TECHNOMETRICS, MAY 1989, VOL. 31, NO. 2

Robust Control Charts

David M. Rocke

Graduate School of Management
University of California
Davis, CA 95616

When a process is first submitted to statistical quality control, a standard procedure is to collect 20–40 subgroups of about five units each and then to construct control charts, such as $\overline{X}$ charts and R charts, with limits determined by the data. These control charts are then used to detect problems in control such as outliers or excess variability in subgroup means that may have a special cause. In this article, the robustness of these charting procedures is investigated. If the number of false alarms when the process is in control is held constant, the most sensitive procedures for detecting the out-of-control state are those that plot a subgroup statistic that is sensitive to outliers (e.g., mean or range) but determine the control limits in a resistant fashion. Ordinary charting procedures, such as the standard $\overline{X}$ and R charts, perform less well, and the worst performance is turned in by procedures in which the subgroup statistics are themselves resistant (e.g., median charts). To illustrate the point that robustness depends not only on resistance of the statistical tools to outliers but also on the purpose of the analysis, robust cumulative sum charts are briefly discussed. When outliers may be present, unlike $\overline{X}$ and R charts, an overall better performance is obtained when a resistant subgroup statistic like the trimmed mean is used.

KEY WORDS: Cumulative sum chart; Outlier resistance; R chart; X-bar chart.

1. INTRODUCTION

When an industrial process is first submitted to statistical quality control, an early step is usually the construction of control charts such as the $\overline{X}$ and R charts. A typical procedure might be to rationally select 20–40 subgroups of about five units each and calculate the mean and range of each subgroup. Then control limits can be established from the mean of the subgroup means and the mean of the ranges in such a way that the subgroup statistics would fall outside these limits only rarely if the process were in a state of statistical control. This procedure allows one to find unusual subgroups that may indicate intermittent problems and also to detect components of variation that are not reflected in the intragroup variation. For these procedures to have the greatest effectiveness, it is important that the charting procedures be as sensitive as possible to problems of control.

The presence of outliers tends to reduce the sensitivity of control-charting procedures because the control limits become stretched so that the detection of the outliers themselves becomes more difficult. Furthermore, these stretched control limits mean that other types of out-of-control behavior become more difficult to detect. The purpose of this article is to investigate this phenomenon and to provide techniques to ameliorate the difficulties caused by the possible presence of outliers in the data used to establish the control limits.

In the previous literature on the use of resistant statistics in control charts, some of the early works were motivated by ease of computation as well as robustness. An early work in which both motivations were apparent was that of Ferrell (1953), in which the author proposed that subgroup midranges and ranges be used with control limits determined by the median midrange and the median range. Although ease of computation was mentioned among the motivations for this procedure, he also pointed out that the wild shots could be more easily detected than with conventional methods. Along the same lines, Langenberg and Iglewicz (1986) proposed mean and range charts with control limits determined by the trimmed mean of the subgroup means and the trimmed mean of the ranges. Both of these methods should perform better than the standard methods in the presence of outliers.

Some previous work used resistant subgroup measures also. Clifford (1959) proposed median and midrange charts, mostly for ease of calculation. Iglewicz and Hoaglin (1987) and White and Schroeder (1987) independently proposed plotting subgroup boxplots, which involves the use of the subgroup median and the subgroup interquartile range (IQR). These procedures are less able to detect outliers than either conventional methods or those discussed in the previous paragraph. In addition, the inefficiency of the median makes the median chart less able to detect other departures from a state of control.

In this article, another way of computing resistant control limits is proposed. Instead of calculating the subgroup range (or standard deviation), a resistant measure of subgroup spread is used to determine the control limits. One example would be to plot $\overline{X}$ and R charts with limits determined by the mean of the subgroup IQR's. As will be seen, this procedure is at least as effective as any other proposed and has some advantages. If desired, one may use both techniques of producing resistant control limits by use of the trimmed/IQR or median/IQR charts.

There are several independent choices that can be made in selecting a control-charting procedure to use on a process being newly brought under statistical quality control. First, one must decide how subgroups are to be chosen given the known and suspected causes of variation in the process. Given the subgroup size selected, one must decide on a summary measure of location and spread for each subgroup, such as the subgroup mean or median and the subgroup range or IQR. Then these individual subgroup summary measures can, in turn, be summarized by using the mean, median, or trimmed mean to set the control limits.

This article does not address the first step in the process, although it may be the most important. Regarding the choice of subgroup measures and summary procedures, there is a large number of possibilities. After some preliminary trials in which many other combinations were tried, six specific types of control charts were selected for examination in this article, depending on the subgroup statistic charted and the summary statistic used to compute the control limits. These methods are summarized in Table 1, in which the method of nomenclature is X/Y, with X representing the method of summarizing the subgroup data to arrive at the control limits and Y representing the subgroup spread statistic used. The first method (mean/range) is the standard $\overline{X}$ and R chart with control limits determined from the mean of subgroup means and the mean of the subgroup ranges. The second type of chart (trimmed/range) consists of $\overline{X}$ and R charts with the 25% trimmed mean of the subgroup means and the 25% trimmed mean of the subgroup ranges used to produce the limits, as in Langenberg and Iglewicz (1986). The third type (median/range) similarly uses the median to summarize the subgroup statistics, as in Ferrell (1953). The fourth type (mean/IQR) uses the mean of the subgroup means and IQR's to set limits, and the fifth (trimmed/IQR) uses the 25% trimmed mean of the means and IQR's; both of these are newly proposed in this article. The final chart (median chart) consists of median and IQR charts with the mean of the subgroup medians and the mean of the subgroup IQR's used to set limits; this resembles part of the procedure used by Iglewicz and Hoaglin (1987) and White and Schroeder (1987).

There are many other control-charting procedures that have been proposed (e.g., Wadsworth, Stephens, and Godfrey 1986). The main variations in behavior in the presence of outliers are all represented in the six procedures chosen, however. For example, the standard deviation behaves essentially the same as the range with respect to efficiency and sensitivity to outliers, at least in the subgroup sizes considered here.

For the purpose of the control charts used in this article, the IQR is defined as $X_{(b)} - X_{(a)}$, where ( ) denotes order statistics, $a = [n/4] + 1$, and $b = n - a + 1$. This definition differs slightly from the usual definitions for the IQR. It is designed so that the IQR is free of the extreme order statistics for subgroup sizes as small as four and so that it is simple to implement. The IQR is, therefore, the difference of the second largest and second smallest observations when n is 4, 5, 6, or 7; it is the difference of the third largest and third smallest when n is 8, 9, 10, or 11.

In the remainder of the article, the six control-charting procedures are examined. After careful standardization, the procedures produce about the same number of false alarms when there is no special cause of variation and the errors are normally distributed. The median chart detects special causes of variation with less sensitivity than the other procedures because of the inefficiency of the median, a point also noted by Iglewicz and Hoaglin (1987). For detection of outlier-contaminated error distributions

Table 1. Control-Chart Procedures Considered in This Article

Short title	*Subgroup statistics*	*Control limits from*
Mean/range	Mean, range	Mean of means, ranges
Trimmed/range	Mean, range	25% trimmed mean of means, ranges
Median/range	Mean, range	Median of means, ranges
Mean/IQR	Mean, range	Mean of means, IQR's
Trimmed/IQR	Mean, range	25% trimmed mean of means, IQR's
Median chart	Median, IQR	Mean of medians, IQR's

and detection of special causes of variation in the presence of outliers, the best procedures plot a responsive statistic such as the mean or range but use robust summaries to determine the control limits. Of the procedures examined, a good combination of robustness and efficiency is found in the mean/IQR chart, which plots the mean and range and determines control limits from the mean of the subgroup means and the mean of the subgroup IQR's.

