THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

ON THE SAMPLE SIZE FOR $\overline{X}$ AND X SHEWHART CONTROL CHARTS FOR A MEAN AND IMPLICATIONS FOR DESIGNING Q-CHARTS

by

Charles P. Quesenberry

Institution of Statistics Mimeograph Series No. 2220

May, 1992

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

ON THE SAMPLE SIZE PROBLEM FOR $\overline{X}$ AND X SHEWHART CONTROL CHARTS FOR A MEAN AND IMPLICATIONS FOR DESIGNING Q-CHARTS

Charles P. Quesenberry

North Carolina State University, Raleigh, NC 27695-8203

In this work we study the question of how much data is required in order for classical $\overline{X}$ and X charts with estimated control limits to perform essentially like charts with the true control limits. The results indicate that using estimated control limits results in charts on which the events that individual points exceed control limits are dependent events, and that one result of this dependence is to increase the rate of false alarms after short runs, even though the ARL is increased. The important implications of this to designing Q-charts is discussed.

Introduction

The usual approach to setting up an $\overline{X}$ chart entails collecting m samples of size n each and using these values to compute an estimate $\overline{\overline{X}}$ of the process mean μ and an estimate $\overline{S}/c_4$ of the process standard deviation. Many writers have recommended the number of samples m and the sample size n needed in order to estimate the parameters and control limits for classical Shewhart 3-sigma control charts. Most of the recommendations are that 20 to 30 samples of size 5, taken while the process is stable, are enough data to treat the estimated control limits as though they are the correct limits. These recommendations are apparently based upon empirical experience. We study here in some detail the question of the number m of samples of size $n = 5$ required to estimate the control limits for a classical 3-sigma $\overline{X}$ chart, and the total number of observations needed to estimate the control limits of an individual measurements chart using a moving range estimate of the process standard deviation σ. The results given here indicate that the usual recommendations are not sufficient to assure that the estimated control limits are close enough to the true limits to avoid having the charts sometimes give misleading results.

Quesenberry (1991) gave formulas that can be used to plot a class of Shewhart type control charts to control a process mean or standard derivation and the plotted statistics have exactly or approximately standard normal distributions for all cases when the process is itself a stable normal process. When the process parameters are "known" the charts to control the process mean are equivalent to the classical $\overline{X}$ or X charts for the cases when data are either in samples or by individual observations. Quesenberry (1991) also gave the first known Shewhart type charts that are distributionally correct for a stable process when the process parameters are unknown. These charts permit the valid charting of start-up processes and short runs. Of course, the charts for the cases when parameters are unknown do not compete with the charts for the case when parameters are known. We will always use known parameters when they are available. Now, in fact, the parameters will be known in practice only from having quite a lot of data taken while the process is running in control.

A question that arises in utilizing Q-charts is: How many data are required from a stable normal process for estimating the control limits for the "known" parameters case? Note that this is exactly the same problem as determining how many data are required to set up reliable classical $\overline{X}$ and X Shewhart charts. We will therefore discuss the problem in that familiar classical context. Our purpose in this work is to study the sample size determination issue in some detail, because it is crucially important in using Q-charts. We need to know at what point it is safe to switch from the Q-charts with known in-control normal distributions of the plotted statistics to charts that assume that the estimated limits are the correct limits.

The Classical $\overline{X}$ Chart

Suppose that X_{ij} for $i = 1, \cdots, m$ and $j = 1, \cdots, n$ represents m samples of size n each of past data that are from a stable process with a $N(\mu, \sigma^2)$ distribution. Then the sample means

$$\overline{X}_i = \frac{1}{n} \sum_{j=1}^{n} X_{ij}, \; i = 1, \cdots, m$$

can be plotted on a classical 3-sigma Shewhart chart with control limits

$$\begin{aligned} UCL &= \mu + 3\frac{\sigma}{\sqrt{n}} \\ LCL &= \mu - 3\frac{\sigma}{\sqrt{n}} \end{aligned} \qquad (1)$$

Suppose the parameters μ and σ are unknown, but that m samples of size n each are available to estimate the control limits (1). The usual formulae for estimating UCL and LCL of (1) are

$$\overline{UCL} = \overline{\overline{X}} + 3\frac{\overline{S}}{c_4\sqrt{n}}, \; \overline{LCL} = \overline{\overline{X}} - 3\frac{\overline{S}}{c_4\sqrt{n}} \qquad (2)$$

for

$$\overline{\overline{X}} = \frac{1}{m}\sum_{i=1}^{m} \overline{X}_i \text{ and } \overline{S} = \frac{1}{m}\sum_{i=1}^{m} S_i$$

where S_i is the *ith* sample standard deviation, and c_4 satisfies $E(S_i) = c_4\sigma$, and is a function of n only.

