STATISTICAL PROCESS CONTROL

(This is an excerpt from Rhinehart and Bethea, Applied Engineering Statistics, 2nd Edition, CRC Press, Boca Raton, FL, 2022, ISBN: 9781032119489.)

1 SPC CONCEPTS

Statistical process control (SPC) is an application of statistics to manufacturing processes. SPC provides evidence that feeds back information to the process manager that something has changed, relative to normal variation. It is an element in 6-sigma practices. SPC is not automatic feedback control that keeps a process at a set point. Instead, it indicates when there is high confidence that something has happened, and triggers an investigation to change something to either prevent the undesired event from happening again, or to ensure a fortuitous event becomes part of normal practice. Progressive prevention of upsetting events leads to improved uniformity, quality, and process yield.

Imagine a process that has noise but has neither measurable changes nor systematic drifts. In such a case, there is no need for either automatic or human control action. A graph of some process variable (PV) versus time may appear like that illustrated in Figure 1.

Control action based on normal noise is both unnecessary and unwanted. There are several reasons: Such action increases valve wear, servomechanism wear, etc. and can upset other process variables. Further, if the process is on target, input adjustment causes real change, which requires subsequent counter adjustment, and this “tampering” actually increases the process variability and burdens the organization. (To visualize the concept of tampering, search the Internet for W. Edwards Deming’s Funnel Experiment.)

The process of Figure 1 is in statistical control because there is no evidence that the PV has significantly changed in level or variation. The trend is characterized as random variation with the same noise characteristics about the same mean. There is no evidence of an *assignable cause*. An assignable cause means that something has happened and created statistical evidence that either the level or the variation change. The term *assignable* means that the cause might not be known, but it could be uncovered with investigation. Managerial control action is warranted only when an assignable cause happens.

Figure 2 illustrates some process variable response to some systematic change. One example of a systematic change or assignable cause is an orifice flowmeter calibration change due to erosion. The result might be a change in blended composition or *level* of one of the products in a refinery.

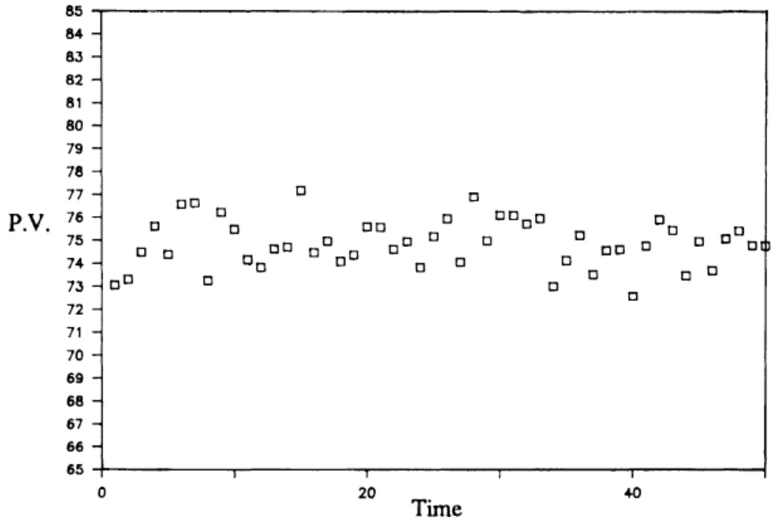

Figure 1 Process variable versus time (no changes in level or variability).

Figure 3 illustrates some process variable response to some other assignable cause that changed the process variability. In this case, perhaps loosened bolts caused increased vibration in a man-made fiber extrusion process, resulting in an increase in fiber thickness variability.

Trends in overall level, sudden shifts in level, cycling in level, and changes in variability all indicate the presence of an assignable cause. The assignable cause need not be known. Nor need it be identified. However, it represents some real change, as it either shifted the mean or changed the variability (or both) and therefore deserves action.