2. DETERMINATION OF CONTROL LIMITS

Suppose that one has N subgroups of n items each. Given a location statistic G_i and spread statistic S_i for each subgroup i and a method $T(\ )$ of summarizing the subgroup statistics $\mathbf{G} = (G_1, G_2, \ldots, G_N)$ and $\mathbf{S} = (S_1, S_2, \ldots, S_N)$ to form aggregate measures of location and spread, the standard method of constructing the upper control limit (UCL) [the lower control limit (LCL) is analogous] is to use

$$\mathrm{UCL} = T(\mathbf{G}) + 3T(\mathbf{S})\sigma_G/\mu_{T(\mathbf{S})}, \tag{1}$$

where σ_G is the standard deviation of the subgroup statistic G and $\mu_{T(\mathbf{S})}$ is the expected value of the summary of the spread statistic. Potentially, one could use different methods to summarize the location statistics and the spread statistics, but this is not done here. For example, if G is the subgroup mean, S is the subgroup range, and $T(\ )$ is the mean, then under normality (1) becomes

$$\begin{aligned}\mathrm{UCL} &= \overline{\overline{X}} + 3\overline{R}(\sigma/\sqrt{n})/(d_2\sigma) \\ &= \overline{\overline{X}} + 3\overline{R}/(d_2\sqrt{n}) \\ &= \overline{\overline{X}} + A_2\overline{R},\end{aligned} \tag{2}$$

using the standard notation d_2 and A_2 for these constants (Wadsworth et al. 1986).

This definition is sufficient to guarantee that the number of false alarms when the process is in control will be low (a few per thousand). This is not sufficient standardization for the purpose of comparing different control-charting procedures, however. Although this method allows for the efficiency of the subgroup location measure, it does not control for the efficiency of the subgroup spread measure or for the efficiency of different methods of summarizing to produce control limits. Thus a more delicate approach is required.

Suppose that one wishes to use a UCL of the form

$$\mathrm{UCL} = T(\mathbf{G}) + kT(\mathbf{S})\sigma_B/\mu_{T(\mathbf{S})}, \tag{3}$$

in which k is chosen so that the proportion of subgroups above the UCL is as close as possible to some specified figure α. In this article, I use $\alpha = .002$, corresponding to a standard normal deviate of 2.878; this figure was chosen because it yields approximately $k = 3$ for $\overline{X}$ charts when location and spread are estimated from small samples (see Table 2). The constant k is determined approximately for location charts using a method presented in the Appendix based on the delta method. The upper and lower control limits for spread charts plotting the subgroup statistic G_i were set by

$$\begin{aligned}\mathrm{UCL} &= T(\mathbf{S})\mu_G/\mu_{T(\mathbf{S})} + kT(\mathbf{S})\sigma_G/\mu_{T(\mathbf{S})} \\ \mathrm{LCL} &= \max(0,\ T(\mathbf{S})\mu_G/\mu_{T(\mathbf{S})} - kT(\mathbf{S})\sigma_G/\mu_{T(\mathbf{S})}),\end{aligned} \tag{4}$$

where k is set using an F approximation outlined in the Appendix.

Tables 2 and 3 show the resulting values of k for the various location and spread-charting procedures examined in this article. The expectations required were estimated by a Monte Carlo study using 100,000 subgroups. Note that use of a single multiplier such as $k = 3$ is not always a satisfactory approximation to the values needed for construction of fair comparisons among control-charting procedures. Of course, for practical use corrections of this kind may not be necessary, since all that may be required is that the number out of limits when the process is in control be small.

The major results in this article are based on a series of simulations of the process of initiating a control-charting procedure. In each simulation trial, 20 or 40 subgroups of five or eight each were generated from a normal distribution or a contaminated normal distribution in which 95% of the observations are standard normal and 5% are normal with a standard deviation of 5 (this distribution is abbreviated 5%5G). In some of the trials, a standard normal variate was added to each subgroup mean to represent a special cause of variation that should be detected by the control charting procedure.

Figures 1a and 1b show the results of a simulation in which 20 or 40 subgroups of five or eight each were generated from a standard normal distribution

Table 2. Multiplier k Required to Achieve an Error Rate of .004 for Location Charts

	$n = 5$		$n = 8$	
Control chart	$N = 20$	$N = 40$	$N = 20$	$N = 40$
Mean/range	3.032	2.955	3.000	2.937
Trimmed/range	3.068	2.973	3.023	2.949
Median/range	3.114	3.001	3.057	2.970
Mean/IQR	3.136	3.006	3.081	2.981
Trimmed/IQR	3.215	3.050	3.132	3.009
Median chart	3.135	3.005	3.081	2.982

NOTE: Each entry represents the multiplier that should result in a fraction .004 of the subgroups being out of limits when the process is in control. Results are based on a Monte Carlo determination of certain constants that used 100,000 subgroups. Details may be found in the Appendix.

Table 3. Multiplier k Required to Achieve an Error Rate of .004 for Spread Charts

Control chart	$n = 5$, $N = 20$	$n = 5$, $N = 40$	$n = 8$, $N = 20$	$n = 8$, $N = 40$
Mean/range	3.265	3.146	3.152	3.052
Trimmed/range	3.332	3.182	3.195	3.073
Median/range	3.413	3.232	3.260	3.107
Mean/IQR	3.644	3.308	3.507	3.231
Trimmed/IQR	3.935	3.442	3.695	3.320
Median chart	3.712	3.483	3.520	3.348

NOTE: See note to Table 2.

and control limits were calculated by several methods, with k's determined as just described. Then 20 or 40 new subgroups were generated and the number of subgroups out of limits was calculated. This procedure was repeated 500 or 1,000 times for a total of 20,000 subgroups. The plotted figures show the fraction of the 20,000 subgroups that fell outside the control limits. As can be seen, the fraction out of limits is very nearly constant for all methods and is quite near the nominal level of .004.

Another view of the performance of these procedures can be gotten by examining Figures 1c and 1d, which illustrate the same sort of simulation except that the number out of limits was calculated from the same data that were used to determine the limits themselves. The correlation between individual subgroup statistics and the control limits, which differs for each procedure, results in somewhat greater variability across methods when the process is in control. Nonetheless, there is an approximate correspondence between the various methods; the method proposed in this article (mean/IQR) performed essentially identically to the standard method. The viewpoint taken in the remainder of the article is that the process of initially applying control-chart methods is best modeled by evaluating the procedures on the same data as were used to determine the control limits. The results were essentially the same when evaluated on new data, however.