We consider next the distribution of the estimate $\overline{UCL}$. Note that $\overline{\overline{X}}$ is a normal random variable and $\overline{S}$ is also approximately normal since it is a mean of m independent random variables. Since $\overline{UCL}$ is a linear combination of $\overline{\overline{X}}$ and $\overline{S}$ it is also normal, to good approximation. By direct evaluation we have the mean and standard deviation of $\overline{UCL}$

$$E(\overline{UCL}) = \mu + \frac{3\sigma}{\sqrt{n}} = UCL$$

$$SD(\overline{UCL}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{1}{m}\left[1 + \frac{9(1 - c_4^2)}{c_4^2}\right]}$$

The probability than an individually plotted mean $\overline{X}_i$ will exceed $\overline{UCL}$ for a stable process is given by

$$P(\overline{X}_i > \overline{UCL}) = P(\overline{X}_i - \overline{UCL} > 0)$$

$$P(\overline{X}_i - \overline{UCL} > 0) = 1 - \Phi\left[\frac{3}{\sqrt{1 + \left[1 + 9(1 - c_4^2)/c_4^2\right]/m}}\right]$$

Moreover, since $P(\overline{X}_i - \overline{LCL} < 0) = P(\overline{X}_i - \overline{UCL} > 0)$ and the events are mutually exclusive, the probability that a point will plot outside these estimated control limits is

$$P(\text{False Signal}) = 2\left\{1 - \Phi\left[\frac{3}{\sqrt{1 + \left[1 + 9(1 - c_4^2)c_4^2\right]/m}}\right]\right\} \quad (3)$$

Note the important point that this probability depends only upon m and n, and therefore for fixed m and n this probability is a constant. Also, the argument of the standard normal distribution function Φ is less than 3, so that the probability of a false signal always exceeds 0.00270. In particular, for $m = 30$ and $n = 5$ (values often recommended), we have $c_4 = 0.9400$ and

$$P(\text{False Signal}) = 2\{1 - \Phi(2.8964)\} = 0.00378$$

We can readily use the formula of equation (3) to evaluate the probability of a false signal for any m and n, however, as we shall demonstrate, these probabilities have only limited value is assessing the properties of charts using $\overline{LCL}$ and $\overline{UCL}$ as control limits. The reason for this is that these are the probabilities of the individual marginal events that points exceed control limits, but since the events are *not* independent these probabilities do not translate into average run lengths between signals. Therefore, to assess the performance of these, and many other charts, we advocate studying the properties of the *distribution of run lengths.*

Using ARL and SDRL to Characterize the Run Length Distribution for $\overline{X}$ Charts

Suppose m samples of size n from a stable $N(\mu, \sigma^2)$ process are used to compute $\overline{UCL}$ and $\overline{LCL}$ of (2). Further independent samples of size n are taken and plotted on a chart with these limits. Suppose that at some point between samples the mean shifts from the value μ to a new value $\mu + \delta\sigma$. Then let Y denote the *run length* which is the number of points plotted until a signal is given by having a point fall outside the control limits. Now, if the control limits were known, since the points $\overline{X}_i$ being plotted are independently and identically distributed for $i = 1, 2, \cdots$; the run length Y would be a geometric random variable with mean ARL, and standard deviation, SDRL, given by

$$\text{ARL} = \frac{1}{1-\beta}, \ \text{SDRL} = \frac{\sqrt{\beta}}{1-\beta} = \sqrt{\text{ARL}(\text{ARL}-1)} \tag{4}$$

for β = P(No Signal) = P(LCL $\leq \overline{X}_i \leq$ UCL). In particular, for an $\overline{X}$ chart with known 3-sigma limits, then $1 - \beta = 0.00270$, and so

$$\text{ARL} = 370.4, \ \text{SDRL} = 369.9. \tag{5}$$

Also, note that the formulas of (4) characterize the distribution of run length for the case when the events of signals on points are independent events with constant probabilities of signals on points. In particular, for this case of independent points note the simple formula relating SDRL to ARL. We suggest that a reasonable way to decide when m and n are large enough for $\overline{\text{LCL}}$ and $\overline{\text{UCL}}$ to be essentially equal to LCL and UCL is to determine where ARL and SDRL satisfy (4) for the case with parameters known. In particular, for the stable case with 3-sigma limits we should have ARL $\simeq$ SDRL = 370 from (5).

