

Enclosure B - Statistical Process Control for process control of microbiological levels

Executive Summary

As mentioned in Appendix D, the purpose of sampling and thus measuring something is to make some type of inference or evaluation of some population property. In Appendix D a short discussion of the effects of sample and measurement errors on evaluation was given. This appendix discusses in more detail one general sampling application: Statistical Process control (SPC), which is a type of quality control (QC) sampling used to control a process. Statistical Process Control (SPC) has been used in the manufacturing setting for many years for controlling the quality of produced items. Recently its applications have been extended to microbiological output for use, successfully we believe, in ensuring the safety of processed foods or other items that might present a hazard to consumers of the product. In addition SPC can be used by laboratories in helping ensure that the measurement process is “in control” – that is, that the measurement deviations from the true value, over time can be considered as being independent of time and within specifications that might have been determined from collaborative or inter-laboratory studies. This can be accomplished using split samples or check samples, and occasionally comparison with another more authoritative method.

There are two features that characterize SPC and differentiate it from other types of sampling, namely acceptance and survey sampling. These types of sampling involve taking samples from a well defined population of units, specifying an upper bound to the number of samples that would be taken, and, from the results obtained from these samples, making a decision or an evaluation about the population that was sampled. As opposed to these types of sampling, SPC sampling does not involve specifying a fixed upper bound number of samples or necessarily identifying clearly a population of units. Rather, SPC involves sequential sampling over time, accompanied by a set of rules or criteria that are used to make decisions or evaluate, not so much a well defined set of units, but rather the process that is creating the units. The second feature that characterizes SPC is that the underlying values of parameters that are used to construct the rules are derived from results from sample units that were created by the process itself. In order for this to be done in meaningful way, the parameter values should be reflecting the process when it is in control. Thus, SPC as a subject matter, involves methodology for judging this – when can it be considered that a process is in control so that the rules that are to be used for evaluating whether or not the process is or remains in control are valid. SPC involves evaluation of the process and not specifically whether produced units or obtained measurements are within some pre-defined specifications.

Laboratories can use QC procedures for assuring that the measured results being produced are within specifications that are defined by repeatability or reproducibility parameters. SPC though offers a degree of flexibility that takes into account the actual system or process of measurement, insofar as the criteria for evaluation are not derived from outside the process but are derived from within the process itself. The full application of SPC entails a continuous examination of the data with the purpose of not

only just judging whether or not a process is not producing as it should, but also that the process has the capability of producing better than it was initially thought it should, by helping identify areas of potential improvement. That is, evaluative criteria can change, taking into account the potential capability of the process.

Thus, this document has a twofold purpose. The first purpose, the primary one and the reason for the document, is to present “performance standards” regarding the application of SPC for microbiological output of a process. However, a second purpose is to provide a simple introductory paper that could serve as a beginning point for learning about SPC and its application for microbiological data. Thus examples in Appendix G1 are given that demonstrate principles that are rooted in the performance standards.

The sampling work group is recommending the following “performance standards” with respect to implementing SPC for microbiological data. The performance standards are not meant to prescribe procedures or criteria that should be used for evaluating processes; rather they are meant to provide guidance and a methodology to be used for developing a SPC sampling plan. Following the performance standards are discussions of them, a conclusion section, and specific examples (seven in all) given as Appendices of the report (BPMM report Appendix F.1). The examples include SPC for qualitative or attribute data, including binomial Poisson –like, and negative binomial distributions; continuous variable data of high levels of generic E. coli; and an example which uses SPC for tracking the occurrence of infrequent events such as the finding of E. coli O157:H7 on samples. Hopefully these examples will serve as useful material.

Performance standards:

1. Charts of plots of the output data over time are not only valuable for verifying calculations and having a visual picture of the variation exhibited by the process output, but also it is an integral tool to be used for identifying sources of unexpected variation in output leading to their elimination. Thus charting is a necessary tool needed to gain the full benefit of doing SPC.
2. Results to be plotted in a control chart, when the process is under control, used for statistical process control should be normal or nearly normally distributed. In cases where this is not true and an alternative known distribution cannot be assumed such as a Poisson, binomial, or negative binomial distributions¹, transformations such as the log transformation for microbiological counts, arcsine transformations for binomial data, or a square root transformation for data distributed nearly as a Poisson distribution should be considered.
3. During some “initial” period of time, it is assumed that the process is operating in a relatively stable manner – or is in control. During this period the distribution of the measurements should be estimated and rules for evaluating the process should be formulated. The statistical “rule of thumb” of using about 20-30 results or more for computing means and standard deviations or other summary statistics

needed to estimate the distribution of results and construct control limits is a recommended and desirable goal.