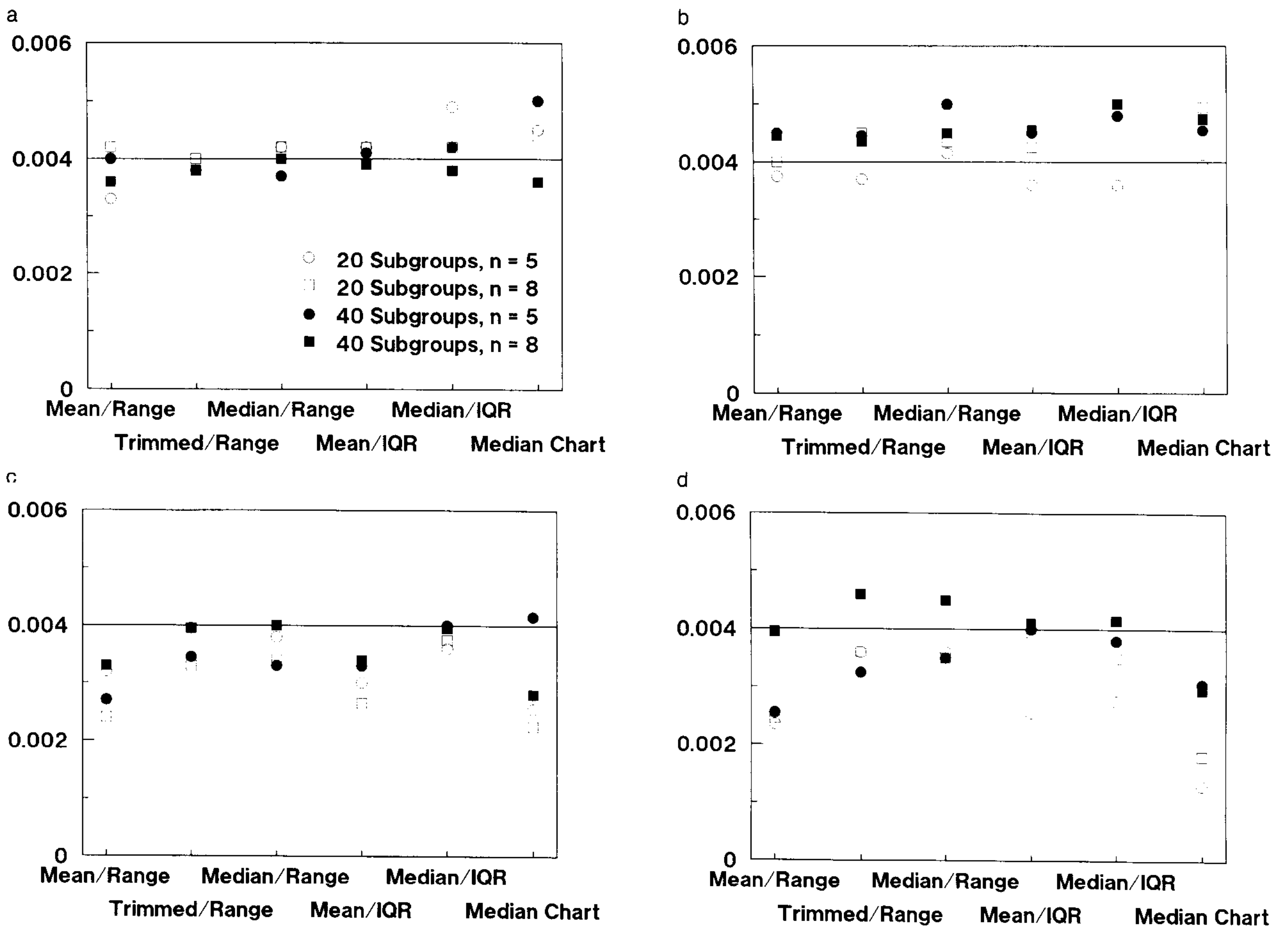


Figure 1. Fraction of Subgroup Statistics Out of Limits When the Process Is in Control, Based on 20,000 Subgroups. The data are iid standard normal. Panels a and c are for location charts, and b and d are for spread charts. In a and b, the statistic plotted is the fraction of subgroups out of limits from new data not used to determine the control limits; in c and d, the data are the same as those used to determine the limits. The horizontal line is the nominal fraction of subgroups expected to be out of limits.

3. ROBUSTNESS OF CONTROL-CHARTING PROCEDURES

3.1 Simple Location and Spread Charts

In the viewpoint of this article, there is an important difference between the terms *robustness* and *outlier resistance*. Outlier resistance is a neutral term that indicates that a particular statistic is not much changed by the presence of outliers—the median is the classic example of an outlier-resistant statistic. Robustness means that a procedure is still able to perform its intended purpose even if the assumptions under which it was developed are slightly incorrect. This requires one to know not only what the procedure is but for what purpose it is intended.

In the case of control charts used for the initial analysis of a process not previously under control, the purpose is to detect unspecified departures from the state of control so that underlying problems may be discovered and corrected. One such departure is the occurrence of outliers within the subgroups. These may be thought of as blunders that should be detected or they may be thought of as coming from a different population (which also should be detected). If one fixes the number of false alarms, then one procedure is more robust than another if it can detect outliers better and, more important, if it can better detect other problems, such as variations in the subgroup means, when outliers are present.

Figure 2a compares the abilities of the six location-charting procedures in the detection of a special cause in the form of variations in subgroup means. In the case of this simulation, this special cause was represented by an additional component of variation, in which the subgroup means were taken as standard normal. Specifically, the data point i from subgroup j was generated from the model $X_{ij} = \alpha_j + \varepsilon_{ij}$, where the α_j are iid standard normal and where ε_{ij} is either standard normal or 5%5G. The standard procedure (mean/range) performs best in that it detects the