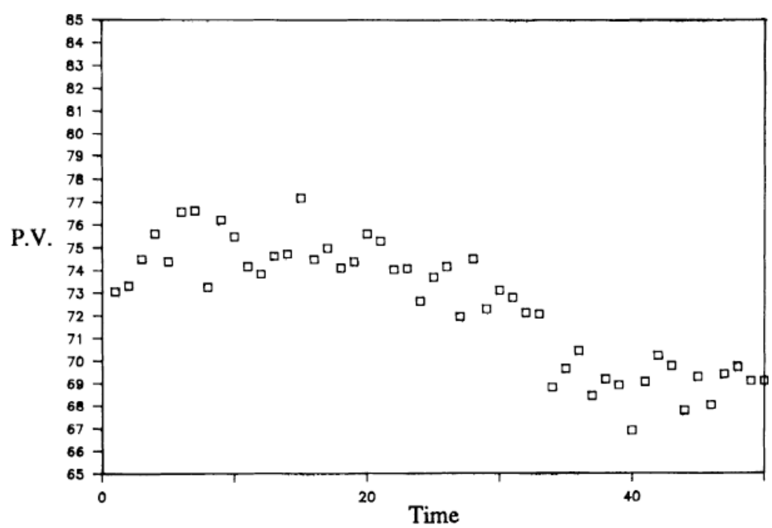

Figure 2 Process variable versus time (change in level).

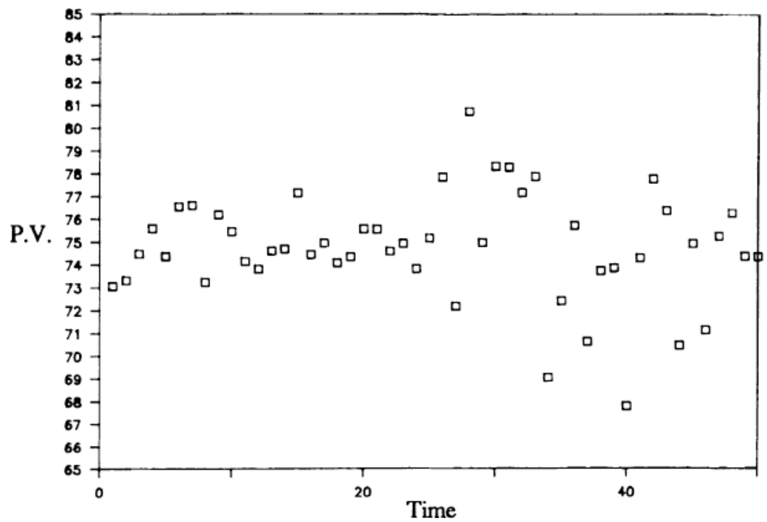

Figure 3 Process variable versus time (change in variability).

One of the objectives of SPC is to determine where the assignable cause occurred so that the operator knows where in the process to take corrective action. This procedure requires sampling appropriate variables at strategic points in the process. Another of the objectives of SPC is to determine when an assignable cause occurred. So, in opposition to classical automatic control that works constantly, SPC is a supervisory strategy that identifies when operator or management intervention is justified. Procedures for rational diagnosis to determine what happened and to determine appropriate ways to fix things are part of the body SPC and of Six-Sigma. This chapter will focus on the statistical procedure for detection.

Statistically, an assignable cause occurs when an observed change in either process variable level or variability has about a 0.26% or less chance of occurring.

Why was $\alpha = 0.0026$ chosen as the significance level? It represents the $\pm 3\sigma$ limits and was chosen from experience to balance the occurrence of Type 1 and Type 2 errors, i.e., justified intervention action and not reacting to false alarms. If, for instance, the $\pm 2\sigma$ limits (95.45% CI, $\alpha = 0.0455$) were chosen, then normal process variability would appear to have an assignable cause 4.55% of the time. If 1000 process variables were being observed in a plant each day, 46 assignable causes would seem to occur and 46 possibly unjustified control actions would be taken. On the other hand, if the $\pm 4\sigma$ limits (99.994% CI, $\alpha = 0.00006$) were chosen, then unjustified control action would be taken only twice per month. However, at the $\pm 4\sigma$ limits, some assignable causes would go unrecognized until there was enough information to be 99.994% sure that the process variable behavior was not simply due to normal variability.