4. Rules for evaluating process control should be set with aids assessing the two types of errors: Type 1, declaring the process out of control when it is not, and Type 2, not declaring a process out of control when it is. Typically there are two measures, depending upon the nature of the rule, that are used for assessing these errors: 1) the probabilities of the two types of errors at a given time (referred to as α - and β - probabilities, respectively); and 2) the average run length (ARL) – the expected number of samples before an out of control signal (one of the rules being not met) is seen.
5. When a process is thought to be “in control,” the limits for assessing individual results are set at some distance from the average, expressed as standard deviation units from the mean or process target value. The default distance is 3 standard deviationsⁱⁱ. Limits other than these should be implemented when taking into consideration economic and public health costs of incorrect decisions regarding whether the process is in control. When developing rules, the α -probability (for the Type 1 error) should be kept low, for example, below 1%.
6. There are numerous run/trend rules that can be used, such as runs test, moving averages and CUSUMS, for detecting shifts in the process mean; and rules for detecting shifts in the process variation or other auto-correlated patterns that could be due to systematic source of variation. The use of any of these may depend upon particular expected conditions that arise when the process is out of control, and the sensitivity desired for detecting such conditions. In assessing the use of these rules, one should consider the ARL. It is recommended, when the process is in control, that an ARL should exceed 100 (corresponding to a less than a 1% α - error).
7. Specification Limits are not Statistical Process Control limits; specifications are either customer, engineering, or regulatory related. Statistical Process Control limits are process related. Specification limits should not be placed on a control chart insofar as these might be considered as process goals thus influencing the efficacy of SPC procedures for ensuring a controlled process, and thereby undermining the safety of the product.

Performance standard 1 – the necessity of charting

Statistical process control (SPC) involves two aspects: use output data from a process to establish an expected distribution of values of some variable which is used for judging the control-status of a process when the process is (thought to be) in control; and a set of rules or criteria for which (future) output values from the process must satisfy in order not to declare, or declare presumptively, the process is out of control. In establishing the distribution to be used for determining the control status of the process, besides the output data, various other, regulative, type judgments are used that can affect the assumed distribution and the rules that are used for evaluating the process.

One feature that is included in the SPC methodology is charting – plotting of output process data values that are used for evaluating the process versus time or sample number, and examining the charted or plotted data. A question might arise is: why is this charting necessary? The implication of the question is that it may not be necessary, particular so with today’s computer technology – all that is needed is to somehow feed the data into a computer program and the program would make the calculations, determine whether or not the rules were violated and thus provide the control-status of the process. Various answers to this question can be given. One answer could be that charting provides a confirmation of the calculations; however, with today’s computer technology there are many other ways of ensuring that the calculations are correct to the extent that if there was a noted discrepancy between the plotted data and the computed results it more likely would be due to an error in plotting rather than in calculations. Thus, the answer to the question involving “looking” at a chart for the purposes of confirmation does not provide a good reason for the necessity of charting. Another answer might be based on psychology – the chart provides management with a visual picture of what is happening and this would give them a greater understanding of the process than what could be gained by examining sets of numbers and adherence of them to a set of rules. This answer by itself though would not provide a necessary reason for charting, at least not one in which a requirement of charting is recommended since there really would not appear to be a concrete gain from plotting.

However, this last answer is getting closer to the reason that compelled us to recommend, necessarily, charting, rather than just pointing out that charting is useful for the above stated reasons. The “seeing” of the chart can convey an understanding of the process that adherence to a set of rules cannot. Thus while the “looking” at charts can provide the confirmatory and psychological assurance, the “seeing” – meaning, a more in depth examination of the charted data - can provide additional information about certain aspects of the process that might have been unanticipated initially so that prior “rules” reflecting these aspects were not constructed. From “seeing” a chart, new insights might be gained that could show the inadequacy of the selected rules or could provide motivation for the development of new rules that lead to identifying unanticipated sources of error and an improvement of the process; on the other hand, however, it could lead to explorations that do not lead to improvements and thus could lead to an inefficient use of time and resources. Thus, to help prevent incorrect decisions statistical analysis (retrospectively) of data should be performed (See Appendix 2). The “look and see”

approach to charting is emphasized in SPC, notwithstanding possible pitfalls associated with this.