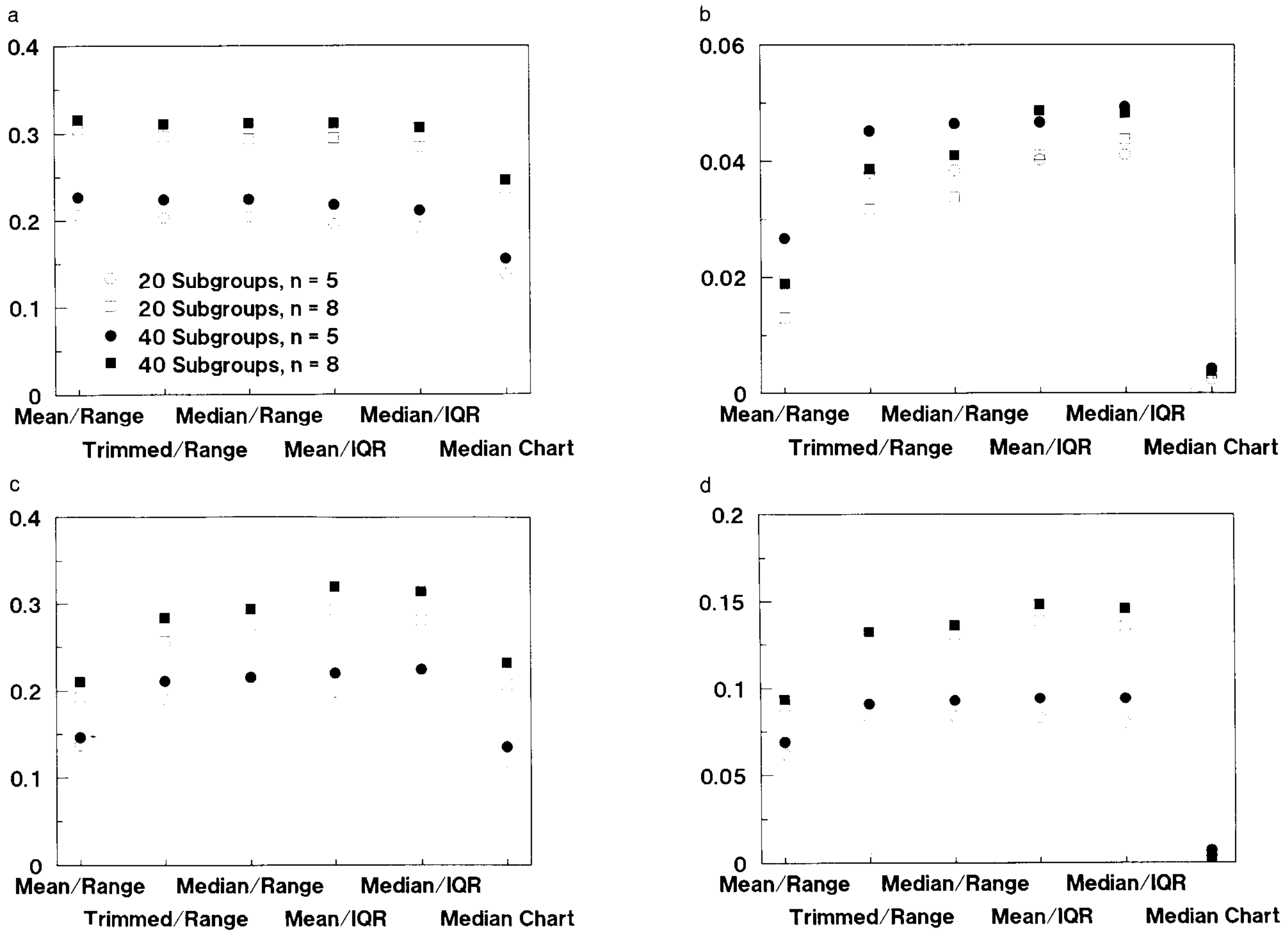


Figure 2. Fraction of Subgroup Location Statistics Out of Limits When Process Is Not in Control, Based on 20,000 Subgroups: (a) results for location charts when the within-subgroup errors are standard normal and there is a normally distributed special cause in the form of additional subgroup-to-subgroup variation, (b) location charts when the within-subgroup errors are a mixture of 95% standard normal and 5% normal with a standard deviation of 5(5%5G) but there is no special cause, (c) location charts when the within-subgroup errors are 5%5G and there is a normally distributed special cause, (d) spread charts when the within-subgroup errors are 5%5G.

most subgroups out of limits. The robust procedures are only slightly worse, but the median-chart method is distinctly inferior, a reflection of the inefficiency of the median. Similar results would be expected for other types of out-of-control behavior such as a sudden shift in the subgroup means.

Figure 2b compares the location-charting procedures on their ability to detect outliers. The error distribution was taken to be 5%5G, in which 95% of the data are standard normal and 5% are normal with mean 0 and standard deviation 5. The median-chart method essentially detects none of the outliers—the outlier resistance of the median has a deleterious effect on the ability of this charting method to find outliers in the subgroups. The standard mean/range method is considerably inferior to the remaining methods, and these four methods are comparable. The IQR methods apparently do a little better than the trimmed/range and median/range procedures, and this slight superiority is strengthened by the fact that the latter two methods produce somewhat more false alarms. Note that none of these procedures is very effective if considered simply as a method of detecting outliers in the subgroups. With subgroups of size 5 and 40 subgroups to the trial, one has 200 observations of which approximately 10 would have been "outliers"—that is, originating from the contaminating part of the distribution. Of these 10, about half would have been expected to be beyond the 3σ limits of the uncontaminated part of the distribution (if this were known). Ignoring, for the sake of this approximate calculation, the chance that two of these five outliers would have been in the same subgroup, an ideal outlier-detection procedure would have detected an outlier in about $5/40 = 12.5\%$ of the subgroups; none of the methods did even half this well. Nonetheless, it is unfair to judge a general-purpose procedure such as a control chart against a statistical test developed against a specific alternative; the control chart must function against a wide variety of problems.

Figure 2c illustrates an even more important difference that results from the use of resistant control limits. If there are outliers in the subgroups, then the resulting stretched control limits mean that other departures from the state of control can be detected less well. In this case, a component of variation between subgroups is poorly detected by either the mean/range charting method or the median chart. Of the remaining four, the mean/IQR method is marginally the most sensitive. Since it also has the fewest false alarms of the four robust charting methods, it is probably to be preferred.

Control charts are also used to find patterns in the subgroup statistics that do not necessarily cause excursions beyond the control limits but may nonetheless represent a real cause of out-of-control behavior. Since the proposed charts plot the same statistics as the conventional $\overline{X}$ and R charts and since the control limits will tend to be about the same (with normal errors) or tighter (when outliers are present), robust control charts will be just as effective as the standard procedures for detecting such patterns—more effective when outliers are present.

In assessing the sensitivity of spread charts, variations in the subgroup means are, of course, irrelevant. The occurrence of outliers in the subgroups has, in this case, the alternative interpretation of occasional disturbances in the spread. For example, when $n = 5$ and the errors are 5%5G, most of the subgroups (77%) consist entirely of standard normal observations, so that the standard deviation is 1. About 20% of the subgroups will have one outlier out of five, for an aggregate standard deviation of 2.41, and about 2% of the subgroups will have two or more outliers, with standard deviations of 3.26 (or more). Figure 2d shows the ability of the various spread-charting procedures to detect this kind of departure from control. Again, the median-chart method (which is, in this case, a chart of IQR's) cannot detect this kind of departure. The standard mean/range chart is inefficient compared to the four robust procedures, and the mean/IQR method is slightly to be preferred of the four. Two differences are obvious, however, between this situation and the analogous results for location charts given in Figure 2b. First, R charts are much more effective at detecting outliers than $\overline{X}$ charts. Second, the differences among the procedures are less marked, although in the same direction as for location charts. In Section 3.2, these properties are used to improve the behavior of location charts by preliminary rejection of groups before calculation of control limits.

A final point concerns the behavior of these procedures when the underlying distribution is nonnormal. If the error distribution is symmetric but long-tailed and the extreme observations are not to be considered outliers, then the robust procedures will reject too many subgroups when there is no special cause. It would be highly unusual to treat this situation as being in control, however, since definitionally the state of control precludes the frequent presence of extreme observations. The situation is more complex when the errors have a skew distribution. If this is known or is detected in an initial analysis of the data, one possibility is to transform the individual observations to normality by use of logarithms or square roots, for example. If this approach is inappropriate, then control limits can be established from the IQR using the relationship between the IQR and standard deviation for the appropriate class of distributions. In any case, the distribution of the mean will be approximately normal, even when the underlying distribution departs moderately from normality.

Overall, the robust charting methods perform much better than the others in the presence of outliers at little cost when the errors are normally distributed. This suggests that they should be routinely used in situations such as the one modeled here in which a process first comes under statistical quality control.

3.2 Two-Stage Procedures

One method that is sometimes used when analyzing data from a process being newly brought under control is to construct $\overline{X}$ and R charts, identify groups that seem out of control from statistical or engineering considerations, and compute control limits from only those subgroups that were not eliminated. In fact, this amounts to using a kind of robust estimator to summarize the subgroup location and spread values. The primary difference between this procedure and the median/range and trimmed/range methods is that the location chart and the spread chart are used together as a pair. Among other things, this allows the superior ability of R charts at detecting outliers to be used to improve the performance of $\overline{X}$ charts. On the other hand, since only the very most extreme subgroups are eliminated, this procedure alone may provide inadequate protection from contamination.

