And if your process is not in statistical control, you have something to work on – Cpk and Ppk are pretty well meaningless – except for the fact that values of Cpk and Ppk that are widely different are indications that the process is not in statistical control.

But you know that already because you are following the process capability checklist from Dr. Wheeler. Always start by looking at the data in control chart format.

Quick Links

Visit our home page

SPC for Excel Software

SPC Training

SPC Consulting

SPC Knowledge Base

Ordering Information

Thanks so much for reading our publication. We hope you find it informative and useful. Happy charting and may the data always support your position.

Sincerely,

Dr. Bill McNeese
BPI Consulting, LLC