Performance standard 2 – The control distribution

Statistical Process Control, (SPC) has been used successfully to control quality and costs of manufactured products since the late 1920's. This statistical tracking system used for monitoring processes performance was developed by Dr. Walter Shewhartⁱⁱⁱ. He discovered that variation observed in manufacturing output was visually “different” from the variation that he would expect to see for similar type characteristics in nature for a stable system. Dr. Shewhart speculated that the variation that was not expected was due to processing errors by either labor or management. In other words, if the process was “under control,” the deviations from a mean value of statistical measurements that “track” some feature or output of the process would be distributed in a “random” looking fashion without any clear patterns, “unimodally” or at least displaying some degree of “regularity” or “stability” with very few outlier values. Further, it was assumed that the errors would be symmetrically, or nearly symmetrically, distributed around the mean value. In other words, normality, or near normality, is a natural distribution to assume when a process is under control since it is then assumed that the deviations are “caused” by many, inherently uncontrolled factors, each contributing only a small amount to the magnitude of the deviation. Historically then, in the manufacturing setting, rules or control limits for assessing a process to be out of control were set symmetrically with respect to the mean value – the assumption being that a result could be equally likely above as below the mean value. Thus, the distribution of the plotted values for the control chart was assumed to be normal and the operating characteristics of the rules - the probability of declaring the process out of control as a function of the true process mean - were evaluated assuming the underlying distribution of results is the normal distribution.

For microbiological data the above assumptions may not be true – rather, often (explicit examples are given in Appendix F.1) distributions seen will not be symmetric. If the non-symmetric distribution is known, then it is possible to use this distribution directly with the accompanying mathematical calculations to derive control limits with certain desirable operating characteristics. In such a situation parameters of these distributions can be estimated by maximum likelihood estimation or other statistical procedures and control plans can be determined directly using estimated distribution. However, often these specialized assumptions cannot be made, since with processing and measurement there would be expected unavoidable differences over time that could be caused by factors related to slight variations of equipment settings, environmental conditions and personnel that cannot easily be controlled or completely eliminated. For example, it might be assumed that under ideal conditions, the plate count distribution would be Poisson, with a parameter, λ - representing, in this case, the expected value. But value of this parameter may not be constant from day to day, or sample to sample, rather, λ itself would be a random variable, taking on possibly different values for different samples. Because of this (λ being a random variable), the total variation seen in the obtained results would not be expected to be equal to the expected variation of results seen from a Poisson distribution. The distribution of the results thus might be represented

well as a mixture of Poisson distributions. One such distribution is the negative binomial distribution, which has two parameters.

In general though, the expected distribution when the process is in control may not be known other than it most likely would not be symmetric. And for the classical SPC control procedures (as described below for Performance Standard 3), the limits are set using sample mean and standard deviation values for results on sample collected from a process assumed to be in control or nearly so, as if the distribution of these results were generated from a nearly normal distribution. If the distribution of results is not nearly symmetric, then transformations of the output variable, for example, taking the logarithm of microbial plate counts, may induce a more symmetric looking distribution. There is often another advantage of using the transformed variable: namely, the expected standard deviation would be less dependent on the expected mean value of the particular result. Thus, if plate counts were thought to be distributed as nearly lognormal, then a log transformation would make the distribution nearly normal and the variances of each transformed result would be nearly uniform for the data. Similarly if the data results were thought to Poisson-like distributed, a square root transformation of the results would make the results more symmetrical and make the variance more uniform (Appendix 3); for the binomial distribution, the arcsine transformation, $\sin^{-1}[(x/N)^{1/2}]$; and for the negative binomial, the inverse hyperbolic sine transformation, $N^{1/2}\sinh^{-1}[(x/N)^{1/2}]$ would make the distribution more symmetric and the variance more uniform (Johnson and Kotz, 1969).

While a normal distribution of the deviations from the mean value is not an absolute necessity for applying the control techniques discussed in this paper, historically the stated probabilities describing the operating characteristics of the control plan are computed assuming normal distributions and used for motivating decision rules. As a result of these considerations, performance standard 2 is recommended.

Performance standards 3 and 4 – Establishing the control distribution and rules for process evaluation

SPC is applied as follows:

- 1) During some “initial” period of time, it is presumed that the process is operating in a relatively stable manner, as described in the preceding paragraphs. This is a very important presumption and in actuality to reach this point when the process controls and parameter values are set, it may be needed an extended period of experimentation or trials. Whenever possible, independent validation of the presumption of process control should be made by other means, different from the statistical process control planning to be used, such as, for laboratory QC, the use of reference standards or cultures with known characteristics. If the distribution of results is expected to be nearly normal, then during this period statistical measurements should be distributed randomly around a mean value, μ with a standard deviation, σ . Values for these parameters are estimated during this time.