To compare this technique to the six simple charting procedures, another simulation was run. This was done for each of the six methods, but only for $n = 5$ and $N = 40$. In these simulations, initial location and spread control charts are constructed. Any subgroup that lies outside the control limits on either the location chart or the spread chart is eliminated, and the control limits are recomputed for both charts from the remaining subgroups. Figure 3 shows the results. Use of the two-stage procedure does not increase the false alarm rate markedly (Figs. 3a and 3e) and decreases the difference between the standard $\overline{X}$ and R charts and the robust estimators. This is quite logical, since in fact, this compound procedure is a robust method. The two-stage robust charts are still superior to the two-stage conventional charts both at detecting outliers (Figs. 3b and 3f) and at finding special causes of variation in the presence of outliers (Fig. 3d), although the margin has clearly shrunk.

Essentially, the use of this two-stage procedure creates a robust control chart that is nearly as good as the ones proposed in the article. The argument in favor of the mean/IQR method is its somewhat greater effectiveness and its somewhat greater simplicity. In addition, the mean/IQR method can cope with contamination that does not force the subgroup means to the 3σ level. As the examples will make clear, this can often be critical. Nonetheless, examination of the data and recomputation of control limits adds to the effectiveness of any of the procedures that one may choose.

4. EXAMPLES

Both examples use data from Wadsworth et al. (1986, pp. 207 and 230). The first concerns the melt index of an extrusion-grade polyethylene compound. As part of a study of the process, 20 subgroups of four each were taken. Figure 4a shows the $\overline{X}$ chart with limits determined from the mean or 25% trimmed mean of the range or the IQR. The center lines are omitted to avoid clutter. Although no points are out of limits on the $\overline{X}$ chart with mean/range limits (dotted lines), four are out if mean/IQR limits are used (dashed lines). Two of these were identified by Wadsworth et al. as problem points using several auxiliary rules (seven of the last eight on one side of the center line, etc.). Although these auxiliary rules increase the sensitivity of the charting process to different forms of adverse behavior, they also increase the false-alarm rate. Use of robust control limits could have detected the problems without the use of these auxiliary rules. Of course, if it is desired, these rules can be used just as easily with robust control limits as with standard ones, and with about the same effectiveness at about the same cost. The R chart of the melt-index example is given in Figure 4b; only the upper limits and the center lines are shown. One so easily sees such dramatic out-of-control behavior that it does not require a great deal of sensitivity to find it. Nonetheless, two points are out and two points are near the upper limit if the mean/IQR limits are used, twice the number found by the standard limits.

After application of the two-stage procedure described in Section 3.2, new control limits were obtained, as shown in Table 4. The mean/IQR method gives the tightest limits and thus the greatest sensitivity, with or without recomputation.

The second example concerns the production of cotton yarn. "Samples are taken from the spinning frames at eight positions daily. Four measurements of yarn 'count' are obtained for each position to form a subgroup. Three days' results were used" (Wadsworth et al., 1986, p. 229). Figures 5a and 5b show the results (values were multiplied by 10 to eliminate decimals). In both the $\overline{X}$ and R charts, the out-of-control behavior is readily apparent only when the robust control limits are used. In this example, the best results were obtained using the trimmed/IQR control limits (dot-dashed lines), but in the previ-

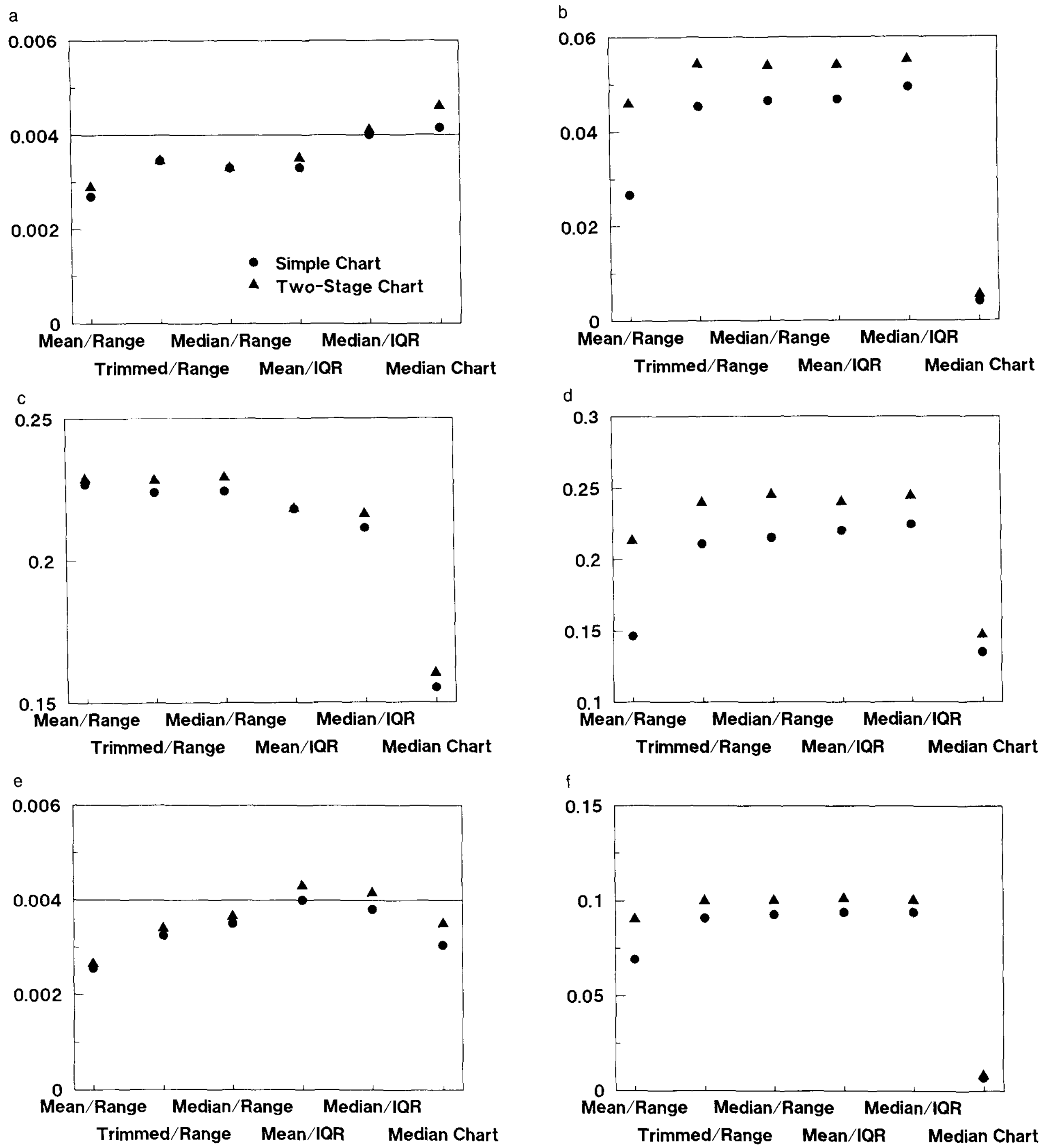
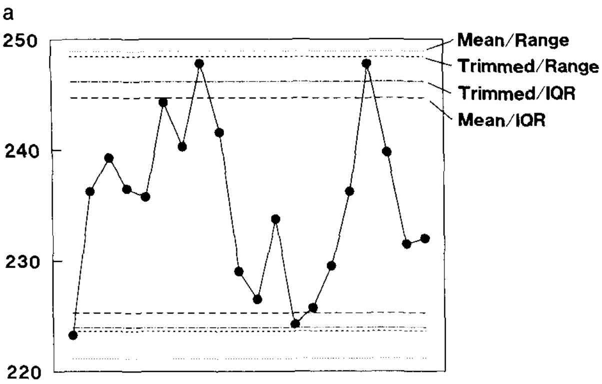