To study the question of total sample size in the common situation of n = 5 we have generated ARL's and SDRL's by simulation for a range of values of m and δ. Table 1 gives the results of this simulation. For each entry in Table 1 m samples of size n were generated, the control limits $\overline{\text{LCL}}$ and $\overline{\text{UCL}}$ computed, and then samples were generated from $N(\mu + \delta\sigma, \sigma^2)$ until a point outside ($\overline{\text{LCL}}$, $\overline{\text{UCL}}$) was found. Then this procedure was repeated many times. The exact values for the known parameters case is given in the last row (for m = ∞) for comparison. For each value of $|\delta|$ and m the first value given is the estimated ARL and the entry beside it in parentheses is its estimated standard error. The value below ARL is the estimated value of SDRL. For example, for m = 30 samples of size n = 5 the estimated ARL is 403, with a standard error of 3.9, and the SDRL is 563. These should be compared with the nominal values of approximately 370 when parameters are known.

Certain points should be noted in considering Table 1. First, a general effect of estimation of the control limits is to increase both ARL and SDRL from the nominal values given in the last row. Note that these increased values of ARL are the *opposite* of what would be expected from the evaluation of marginal probabilities in the last section for the stable case. This effect is due to the *dependence* of the events that is in turn due to using estimated limits. Actually, for a few cases when m is large--at least 75--this is not true, but we think these cases are due to sampling error of the simulation. Also, note that for smaller values of m there is a pronounced tendency for SDRL to exceed ARL. This is true even for large values of m for the $\delta = 0$ column, which represents the extremely important case of a stable or in-control process. However, from (3) we see that for a stable process the probability of a signal on a point is constant over points for fixed values of m and n. Therefore if the events of signals were independent, then run length would be a geometric random variable and from (4) we know that SDRL = $\sqrt{\text{ARL}(\text{ARL}-1)} <$ ARL. Thus the values of SDRL and ARL in the $\delta = 0$ column of Table 1 reflect the dependence of the plotted points. When SDRL is larger than ARL by a

Table 1. ARL and SDRL for m Samples of Size n = 5

m	$\lvert\delta\rvert$: 0	0.5	1.0	1.5	2.0
5	1347(40.5) 9173	202.32(29.9) 1888	9.91(0.58) 41.15	2.082(0.049) 3.11	1.159(0.012) 0.538
10	611(17.1) 3127	70.33(3.3) 208.4	6.68(0.31) 13.74	1.806(0.039) 1.728	1.097(0.008) 0.359
20	434(4.8) 754	46.68(1.40) 88.29	5.23(0.13) 5.89	1.658(0.026) 1.164	1.095(0.007) 0.333
30	403(3.9) 563	43.74(1.00) 63.29	5.03(0.12) 5.55	1.599(0.024) 1.060	1.085(0.007) 0.296
50	388.6(2.28) 484.3	39.21(0.81) 51.09	4.73(0.10) 4.49	1.625(0.024) 1.069	1.085(0.007) 0.306
75	384.2(3.2) 452.5	35.65(0.65) 41.27	4.47(0.089) 4.00	1.581(0.023) 1.013	1.084(0.006) 0.309
100	380.6(2.96) 419.2	34.18(0.61) 38.34	4.50(0.091) 4.06	1.584(0.023) 1.020	1.077(0.005) 0.292
200	367.9(2.71) 383.7	35.05(0.59) 37.34	4.56(0.093) 4.18	1.529(0.028) 0.900	1.078(0.009) 0.300
300	368.9(2.66) 380.9	34.46(0.56) 35.55	4.61(0.130) 4.14	1.544(0.030) 0.950	1.079(0.009) 0.288
500	372.4(2.33) 381.0	34.39(0.54) 34.44	4.48(0.125) 3.95	1.527(0.029) 0.904	1.069(0.009) 0.269
1000	365.5(3.6) 369.5	34.23(1.33) 32.63	4.60(0.125) 3.96	1.586(0.031) 0.970	1.077(0.009) 0.288
∞	370.4 369.9	33.42 32.92	4.50 3.97	1.57 0.95	1.076 0.286