- 2) Over time, the statistical measurements are plotted on a graph, called a Shewhart chart (see Appendices for examples), showing the distribution of the statistical measurements. The Shewhart chart is basically the plot of the measured values versus sample number, starting with some sample labeled 1.
- 3) If the plotted statistical measurements do not meet any one of a set of criteria the process is considered to be “out of control.”

The criteria are chosen to reflect different manifestations of “out of control” of interest to the producer. Particular types of “out of control” signals are: a) “short term” non-systematic errors that might occur that result in an unacceptable product for a given day or lot; b) persistent errors that cause a systematic deviation from the pre-designated target value, μ ; and c) persistent errors that cause an increase in variability (σ) of process output.

Decision errors in regard to deciding whether or not a process was under control are similar to decision errors guarded against by the use of statistical procedures when testing two competing hypotheses in science. That is, a Type 1 error is made by deciding that the process is out of control when, in fact the process is in control and thus would not require adjusting; and a Type 2 error occurs when a process is not adjusted (actually is out of control) but it is decided that it is not out of control and the process is left as is. The probabilities of these errors are, respectively, referred to as α - and β - probabilities. Both of these errors could contribute to processing inefficiencies.

Processes can be affected by Type 1 and 2 errors because management and hourly workers often make adjustments that should not be made or fail to make adjustments that should be made in the attempt to “improve” the process output. The psychological forces that lead to changes or no-changes and thus errors influence the output of a process. A belief could develop, particularly the more one gains experience with the process, that ad-hoc adjustments based on one’s expert judgment would lead to a better process and output than just relying on pre-set rules as implied by charting and SPC. While in certain circumstances this may be true, often times it would not be so, and such a belief (of the advantages of following expert judgment) is not a reason to resist placing control charts on a process and using SPC. If nothing else, the use of control charts and SPC helps establish objective criteria for making adjustments (once the limits are established).

In other words, SPC and the use of rules for evaluating the process, determining α - or β -probabilities of the rules are not meant to eliminate expert judgment; rather these activities should be viewed as an aid for making judgments helping to prevent unwarranted actions that lead to a Type 1 or Type 2 error. “Out of control” signals can be considered “presumptive” regarding whether the process is out of control; and the examination of the data once plotted can lead to judgments of “an out of control process” that the charting “rules” have not reflected. Thus performance standards 3 and 4 we consider to be necessary for preventing the dominance of expert judgment in the evaluation of a process, but is not meant to eliminate it.

Consequently, SPC in its fullest sense involves preliminary analyses or testing of the process to a point where process parameters have been determined, such that it is believed that when the process is operating in accordance with the parameter specifications, the distribution of the measured output that is being used for evaluating process control would have the characteristics described above (random, stable, nearly symmetric or with some other designated distribution). In the developmental stage of the process this assumption may not be true. The SPC and plotting techniques described here can also be used in the developmental stages; however, in this situation the criteria for out of control may need to be changed.

Performance standard 5 – Control limits for individual values

Dr. Shewhart understood that in order for a control system to work effectively the rules or criteria used for determining the control-status of a process should meet a couple of requirements. These requirements include:

1. The α -probabilities (of incorrectly saying a process was out of control when it was not) must be low enough so to not unnecessarily create delays in processing (which could be costly) and fatigue workers and management from looking for causes of variation that do not exist;
2. The criteria must be robust enough so that a number of probability distributions can be accommodated by the procedures; and
3. The criteria should be simple and easily “seen” on a graph^{iv}.

As a consequence of the above considerations, Dr. Shewhart settled on his most well-known criterion that placed, what he termed, “Control Limits” a distance of three standard deviations from the process average; that is, control limits were set such that if a single measured value, (labeled often as X_i), was either greater than $\mu + 3\sigma$ or less than $\mu - 3\sigma$, with μ being the process average or intended process target then the process was to be presumed out of control. When the underlying distribution is normal, then the probability of exceeding one of the limits is 0.135%, so that the two-sided α -error is 0.27%. For most distributions expected for processes under control, the likelihood of seeing measured observations that do not satisfy these criteria is small^v thus satisfying requirements 1 and 2 above. Also, the third requirement is clearly met because the limits are just horizontal lines on the chart, and it can be easily seen if a plotted point is not between the two lines, indicating “out of control.” This criterion would “catch” a processing error that might not be systematic, and, when not met, would imply that there is some aspect of processing that might not be controlled.