Figure 3. Fraction of Subgroup Location Statistics Out of Limits for Simple and Two-Stage Charts, Based on 20,000 Subgroups: (a) location charts when the within-subgroup errors are normally distributed and there is no special cause, (b) location charts when the within-subgroup errors are 5%5G and there is no special cause, (c) location charts when the within-subgroup errors are normally distributed and there is a normally distributed special cause in the form of additional subgroup-to-subgroup variation, (d) location charts when the within-subgroup errors are 5%5G and there is a normally distributed special cause in the form of additional subgroup-to-subgroup variation, (e) spread charts when the within-subgroup errors are normally distributed; (f) spread charts when the within-subgroup errors 5%5G. The data are iid standard normal with 40 subgroups of five units each. The horizontal line is the nominal fraction of subgroups expected to be out of limits.

ous example the best results were obtained using the mean/IQR (dashed lines). Thus in some cases it may be preferable to use both methods of producing robust control limits by using resistant subgroup summaries and summarizing them in a resistant fashion. Note that the two-stage method used with standard $\overline{X}$ and R charts would not duplicate the performance of the robust methods; since none of the subgroups



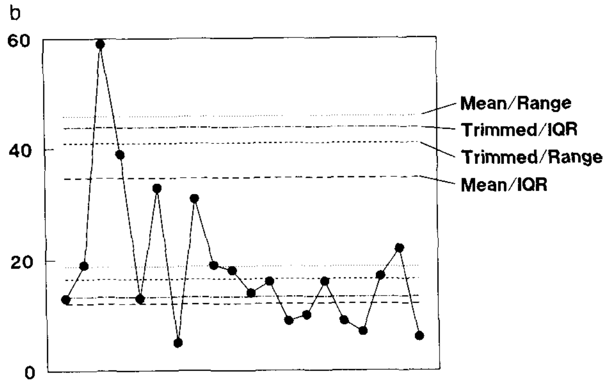


Figure 4. $\overline{X}$ and R Charts for Melt-Index Data: (a) subgroup means, (b) ranges.

is out of limits on either the standard $\overline{X}$ or R charts, the control limits would remain the same. As Table 5 shows, however, the two-stage method gives even tighter limits for the two IQR methods.

5. CUMULATIVE SUM CHARTS

The viewpoint taken in this section is that a cumulative sum (CUSUM) chart has a different purpose than the other control charts considered so far. $\overline{X}$ and R charts are general-purpose data-analytic devices for finding problems of control of many different kinds without the necessity of specifying in advance what kind of problem may occur. The CUSUM chart is designed specifically to detect a difference in the mean of the process from a specified value and range within which it may lie harmlessly. From this point of view, outliers in the subgroups should not cause a signal to occur since they do not directly represent a shift in the mean. If outliers may occur, then one can detect a shift in the mean more efficiently using a resistant subgroup summary rather than the mean.

The choice of subgroup location estimators for CUSUM charts is a subject too complex to deal with here. As an illustration, CUSUM charts using subgroup trimmed means will be developed and compared with standard methods. Trimmed means have the virtue of simplicity and are reasonably robust, but they are less efficient than more sophisticated robust estimators that might be used for this problem (Andrews et al. 1972). The variance of a trimmed mean can be accurately estimated using the Winsorized variance (Dixon and Tukey 1968; Tukey and McLaughlin 1963). Thus we will consider two procedures: The first is the usual procedure using subgroup means, and the second is the analogous method using 25% trimmed means.

Note that the situation considered here differs from that of Lucas and Crosier (1982). They supposed that the subgroup means had a nonnormal distribution around their grand mean, whereas in the present study the subgroup means are constant with a possibly nonnormal distribution of errors within subgroups. A more extensive study would be needed to examine the joint effects of these factors.

Let M be the nominal mean that the process should obey, and suppose that a preliminary sample has been taken so that $\sigma_{\overline{X}}$ can be estimated. If it is important to detect quickly a change in mean size of $2F$, then one may use the upper and lower CUSUM statistics

$$S_i = \max(0, S_{i-1} + (\overline{X}_i - M - F)/\sigma_{\overline{X}})$$
$$T_i = \min(0, T_{i-1} + (\overline{X}_i - M + F)/\sigma_{\overline{X}}). \quad (5)$$

For these definitions, the change to be detected has been specified in absolute terms rather than in standard-deviation units as is usually done. This is necessary to compare procedures using statistics with different variances. The two-sided CUSUM procedure signals when either $S_i \geq h$ or $T_i \leq -h$. The

Table 4. Control Limits for Melt-Index Data for Simple and Two-Stage Methods

	$\overline{X}$ LCL		$\overline{X}$ UCL		R UCL	
	Simple	*Two-stage*	*Simple*	*Two-stage*	*Simple*	*Two-stage*
Mean/range	221.15	222.50	248.92	247.13	45.81	40.63
Trimmed/range	223.61	223.61	248.39	248.39	40.93	40.93
Mean/IQR	225.34	223.94	244.74	244.92	34.85	37.69
Trimmed/IQR	223.96	222.24	246.14	245.63	43.77	46.16

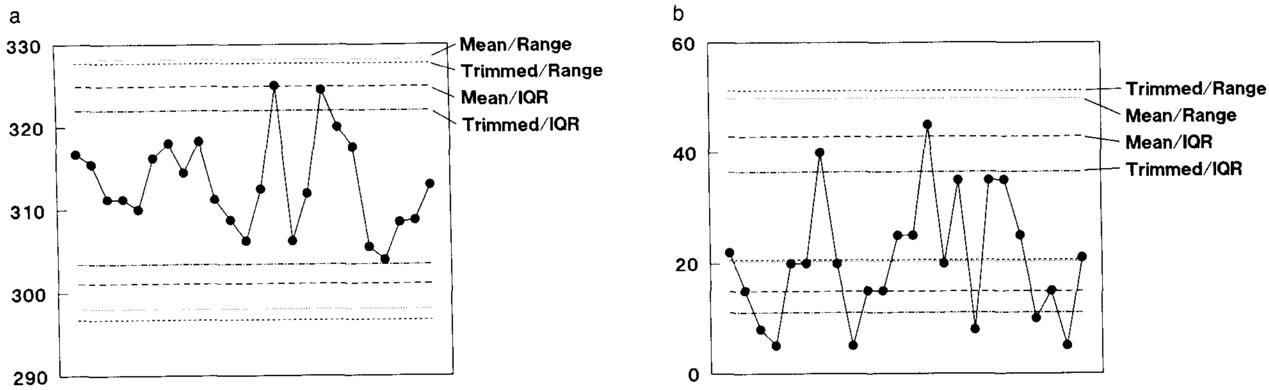


Figure 5. $\overline{X}$ and R Charts for Yarn-Count Data: (a) subgroup means, (b) ranges.

standard error of $\overline{X}$ can be estimated from the preliminary sample as $\hat{\sigma}_{\overline{X}} = s/\sqrt{n}$ or $\overline{R}/d_2$.

