PROCESS CONTROL

By Kevin Caskey

Professor of Operations Management and Quantitative Methods, SUNY New Paltz

Following statistics to the right conclusion

All processes have variation, some of it random and some assignable to a cause. Statistics can help us recognize which is which and to make the appropriate fix.

"THERE ARE THREE kinds of lies: Lies, Damned Lies, and Statistics" is a famous quote attributed by Mark Twain to the British statesman, Benjamin Disraeli. Whoever actually first said this, the sentiment seems to involve the use of twisted numbers to delude. I want to talk about the use of numbers to avoid deluding ourselves.

In our businesses and on our factory floors, the 'customer' for statistical analysis is often ourselves. Let's look at using Statistical Process Control (SPC)—often called Statistical Quality Control (SQC)—to help us avoid reaching unwarranted conclusions. Have we not all heard statements like:

"John is not working out, there are more complaints against him"

"There is something wrong with the fourth line, we keep needing to reject parts"

"This last week was great, we are rejecting fewer completed lots"

All of these statements could be correct. On the other hand, it may be that we are being misled by random variation in the process. Let's look at process, random variation, and then variation that seems to have changed (so that we want to find the cause).

We call this 'process control'. We can view pretty much everything we do as a process. We take our inputs, do something, and deliver either a good or a service. The 'do something' is our process. Whatever we are doing, there is going to be some variation, or spread, in the results. We may not be able to see it, it may not matter to the customer, but all processes have variation. Customer service may always answer between the first and second ring, but that is not the same as always answering at the same time. The machine may drill every hole within specification, but that is not the same as all holes being identical. We call this the common or random variation.

On the other hand, it could be that things have, indeed, changed. John may very well be the source of more complaints than the typical representative. The fourth line could be out of adjustment. We would call these 'assignable causes' of variation. We want to identify that the change exists, find the cause, and either fix it (if things have gotten worse) or incorporate the change in process into our standard practices (if the change caused things to get better). Our task now is to separate assignable variation from random variation. We will use statistical control charts to do this.

In building control charts, we rely on some properties of normal distributions. We can assume, for example, that means of observations from large enough samples themselves be distributed normally. We know that if something follows the normal distribution:

68% of them will be between ± 1 standard deviation

95% of them will be between ± 2 standard deviations

99.7% of them will be between ± 3 standard deviations

It's this last one we will use. We can read that as 'pretty much all of them' should be within 3 standard deviations. With this knowledge, we now collect data and build our charts. The charts differ somewhat

depending on our situation. The three statements above represent the three basic situations:

- Are we dealing with something we can measure (such as diameters of the parts coming off line 4)?
- Are we dealing with something that is a percentage (such as the percent of lots we reject in a given hour, day, or week – at most 100 percent)?
- Are we dealing with something that we can only count (such as the complaints received in a department per day – there is no concept of 100 percent)?

Eggs break. But are we breaking more eggs?

As an illustration, let's discuss dealing with something that is a percentage. Consider the following data, tracking the number of broken eggs per package:

Package	# broken	% broken
1	2	0.167
2	1	0.083
3	3	0.250
4	1	0.083
5	2	0.167
6	4	0.333
7	2	0.167
8	6	0.500
9	4	0.333
10	3	0.250

The average (mean) percent of broken eggs is 0.233.
The standard deviation formula for percentages is

$$Std\ Dev_p = \sqrt{\frac{(\bar{p} \times (1 - \bar{p})}{n}}$$

Which, in our case, is 0.122.

We now need to find the centerline of our chart and the upper and lower control limits. Remember that we expect that, under typical conditions, observations remain within plus or minus three standard deviations. So, the upper and lower limits are:

Upper Control Limit: $\bar{p} + 3 \times \sigma_p = 0.599$

Lower Control Limit: $\bar{p} - 3 \times \sigma_p = -0.133$, which we treat as zero because a proportion cannot be negative.

All of the 10 proportions given in the table above are between the upper and lower limit. So, barring any other patterns, we can say the process is in control.

The difference between control and a good job.

Now let's talk about the difference between being in control and doing a good job. The results of building the chart above show that the process is in control. That is, no point is outside of what we would expect under our normal operating conditions. This does not mean that we are doing a good job. Breaking almost a quarter of our eggs on average cannot be seen as good. All we know from our control charts is we are not occasionally doing significantly better or worse than what is typical. Our problem is not that the process is out of control (in SPC terms), it is that we don't have a process that performs well. We need to find ways to break fewer eggs.

To illustrate this, let's say we trained our workers in better handling practices and then we collected new data, shown below.

Broken Eggs per Package (Dozen)

Package	#. broken	% broken
1	0	0.000
2	1	0.083
3	0	0.000
4	0	0.000
5	1	0.083
6	3	0.250
7	0	0.000
8	1	0.083
9	0	0.000
10	0	0.000

Now, the average (mean) percent of broken eggs is 0.05.

Using the standard deviation formula for percentages yields a value, in this case, of 0.063. And

Upper Control Limit: $\bar{p} + 3 \times \sigma_p = 0.239$

Lower Control Limit: $\bar{p} - 3 \times \sigma_p = -0.138$, so once again we use zero as the lower bound.

Our process is much improved. The variability is much lower, and the control limits are much tighter. Also, the sixth proportion is 0.25, which is now above the upper bound. The process, while now improved, is out of control. We would not be fooling ourselves if we concluded it is worth looking into what possibly could have lead to the high number of broken eggs in the sixth package.

The application above is somewhat artificial. Just how likely is it that a process would be performing poorly but fairly consistently? If it is performing poorly, it probably has bad and worse. SPC would allow us to identify the worse, eliminate those causes, and tighten the limits. Continuing this process would allow us to continue to get better.