The $\pm 3\sigma$ limit is a subjective choice. There are other conventions. In some regions, the convention is to use the $\pm 3.09\sigma$ limits, which encompass 99.8% of the normal data. Further, where subsampling of $n = 2$ is used, the 96.5% confidence limits are often used. Although the 99.74% confidence interval is one convention, you may choose to modify those limits based on your process experience.

SPC will not replace on-line primary regulatory process control; however, SPC-triggered action can be automatically implemented within a feedback control loop as an alternate to filtering to temper response to noise.

The most common statistical tools within SPC are capability indices, $\bar{X}$ and either *S* or *R* charts, and cumulative sum and attribute charts. This discusses $\bar{X}$ and *S* charts. For extended topics, see Chapter 21 in Rhinehart and Bethea, Applied Engineering Statistics, 2nd Edition, CRC Press, Boca Raton, FL, 2022, ISBN: 9781032119489.

2 MEAN AND STANDARD DEVIATION CHARTS

Introduced by Walter A. Shewhart in the 1920s, the $\bar{X}$ (*X*-bar, average) and *R* (range) chart are some of several Shewhart charts devised to detect when an assignable cause occurred in a continuous process variable. Although the $\bar{X}$ and *S* chart is statistically more rigorous, the $\bar{X}$ and *R* chart is fully functional, was preferred in the pre-computer age, has advantages for small subgroup size sampling, and has become an accepted SPC standard. We will develop the $\bar{X}$ and *S* chart with an example to illustrate the principles involved.

Table 1 lists measured values of a continuous process variable that are based on subgroup sampling of size 4. In this example, polyester staple yarn is being measured for denier (a measure of thickness, in decitex or dtex, gm/10,000 m) once per shift. The past 25 shift samplings are listed. The average of each subgroup, $\bar{X}$, and the standard deviation of each subgroup, are also listed. The range value on each subgroup is the difference between the largest and smallest of the 4 samples. The average of the entire set of 100 individuals, $\bar{\bar{X}}$, is 74.6985 dtex and the standard deviation, σ, of the entire set of individual measurements is 1.27297 dtex. The average of the subgroup standard deviations, $\bar{S}$, is 1.049861 dtex, and the average range $\bar{R}$ is 2.2974 dtex.

Table 1 Sampled Yarn Deniers

Sample No.	Shift	Date	Sample values (decitex)				Subgroup average	Subgroup std. dev.	Range
1	2	8/19	73.07	73.33	74.49	75.62	74.1275	1.170934	2.55
2	3	8/19	74.38	76.57	76.63	73.26	75.21	1.669073	3.37
3	1	8/20	76.22	75.49	74.17	73.83	74.9275	1.120209	2.39
4	2	8/20	74.63	74.72	77.18	74.47	75.25	1.290814	2.71
5	3	8/20	74.98	74.10	74.37	75.59	74.76	0.6645796	1.49
6	1	8/21	75.58	74.62	74.96	73.84	74.75	0.7252611	1.74
7	2	8/21	75.19	75.96	74.07	76.91	75.5325	1.202285	2.84
8	3	8/21	75.00	76.11	76.10	75.73	75.735	0.5209285	1.11