To generalize this procedure to the trimmed mean, we replace $\overline{X}$ by the 25% trimmed mean and estimate its standard error using the 25% Winsorized variance as follows:

$$\hat{\sigma}_{T25} = \sqrt{\mathrm{SS}_W/(g(g-1))}, \qquad (6)$$

where SS_W is the sum of squares around its mean of the revised data set in which the lowest 25% of the observations are replaced in value by the next lowest and the highest 25% are replaced by the next highest. The number of unaltered data is g. For example, if $n = 5$, then one version of the 25% trimmed mean has $g = 3$ and

$$\mathrm{SS}_W = 2(X_{(2)} - \overline{X}_W)^2 + (X_{(3)} - \overline{X}_W)^2 + 2(X_{(4)} - \overline{X}_W)^2, \qquad (7)$$

where

$$\overline{X}_W = (2X_{(2)} + X_{(3)} + 2X_{(4)})/5 \qquad (8)$$

and $\hat{\sigma}_{T25} = \mathrm{SS}_W/6^{1/2}$.

As usual, the performance of these two CUSUM procedures is evaluated by computing the average run length (ARL)—that is, the expected number of subgroups examined before the procedure signals. The ARL should be high when the process mean is at or near the nominal value and should be small otherwise. In the simulation reported in Table 6, the ARL was calculated for a standard CUSUM procedure ($F = .5\sigma/\sqrt{n}$ and $h = 3$) with standard normal data and with two selected contaminated normals for $n = 5$. This was done for the actual mean equal to the nominal mean and for one-half and one unit larger. The ARL values in the table are based on up to 10,000 simulation runs each, in which the exact number was chosen to give a standard error for the ARL that is 1% of its value. Since it was not easily possible to standardize these procedures so that the ARL under normality was the same, comparisons will employ the ratio between the ARL with a mean shift and the ARL without. Although this is a somewhat crude expedient that would need to be improved for serious work on robust CUSUM's, it will serve to make the necessary point.

With normal errors, the trimmed-mean CUSUM performs worse than the standard procedure. It has somewhat more false alarms than the standard method (ARL = 51.6 vs. 58.4) and responds less quickly to changes in the process mean. For example, when the mean shift is actually 1.0, the ARL ratio is 24.4 for the standard method; that is, the ARL in the absence of the shift is 24.4 times that when it is present. For the trimmed-mean CUSUM, this ratio is slightly worse at 20.4. Essentially, this poorer performance is due to the lesser efficiency of the trimmed mean of five data points as a location estimator. Use of a more efficient robust estimator could make the performance difference even less.

The slight superiority of the mean CUSUM method under normality is counterbalanced by a serious de-

Table 5. Control Limits for Yarn-Count Data for Simple and Two-Stage Methods

	$\overline{X}$ LCL		$\overline{X}$ UCL		R UCL	
	Simple	Two-stage	Simple	Two-stage	Simple	Two-stage
Mean/range	298.06	—	328.24	—	49.78	—
Trimmed/range	296.76	—	327.74	—	51.17	—
Mean/IQR	301.25	302.03	325.04	323.23	42.73	38.10
Trimmed/IQR	303.51	303.42	321.99	320.75	36.47	34.19

Table 6. Average Run Lengths for Standard and Trimmed-Mean CUSUM Charts

Distribution	Mean CUSUM, mean shift			Trimmed-mean CUSUM, mean shift		
	0	*.5*	*1.0*	*0*	*.5*	*1.0*
Normal	58.4	5.5 (10.7)	2.4 (24.4)	51.6	5.8 (9.0)	2.5 (20.4)
5%5G	22.4	6.4 (3.5)	2.9 (7.6)	45.8	6.2 (7.4)	2.8 (16.7)
10%10G	10.4	7.4 (1.4)	4.5 (2.3)	43.2	8.1 (5.4)	3.6 (12.0)

NOTE: Values are based on a simulation of up to 10,000 subgroups; the uncertainty in each figure is in the third digit. Parenthesized figures are ratios between the ARL when the mean is on target and the ARL when it is off target.

terioration in the presence of outliers. For example, with 5%5G errors, the ARL ratio of the mean is only 7.6 compared to 16.7 for the trimmed-mean method. Thus the use of resistant subgroup statistics can make a CUSUM procedure more robust. If it is desired for the CUSUM procedure to detect differences in the mean level even in the presence of outliers, and if these outliers are at all likely to occur, then the trimmed-mean CUSUM will have a better overall performance.

6. CONCLUSIONS

This article has two main points. First, as a general matter, robustness of a procedure to outliers or other violations of the assumptions may or may not require the use of resistant statistics. It is necessary to adapt the ideas of robustness to particular applications with sensitivity towards the purposes of the procedure. The opposite attitude of automatically advocating use of resistant statistics has perhaps led to some skepticism about the importance and validity of robust methods among practicing statisticians. Yet the use of these methods has been shown to be very important in practice (e.g., Rocke, Downs, and Rocke 1982; Stigler 1977).

Second, for the particular circumstances of generating control charts to help bring under control a process that has not previously been well controlled, certain design considerations became clear. Subgroup statistics plotted should be sensitive to outliers so that their existence may be detected; control limits, however, should be calculated robustly. There are several methods for insuring this robustness—the methods of Ferrell (1953) and Langenberg and Iglewicz (1986) and the IQR procedure proposed in this article. Of these, perhaps the last showed itself to be slightly superior, but any of them performed better overall than either the standard procedure or the fully resistant median chart.

Robust variation measures like the IQR and those proposed in Rocke (1983) and Iglewicz (1983) have many more applications in quality control and improvement than there is space here to detail. Applications such as interlaboratory studies (Rocke 1983), process capability, components of variation (Fellner 1986; Rocke 1983), statistical tolerance, and robust parameter design (Box and Meyer 1986; Taguchi 1986) could all benefit from considering the effects of departures from the assumptions. Work on these problems will arm the quality-control professional with tools that are usable under a broader range of circumstances than those in use today.

ACKNOWLEDGMENTS

Part of the work for this article was done while I was visiting at Imperial College, London. I am grateful for helpful suggestions from the referees, the associate editor, and Boris Iglewicz.

APPENDIX: CALCULATION OF CONTROL LIMITS

This appendix explains the method used to set control limits for the charts considered in this article. Suppose that one wishes to use a UCL of the form

$$\mathrm{UCL} = T(\mathbf{G}) + kT(\mathbf{S})\sigma_G/\mu_{T(\mathbf{S})}, \tag{A.1}$$

where k is chosen so that the proportion of subgroups above the UCL is as close as possible to some specified figure α. As usual, assume the normality of the subgroup statistic G_i and the central summary $T(\mathbf{G})$. Then

$$\begin{aligned} &\mathrm{PR}(G_i > \mathrm{UCL}) \\ &\quad = \Pr(G_i > T(\mathbf{G}) + kT(\mathbf{S})\sigma_G/\mu_{T(\mathbf{S})}) \\ &\quad = \Pr((G_i - T(\mathbf{G}))/\tilde{\sigma} > kT(\mathbf{S})\sigma_G/(\mu_{T(\mathbf{S})}\tilde{\sigma})), \end{aligned} \tag{A.2}$$

where $\tilde{\sigma}^2 = V(G_i - T(\mathbf{G})) = \sigma_G^2 + \sigma_{T(\mathbf{G})}^2$. Here we ignore any correlation between G and $T(\mathbf{G})$ either as an approximation or in the view that we wish to control the number of rejections from later data.

