TECHNICAL AIDS

by
Lloyd S. Nelson

Interpreting Shewhart $\bar{X}$ Control Charts

EIGHT tests for special causes were given in last October's Technical Aids (Nelson (1984)). These are applicable to X and $\bar{X}$ charts, and the ideas underlying some of them can also be applied to control charts for attributes. Here we will consider what phenomena will produce signals from each of the tests. With this knowledge it is easier to interpret the test signals. Recall that the chart is partitioned into six zones of equal width (one sigma) between the upper and lower control limits. These are labeled A, B, C, C, B, A with Zones C placed symmetrically about the centerline.

We will follow the convention that a point must be beyond, not on, the upper or lower control limit to be judged indicative of a special cause. The phrase "in Zone A or beyond" (cf. Test No. 5) can be taken to mean "beyond Zone B". Thus, a point lying exactly on the line dividing Zones A and B would not be considered to be "in Zone A or beyond". Although this may appear to be nit-picking, it is comfortable to have the rules unambiguous; then it is not necessary to decide on their interpretation as they are being used.

Imagine a process that is in statistical control. That is to say its behavior is well represented by a model consisting of balls (values) being randomly drawn from a bowl containing a large number of essentially normally distributed values (say extending out to $\pm 4\sigma$). In order to avoid depletion of the contents of the bowl we will assume either that there is an infinite supply of balls or, what amounts to the same thing, that we replace each ball after recording its value. Forming subgroups of any size and plotting the resulting means will provide an example of a control chart (actually a process) that is in statistical control.

Theoreticians don't like such phrases as "well represented" and "essentially normally distributed" in the second sentence of the preceding paragraph because they are not quantitative. Nevertheless achieving the goal of being exquisitely precise does not seem to be required here.

Looking first at the overall problem of interpreting a control chart we are led to consider what could be described as a natural data pattern. The normal distribution (model) of the data can be expected to yield values that tend to be close to the centerline. Fewer and still fewer values will occur in regions farther and still farther away from the centerline. It is expected that only one out of 741 points (1/0.00135) will lie above the upper control limit.

Data patterns that can be thought of as unnatural will deviate from the natural in any of a number of ways. Many kinds of unnatural patterns can be recognized at a glance. Others escape all but the most expert eye. Formal tests are available to bring the novice to the expert level and to make the expert consistent. Use of such tests puts the interpretation of control charts on a uniform and scientific basis.

Test 1 (One point beyond Zone A)

This is the test originally devised by Shewhart (1931) and called "Criterion I". In many applications it is the only test used. Ease of instruction in its use would seem to be the only reason for this. Test 1 will give a signal in response to either a shift in the position of the mean or an increase in the standard deviation of the process. The larger the change the sooner the signal will occur. If a range control chart is being kept and remains in control, an increase in variation can be ruled out. Test 1 will also respond to a single aberration in the proc-

KEY WORDS: Control Chart Interpretation, Mixture Detection, Shewhart Control Charts, Stratification Detection, Tests for Assignable Causes, Tests for Special Causes, $\bar{X}$ Control Charts

ess such as a mistake in calculation, an error in measurement, bad raw material, a breakdown of equipment, and so on.

Test 2 (Nine points in a row in Zone C or beyond)

It is common to supplement Test 1 with other tests for special causes in order to improve the sensitivity of the control chart and/or to provide tests for phenomena that do not affect Test 1. Test 2 is a run test that is sensitive to a shift in the process average. The selection of 9 points was made to make the chance of a false signal roughly equal to the chance of a false signal for Test 1 while not increasing the usual number of 7 (Grant and Leavenworth (1980)) by too much.

Test 3 (Six points in a row steadily increasing or decreasing)

This is a test for trend or drift in the process average. A small trend in the process average will be signaled by this test before Test 1 can react. Causes for a trend might be tool wear, depletion of chemical baths, deteriorating maintenance, improvement in skill, and so on.

Test 4 (Fourteen points in a row alternating up and down)

A systematic effect such as produced by two machines, spindles, operators or vendors used alternately will trigger this test. This phenomenon has been observed many times in measuring the time for sand to flow through an egg timer that is alternately upended (see, for example, Ott (1975)), thus giving the name "egg timer effect" to the pattern. The selection of the number of points was made on the basis of Monte Carlo trials that showed 14 gives roughly (about 0.004) the same chance of a false signal as Test 1.