9	1	8/22	75.96	73.02	74.15	75.25	74.595	1.287234	2.94
10	2	8/22	73.55	74.59	74.62	72.60	73.84	0.9648132	2.02
11	3	8/22	74.77	75.93	75.46	73.49	74.9125	1.06127	2.44
12	1	8/23	74.98	73.72	75.09	75.42	74.8025	0.7454901	1.70
13	2	8/23	74.79	74.79	75.05	73.84	74.6175	0.5326279	1.21
14	3	8/23	73.16	75.97	75.81	73.13	74.5175	1.586221	2.84
15	1	8/24	74.77	75.39	75.63	74.42	75.05251	0.5559604	1.21
16	2	8/24	75.41	73.91	76.02	74.87	75.0525	0.8948886	2.11
17	3	8/24	76.22	74.96	76.19	74.53	75.475	0.8611035	1.69
18	1	8/25	75.09	77.77	74.32	75.74	75.73	1.47867	3.45
19	2	8/25	75.93	75.12	75.35	75.79	75.5475	0.3772181	0.81
20	3	8/25	75.13	74.42	76.23	76.00	75.445	0.8314843	1.81
21	1	8/26	73.43	74.83	73.77	74.51	74.135	0.6465035	1.40
22	2	8/26	74.98	73.56	74.12	72.59	73.8125	1.002642	2.39
23	3	8/26	72.77	71.23	75.16	75.41	73.6425	2.00069	4.18
24	1	8/27	76.48	72.08	71.42	74.22	73.55	2.289948	5.06
25	2	8/27	72.27	72.57	73.39	71.54	72.4425	0.7655659	1.85

Example 1 Develop an $\bar{X}$ chart for the data in Table 1.

Figure 21.4 shows the $\bar{X}$ chart, on which the past 25 subgroup averages are plotted against the sequential sample number. The upper and lower control limits, UCL and LCL, are

$$\text{UCL} = \mu + \frac{3\sigma}{\sqrt{n}} \cong \bar{\bar{X}} + \frac{t_{\alpha,\upsilon}\ s}{\sqrt{n}} \cong \bar{\bar{X}} + A_3\bar{S} \qquad (1)$$

$$\text{UCL} = \mu - \frac{3\sigma}{\sqrt{n}} \cong \bar{\bar{X}} - \frac{t_{\alpha,\upsilon}\ s}{\sqrt{n}} \cong \bar{\bar{X}} - A_3\bar{S} \qquad (2)$$

where n is the subgroup sample size (of $n = 4$) and $\sigma/\sqrt{n}$ represents the standard deviation of the $\bar{X}$, the subgroup means. The ideal first part of the equations presumes that the standard deviation of the individual measurements is known and that $\bar{\bar{X}} = \mu$, the process mean. If the process is in statistical control (only subject to random independent perturbations), the average and standard deviation of the many individuals (100 in this case) could be used as estimates and the t-statistic value would replace the z-value of 3. However, if the process is not in statistical control, if an assignable cause has occurred, then the 100 individuals would not represent the same mean or variance, and a safer estimate of the standard deviation would be the average of the subgroup standard deviations. The last part of Equations (1) and (2) represent how the UCL and LCL values in an X-Bar chart are calculated. The value for A_3 comes from many sources.

Note that UCL and LCL are the process ±3σ capability limits and are *not* the product specification limits. The product specifications are not normally shown on the control chart. The process $\pm 2\sigma$ limits are occasionally illustrated and are termed the upper and lower warning limits (UWL and LWL). The UWL and LWL values are useful when a process capability index is low. In such a case, UWL or LWL violations are used to trigger a second sampling and, if warranted, subsequent action.