From (A.2) we may achieve a more accurate control limit by treating $(G_i - T(\mathbf{G}))/\tilde{\sigma}$ as standard normal, since the variability of $T(\mathbf{G})$ has been taken into account. This would consist of setting k so that the expectation of the right side of the last inequality in (A.2) is $\Phi^{-1}(1 - \alpha)$; that is, $k = \mathbf{\Phi}^{-1}(1 - \alpha)(1 + \sigma_{T(\mathbf{G})}^2/\sigma_G^2)^{1/2}$. The variability of $T(\mathbf{S})$ is still not accounted for, however. In fact, letting $W = kT(\mathbf{S})\sigma_G/\mu_{T(\mathbf{S})}\tilde{\sigma}$, we wish to set k so that $E(\Phi(W)) = 1 - \alpha$, whereas the preceding procedure sets $\Phi(E(W)) = 1 - \alpha$. By Jensen's inequality, this always results in too many rejections.

A second-order correction can be obtained using the delta method. Suppose we define a function $f(W)$ of the random variable by $f(W) = \Phi(W)$. The first two moments of W are

$$\begin{aligned}\mu_W &= k\sigma_G/\tilde{\sigma} \\ &= k/(1 + \sigma^2_{T(\mathbf{G})}/\sigma^2_G)^{1/2}, \\ \sigma^2_W &= k^2\sigma^2_{T(\mathbf{S})}\sigma^2_G/(\mu^2_{T(\mathbf{S})}\tilde{\sigma}^2) \\ &= k^2\sigma^2_{T(\mathbf{S})}/(\mu^2_{T(\mathbf{S})}(1 + \sigma^2_{T(\mathbf{G})}/\sigma^2_G)). \qquad \text{(A.3)}\end{aligned}$$

Expanding $f(W)$ in a Taylor series up to the term of second degree and taking expectations, we obtain

$$\begin{aligned}&E(f(W)) \\ &\approx \Phi(\mu_W) + \tfrac{1}{2}\sigma^2_W\phi'(\mu_W) \\ &= \Phi(k\sigma_G/\tilde{\sigma}) + \tfrac{1}{2}(k^2\sigma^2_{T(\mathbf{S})}\sigma^2_G/(\mu^2_{T(\mathbf{S})}\tilde{\sigma}^2))\phi'(k\sigma_B/\tilde{\sigma}), \qquad \text{(A.4)}\end{aligned}$$

where ϕ is the density and Φ is the cumulative distribution function of a standard normal variate. The equation $E(f(W)) = 1 - \alpha$ may then be easily solved for k using Newton's method.

A different method was used to set the multiplier for control charts for spread. The upper and lower control limits for a spread chart plotting the subgroup statistic G_i are conventionally set by

$$\begin{aligned}\text{UCL} &= T(\mathbf{S})\mu_G/\mu_{T(\mathbf{S})} + kT(\mathbf{S})\sigma_G/\mu_{T(\mathbf{S})} \\ \text{LCL} &= \max(0, T(\mathbf{S})\mu_G/\mu_{T(\mathbf{S})} - kT(\mathbf{S})\sigma_G/\mu_{T(\mathbf{S})}). \qquad \text{(A.5)}\end{aligned}$$

Thus

$$\begin{aligned}\Pr(G_i > \text{UCL}) &= \Pr(G_i^2\mu_{T^2(\mathbf{S})})/T^2(\mathbf{S})\mu_{G^2} \\ &> \mu_{T^2(\mathbf{S})}(\mu_G + k\sigma_G)^2/(\mu_{G^2}\mu^2_{T(\mathbf{S})}). \qquad \text{(A.6)}\end{aligned}$$

We treat $G_i^2\mu_{T^2(\mathbf{S})}/(T^2(\mathbf{S})\mu_{G^2})$ as having approximately an F distribution with numerator degrees of freedom $2\mu^2_{G^2}/\sigma^2_{G^2}$ and denominator degrees of freedom $2\mu^2_{T^2(\mathbf{S})}/\sigma^2_{T^2(\mathbf{S})}$, matching the coefficient of variation of χ^2 variates. This would be exact if G were the standard deviation and $T(\ )$ were the root mean squared average.

[*Received February 1987. Revised November 1988.*]

REFERENCES

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972), *Robust Estimates of Location: Survey and Advances*, Princeton, NJ: Princeton University Press.

Box, G. E. P., and Meyer, R. D. (1986), "Dispersion Effects From Fractional Factorials," *Technometrics*, 28, 19–28.

Clifford, P. C. (1959), "Control Charts Without Calculations," *Industrial Quality Control*, 15, 40–44.

Dixon, W. J., and Tukey, J. W. (1968), "Approximate Behavior of the Distribution of Winsorized *t* (Trimming/Winsorization 2)," *Technometrics*, 10, 83–98.

Fellner, W. H. (1986), "Robust Estimation of Variance Components," *Technometrics*, 28, 51–60.

Ferrell, E. B. (1953), "Control Charts Using Midranges and Medians," *Industrial Quality Control*, 9, 30–34.

Iglewicz, B. (1983), "Robust Scale Estimators and Confidence Intervals for Location," in *Understanding Robust and Exploratory Data Analysis*, eds. D. C. Hoaglin, F. Mosteller, and J. W. Tukey, New York: John Wiley, pp. 404–431.

Iglewicz, B., and Hoaglin, D. (1987), "Use of Boxplots for Process Evaluation," *Journal of Quality Technology*, 19, 180–190.

Langenberg, P., and Iglewicz, B. (1986), "Trimmed Mean $\overline{X}$ and R Charts," *Journal of Quality Technology*, 18, 152–161.

Lucas, J. M., and Crosier, R. B. (1982), "Robust CUSUM: A Robustness Study for CUSUM Quality Control Schemes," *Communications in Statistics—Theory and Methods*, 11, 2669–2687.

Rocke, D. M. (1983), "Robust Statistical Analysis of Interlaboratory Studies," *Biometrika*, 70, 421–431.

Rocke, D. M., Downs, G. W., and Rocke, A. J. (1982), "Are Robust Estimators Really Necessary?" *Technometrics*, 24, 95–101.

Stigler, S. M. (1977), "Do Robust Estimators Work With Real Data?" *The Annals of Statistics*, 5, 1055–1098.

Taguchi, G. (1986), *Introduction to Quality Engineering*, Tokyo: Asian Productivity Organization.

Tukey, J. W., and McLaughlin, D. H. (1963), "Less Vulnerable Confidence and Significance Procedures for Location Based on a Single Sample (Trimming/Winsorization 1)," *Sankhyā*, Ser. A, 25, 331–352.

Wadsworth, H. M., Stephens, K. S., and Godfrey, A. B. (1986), *Modern Methods for Quality Control and Improvement*, New York: John Wiley.

White, E. M., and Schroeder, R. (1987), "A Simultaneous Control Chart," *Journal of Quality Technology*, 19, 1–10.