significant amount there is dependence in the plotted points. Also, by considering the nature of the run length distribution we can also see that when SDRL exceeds ARL that we can expect that the run length distribution will have, in comparison with the nominal geometric distribution, a large number of quite short runs that are balanced by a few very long runs. This means that the net effect of the dependence caused by using estimated control limits $\overline{LCL}$ and $\overline{UCL}$ is that the run length distribution will, for a particular value of ARL, have more probability in the tails of the distribution, however, since this is a distribution on the positive integers, there will be an increased rate of very short runs between false signals. There will also be an increased number of extremely long runs between false signals, however, the ratio of the number of very short runs to the number of extremely long runs is itself large. This is clearly undesirable and we should remain constantly aware of this phenomenon in our charting activities. Studying Table 1 shows that there is a definite tendency of this type at the commonly recommended values of $m = 30$ and $n = 5$. At least a slight trend is still evident at $m = 100$, or even larger.

Readers will note that the considerations just discussed are relevant to a much wider class of situations that the one considered here. For example, for many types of control procedures that signal on points when a defined event occurs, if the signaling events are *dependent*, then we should consider more properties of the run length distribution than just its mean, that is, ARL. Considering just the run length distribution mean and not its standard deviation can be misleading when judging a control procedure.

Characterizing the Run Length Distribution for Individual Measurements Charts

Consider a sequence of observations $X_1, X_2, \cdots, X_r, \cdots$ that are assumed to be identically and independently distributed with a $N(\mu, \sigma^2)$ process distribution. Then Quesenberry (1991) gave formulas that can be used to chart a Q-chart starting from the third measurement, and for a stable process the plotted Q-statistics are independently distributed $N(0, 1)$ random variables. Now, if the process mean μ and standard deviation σ were known, then the control limits for a classical 3-sigma Shewhart chart for μ are given by

$$\begin{aligned} UCL &= \mu + 3\sigma \\ LCL &= \mu - 3\sigma \end{aligned} \qquad (6)$$

If the parameters μ and σ are unknown, but a sample $X_1, X_2, \cdots, X_N$ is available from the process, then the common practice is to estimate μ and σ by the sample mean $\overline{X}$ and the moving range $\overline{MR}$ given by

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i, \ \overline{MR} = \frac{1}{N-1}\sum_{i=2}^{N} |X_i - X_{i-1}|$$

Then $\overline{X}$ is an unbiased estimator of μ and

$$\tilde{\sigma} = \frac{\overline{MR}}{d_2} = \frac{\overline{MR}}{1.128} = 0.8865\,\overline{MR}$$

is an unbiased estimator of σ. Plugging these values into the formulas of (6), we obtain the estimates of UCL and LCL given by

$$\begin{aligned} \overline{UCL} &= \overline{X} + 3\tilde{\sigma} = \overline{X} + 2.6595\,\overline{MR} \\ \overline{LCL} &= \overline{X} - 3\tilde{\sigma} = \overline{X} - 2.6595\,\overline{MR} \end{aligned} \qquad (7)$$

We now consider how large the sample size N must be, i.e., how many measurements must be

Table 2. ARL and SDRL for Individual Measurements Charts

N	$\|\delta\|$: 0	1	2	3	4	5	6
30	33.06(181.4)	207(16.6)	11.98(0.63)	2.63(0.08)	1.29(0.013)	1.048(0.004)	1.0045(0.0011)
	38356	2782	40.11	4.88	0.79	0.241	0.0706
50	1106(43.5)	91.2(2.70)	8.70(0.27)	2.26(0.04)	1.25(0.017)	1.031(0.004)	1.0030(0.0012)
	8923	341	16.93	2.22	0.75	0.178	0.0547
75	663.4(8.74)	68.3(1.41)	8.15(0.17)	2.22(0.03)	1.21(0.012)	1.030(0.004)	1.0015(0.0009)
	1747.9	141.5	10.97	1.98	0.52	0.179	0.0387
100	574.0(6.09)	59.8(1.00)	7.16(0.14)	2.10(0.027)	1.20(0.012)	1.022(0.003)	1.0005(0.0005)
	1217.5	97.9	8.70	1.73	0.53	0.145	0.0224
200	458.7(3.40)	49.6(0.96)	6.70(0.12)	2.08(0.026)	1.196(0.011)	1.028(0.004)	1.0030(0.0012)
	679.8	61.0	7.34	1.63	0.498	0.167	0.00547
300	427.7(2.95)	47.2(0.84)	6.51(0.10)	2.04(0.024)	1.215(0.012)	1.0271(0.004)	1.0005(0.0005)
	561.4	53.3	6.46	1.51	0.516	0.162	0.0223
500	404.9(2.75)	46.4(0.80)	6.43(0.10)	2.03(0.023)	1.182(0.010)	1.026(0.004)	0.0020(0.0010)
	484.8	50.6	6.05	1.48	0.460	0.164	0.0447
1000	387.5(2.45)	45.5(0.73)	6.46(0.10)	2.01(0.023)	1.200(0.011)	1.024(0.003)	1.0020(0.0010)
	409.1	46.0	6.01	1.43	0.497	0.153	0.0447
2000	379.1(2.81)	44.6(0.33)	6.37(0.04)	2.018(0.014)	1.189(0.005)	1.023(0.0015)	1.0012(0.0003)
	397.8	46.1	5.77	1.428	0.474	0.156	0.0346
∞	370.4	44.0	6.30	2.00	1.49	1.023	1.0014
	369.9	43.5	5.78	1.41	0.48	0.153	0.0374