Test 5 (Two out of three points in a row in Zone A or beyond)

A shift in the process average is the usual cause for a signal from this test although it is also somewhat sensitive to an increase in variation. It should be made clear that *any* two of the three points in Zone A or beyond provide a positive test. The point not counted can be anywhere; it can even be missing! Assuming a normal distribution the probability of getting a false signal from this test is $6(0.0228)(0.0228)(0.9772) = 0.003$. This is a little closer to the probability of a false signal for Test 1 than is the probability (0.001) associated with two successive points beyond a two-sigma limit.

Test 6 (Four out of five points in Zone B or beyond)

Similarly to Test 5 *any* four of five points in Zone B or beyond give a positive test. The point not counted can be anywhere. This is also sensitive to a shift in the process average.

Test 7 (Fifteen points in a row in Zones C, above and below the centerline)

At first the novice thinks that this configuration of points looks very good. However, "good looks" is dependent on the scaling used. So we are not concerned with "looks" but rather with patterns of points that imply nonrandomness. The critical value of 15 was found from Monte Carlo simulation and has a chance of a false signal roughly equal to that of Test 1. Test 7 will detect a stratification problem of the following type. For simplicity suppose that subgroups of size 2 are being used and that one member of each subgroup always comes from distribution A and the other member comes from distribution B. If distributions A and B have different averages the result is that control limits based on the range will be too wide. Thus it is not that the points are not varying enough, it is that the control limits are too wide. It should be mentioned that a mistake in arithmetic is at least as frequent a cause for a signal from this test as is stratification.

Test 8 (Eight points in a row on both sides of the centerline with none in Zones C)

If we imagine the two distributions described under Test 7 and suppose that the subgroups are made up of values from either one or the other (but not both), we have a mixture situation that Test 8 is designed to detect.

General Discussion

The probabilities quoted for getting false signals should not be considered to be very accurate for three reasons. First the normality assumption on which they are based is shaky. Second the (unknown extent of) lack of independence among the tests that are affected by the same phenomena alters the probability. Finally the fact that a given test involving runs of points (that is any of them except Test 1) is applied in an overlapping manner to the ongoing sequence of points, thereby affecting the chance of a false signal. Consequently these tests should be viewed as simply practical rules for action rather than tests having specific probabilities associated with them.

Which combination of these tests to use will depend on the circumstances. At the beginning of a training program Test 1 alone might be best. Subsequently Tests 1 through 4 comprise a good set that will react to many commonly occurring special causes. For an engineering study, adding Tests 5 and 6 will increase the sensitivity to changes in the process average. Finally, using Tests 7 and 8 will enable sampling problems (stratification and mixtures) to be detected. More discussion of this is given in the Western Electric *Statistical Quality Control Handbook* (1956).

This battery of eight tests is designed to reveal the more commonly occurring special causes. It is possible, though unlikely, for a process to be out of control yet not show any signals from these eight tests. For example, points might alternate up and down repeatedly except that an adjacent pair of points may move in the same direction beginning at every tenth subgroup. If this pattern recurs many times the existence of a special cause is clearly indicated even though Test 4 gives no signal. One should be alert to any patterns of points that might indicate the presence of special causes; but one should also be aware of the more common failing of imagining that something has happened when it hasn't.

Because the eye is easily fooled by the effect of scaling it is important that, for a given vertical distance between the LCL and the UCL (call this d), the chart neither be too stretched out nor too compacted. I recommend that a horizontal distance equal to d contain no fewer than 6 and no more than 14 points. The drawings in Figure 1 of Nelson (1984) have 12 points in this distance.

References

GRANT, E. L. and LEAVENWORTH, R. S. (1980). *Statistical Quality Control.* McGraw-Hill Book Company, New York, NY, p. 91.

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

OTT, E. R. (1975). *Process Quality Control.* McGraw-Hill Book Company, New York, NY, pp. 35-36.

SHEWHART, W. A. (1931). *Economic Control of Quality of Manufactured Product.* Reprinted in 1980 by the American Society for Quality Control, pp. 304-318.

WESTERN ELECTRIC (1956). *Statistical Quality Control Handbook.* American Telephone and Telegraph Company, Chicago, Ill.