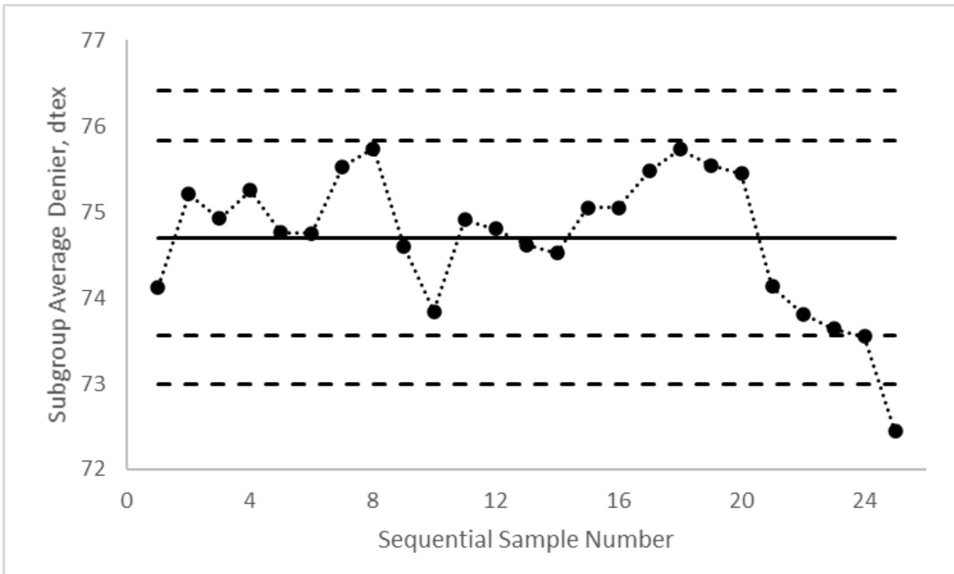

Figure 4 $\bar{X}$ chart for Example 1 (showing UCL, UWL, $\bar{\bar{X}}$, LWL, and LCL values)

The $\bar{X}$ chart is a moving window on the recent history of the data. As each scheduled sampling occurs, all the data points are indexed to the left, the oldest one is discarded, and the new $\bar{X}$ value is added to the right. Also $\bar{\bar{X}}$, UCL, and LCL (and UWL and LWL if used) are recalculated and added. The index is usually sample time, here, it is sample number, with #25 being the most recent.

To use the $\bar{X}$ chart to identify when an assignable cause occurred, one looks for events that have less than a 0.26% chance of occurring during periods in which there are no systematic changes in stationary process behavior. One such event is the violation of a control limit by sample 25 in this example. Another such improbable event is too long a run of data on one side of the $\bar{\bar{X}}$ line. Using Figure 21.4, in the 2nd to 8th sample, 7 data points are on one side of the X-bar-bar line. Using the binomial distribution, the probability of seven contiguous data on one side of the $\bar{\bar{X}}$ line is $\binom{8}{0} 0.5^0\ 0.5^7 = .007812\ldots$ and indicates 99.22% confidence that an assignable cause occurred. Some other conventional rules are listed in Table 2.

Note, the assignable cause may not have occurred at the same time when the pattern indicates something real has happened. If the most recent several points had not been included, the $\bar{\bar{X}}$

value would be greater, and there would not be 7-in-a-row on one side of the $\bar{\bar{X}}$ line. When the most recent data shifted the $\bar{\bar{X}}$ line, then the early data points were also able to flag that something has happened.

Even though this approach is grounded in statistical theory, subjectivity is present in both the sampling program and rules for action. Experience is needed to set the sampling frequency and the subgroup size to responsively observe an assignable cause. Similarly, you must use a time "window" or data horizon of sufficient length that long-term process drifts may be observed with confidence. Further, you must balance the cost of increasing both sampling frequency and subgroup size against the economic benefit of improved control. Conventionally, 100 samples are represented on an X-Bar chart. In Example 1, a subsample size of n = 4 and a horizon of 25 samplings, as illustrated in Table 1, are used.

The action rules in Table 2 generally reflect 99.74% confidence, but some rules reflect a 98% confidence limit based on individual experiences.