made in order for the estimated control limits of (7) to perform essentially like the "known" limits of (6). To study this issue we have performed a simulation study as follows and the results are summarized in Table 2. For each value of N and δ in Table 2 we first generated N observations from a $N(\mu, \sigma^2)$ distribution and compute $\overline{LCL}$ and $\overline{UCL}$ from (7). Then we generated a run of observations from a $N(\mu + \delta\sigma, \sigma^2)$ distribution until a value was found to fall outside these estimated control limits. This was repeated many times to generate a large sample from the run length distribution. The values given for selected values of N and $|\delta|$ are the run length sample mean, sample standard deviation, and standard errors of the sample mean. For example, the estimated ARL for $N = 100$, $\delta =$ is 574.0 with standard error of 6.09 and estimated SDRL of 1217.5. The last row ($N = \infty$) gives the values of ARL and SDRL for the case when μ and σ are known, which for this case with $\delta = 0$ are ARL $=$ 370.4 and SDRL $=$ 369.9. This shows a considerable degree of dependence among the events for this sample of $N = 100$.

The values in Table 2 for $|\delta| = 0$ show that the estimated control limits perform essentially like the limiting case only for very large sample size N of 2000 or so. By the same argument as given in the last section, there is a clear dependence of the plotted points even for N of 500 and 1000. This dependence will have the same type effect as that discussed above for $\overline{X}$ charts with estimated control limits. Namely, we can expect many more false alarms after short runs and a few more very long runs between false alarms. In order to guard against having this dependence due to estimating control limits unduly increase the number of false alarm short runs we recommend that N should be at least 300.

Summary: Implications for Implementing Short-Runs Q-Charts

To this point we have considered chart performance only in terms of the chart's ability to signal a change in the process mean by having one point plot outside the control limits. However, when charts are used as aids to stabilize a process, and bring it into control, it is especially important that we know the type of point pattern to expect when the process is stable. Only when we know the pattern to expect for a stable process can we recognize an anomalous pattern. Indeed, we expect to read more information from a chart than simply whether or not a point is outside control limits. Now, for the short-runs Q-charts this in-control distribution is known exactly and anomalous patterns are readily identified, however, when using estimated control limits care must be exercised in interpreting many point patterns, especially those such as tests 1, 2, 5, 6, 7 and 8 of Nelson (1984), which will in this circumstance be defined using *estimated zones.* Indeed, in these applications the capability to recognize unexpected anomalous patterns is important and this ability is destroyed by dependence of signaling events.

In this work we have given results that indicate that, even when the assumption of independent normal observations is valid, the usual method and commonly recommended sample sizes for estimating control limits result in dependence of the signaling events and affects chart performance. One result of this is an increase in the rate of false alarms after short runs, even though the ARL is increased from nominal value. These are important issues in designing either classical or Q-charts in an SPC program. Of course, there are other important design issues not addressed here. In particular, the sensitivities of the charts to detect various types of out-of-control conditions is an important issue. We will give some results on sensitivity of Q-charts to detect mean shifts in another paper.

REFERENCES

Nelson, L.S. (1984). The Shewhart Control Chart - Tests for Special Causes, *Journal of Quality Technology* 16(4), pp. 237-239.

Quesenberry, C. P. (1991). SPC Q Charts for Start-Up Processes and Short or Long Runs, *Journal of Quality Technology* 23(3), pp. 231-224.