Table 2 Common Rules to Identify an Assignable Cause Occurrence

1. A single violation of the UCL or LCL indicates a change.
2. A run of 6 or 7 consecutive data on one side of $\bar{\bar{X}}$, $\bar{S}$ or $\bar{R}$ indicates a change.
3. A run of 6 or 7 consecutive data with a consistent upward or downward trend indicates a trend.
4. 2 out of 3 consecutive data violating a warning ($\pm 2\sigma$) limit indicate a change.
5. 4 or more crossings of $\bar{\bar{X}}$, $\bar{S}$ or $\bar{R}$ out of 15 consecutive data indicate cycling.
6. 4 out of 5 consecutive data outside a $\pm\, 1\sigma$ line indicate a change.
7. 10 out of 11 consecutive data on one side of $\bar{\bar{X}}$, $\bar{S}$ or $\bar{R}$ indicate a change.
8. 12 out of 14 consecutive data on one side of $\bar{\bar{X}}$, $\bar{S}$ or $\bar{R}$ indicate a change.

Example 2 Develop an S chart for the data in Table 1.

Figure 5 illustrates the S chart for the yarn denier data of Table 1. In this figure, subgroup standard deviations are plotted in chronological order. $\bar{S}$ is the average of the sample standard deviations

$$\bar{S} = \frac{1}{k}\sum_{i=1}^{k} S_i \qquad (3)$$

Which is not the σ of the entire data set. The number of subgroups is k; in this example, k = 25.

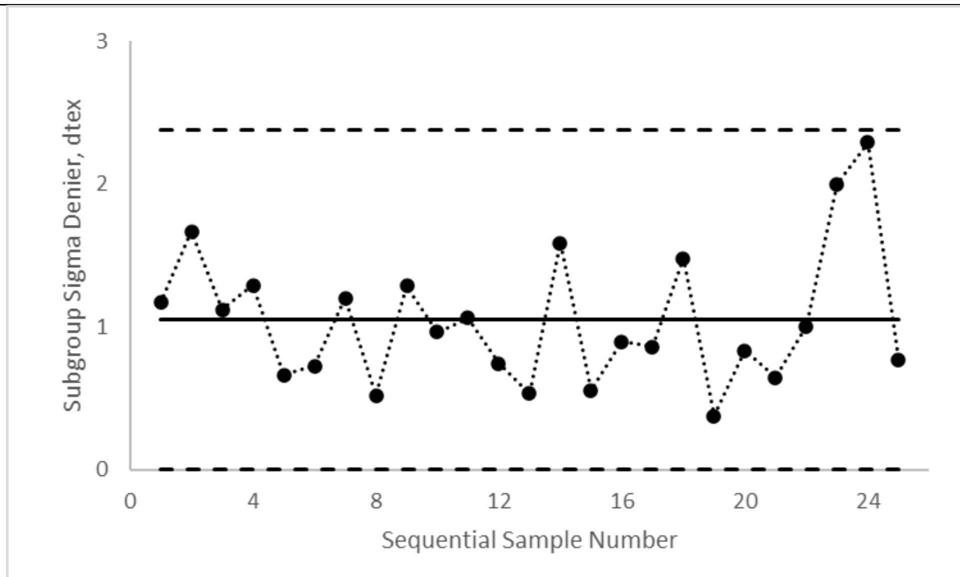

Figure 5 *S* chart for Example 2

There are several ways to set the UCL and LCL. Since S^2 is distributed as chi-squared and has $n - 1 = 3$ degrees of freedom, the 99.8% confidence interval for *S* can be determined by

$$\sigma\sqrt{\frac{\chi^2_{3,0.001}}{3}} < S < \sigma\sqrt{\frac{\chi^2_{3,0.999}}{3}} \qquad (4)$$

This, again, presumes that the true standard deviation could be known. If using the average standard deviation of all subgroups then

$$LCL = B_3\bar{S} < S < B_4\bar{S} = UCL \qquad (5)$$

There are many sources for the B-values.

Use of the *S* chart parallels that of the $\bar{X}$ chart. Violations of the UCL indicate a significant increase in variability due to an assignable cause. Violations of the LCL indicate a significant reduction in variability. If beneficial, the cause should be found and intentionally instituted. However, excursions below the LCL may indicate instrument failure, simply an “apparent” process improvement, and the necessity for instrument repair.