6-1986

A Procedure for Initiating Process Control

Kim Ingrid Melton
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by Kim Ingrid Melton entitled "A Procedure for Initiating Process Control." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Management Science.

Ralph G. O'Brien, Major Professor

We have read this dissertation and recommend its acceptance:

Kenneth Gibson

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Kim Ingrid Melton entitled "A Procedure for Initiating Process Control." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Management Science.

Ralph G. O'Brien, Major Professor

We have read this dissertation
and recommend its acceptance:

Kenneth C. Gilbert

Lyle V. Jones

Lyle B. Ramsey

Accepted for the Council:

Vice Provost
and Dean of The Graduate School

A PROCEDURE FOR INITIATING PROCESS CONTROL

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Kim Ingrid Melton

June 1986

ABSTRACT

The purpose of this research is to develop a procedure to initiate process control. Traditional methods of process control do not allow testing until an initial group of samples (usually 20 or more) is drawn. In addition, the testing for control with respect to the process average is dependent on evidence that the variability in the process is stable.

The procedures developed and studied here allow immediate testing on the process average without assuming that the process is in control with respect to variability. Through the use of a sequential testing procedure, data from a current sample is tested to determine if it indicates that the process average has shifted and, if not, the estimate of the process mean is updated.

Two sets of test statistics are considered for the sequential procedure. Both are modifications of the traditional test statistic used to test the equality of the means of two independent samples. One set of test statistics modifies the estimate of σ^2 by pooling the sample variances; the other set of test statistics uses a version of the Welch t with estimated degrees of freedom.

An empirical study is conducted to evaluate the proposed procedure using all versions of the test statistics. Comparisons are made between the proposed procedure and current practices. Finally, extension of this procedure to estimation of other parameters and to applications outside statistical process control are suggested.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE

CHAPTER 1

INTRODUCTION

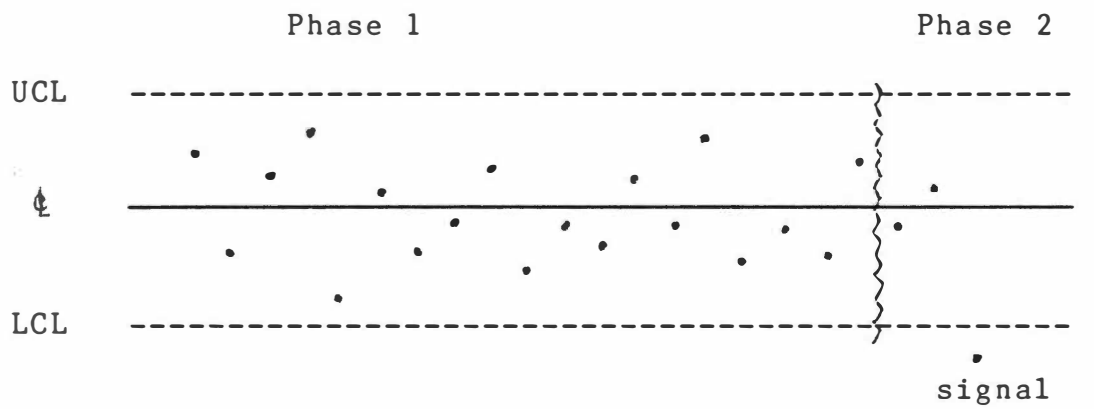
Quality control is a growing concern for American production managers. Applying scientific and statistical methodology to improve quality is known as statistical process control (SPC). Historically, the control chart has been a basic tool in SPC. Control charts for the average, $\bar{X}$, and for the range, R, are two common methods for monitoring production. $\bar{X}$ charts assess whether a process is "on target"; R charts assess whether the amount of spread in the observations is remaining consistently the same. Traditionally, both methods assume that the measure of interest is normally distributed. Whereas $\bar{X}$ chart methods are relatively robust if the measures are not really normally (Gaussian) distributed, interpretation of an R chart can be seriously compromised if used with non-normal data.

Setting Up Control Charts for Average and Variability

Control charts are established based on the premise that a certain amount of variability is inherent in any process. The objective is to differentiate between that natural variability (common cause variation) and variability produced by special causes. Control charts

for average and variability are used together to assess the location and spread in the measurements from a process. The most popular of these charts are $\bar{X}$ and R charts.

The modern control chart, introduced by Walter A. Shewhart in 1924, resembles that shown in Figure 1. The chart has a central line, ℄; an upper control limit, UCL; and a lower control limit, LCL.



Phase 1: Initial data (first k samples)
Phase 2: Continued monitoring

FIGURE 1. Sample Control Chart

In the case of $\bar{X}$ and R charts these lines are established by taking k samples (subgroups) of n observations each. The range and the mean of each subgroup is calculated. Next the average of the subgroup

ranges and the average of the subgroup means are calculated ($\bar{R}$ and $\bar{\bar{X}}$). These serve as initial values for the central lines on the respective charts. Control limits for the R chart are $\bar{R} \pm 3\hat{\sigma}_R$ where $\bar{R} - 3\hat{\sigma}_R$ is estimated by $D_3\bar{R}$ and $\bar{R} + 3\hat{\sigma}_R$ is estimated by $D_4\bar{R}$. The k subgroup ranges are then plotted on the R chart. If the process variability can be assumed to be stable (in control), the variability in the individual measurements can be estimated from the range using $\hat{\sigma}_x = \bar{R}/d_2$. The control limits on the $\bar{X}$ chart can then be set at $\bar{\bar{X}} \pm 3\hat{\sigma}_{\bar{x}}$ where $3\hat{\sigma}_{\bar{x}}$ is estimated by $A_2\bar{R}$. The k subgroup means may then be plotted on the $\bar{X}$ chart. The values for D_3, D_4, d_2, and A_2 can be obtained from Table A-1 and depend on the sample size and the assumption that the individual measurements follow a normal distribution. A derivation of the expressions for D_3, D_4, and A_2 may be found in Burr (1976). These constants are functions of the d_2 and d_3 values calculated by Pearson (1932).

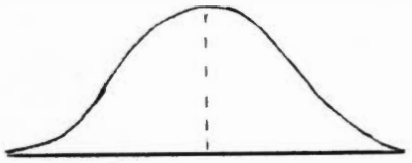
Interpretation of Control Charts

If no signals of instability are found in the initial data, the initial limits can be used to continue monitoring the process. Each time a sample (subgroup) is taken, $\bar{X}$ and R for that sample are plotted. As long as the point plotted falls between the control limits, the

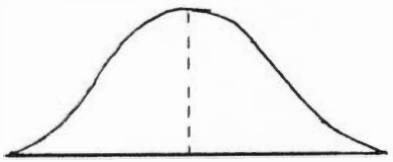
user of the chart will conclude that the sample data are consistent with the hypothesis that the individual measurements were drawn from a normal distribution with mean and standard deviation estimated by $\bar{\bar{X}}$ and $\hat{\sigma}_x$, respectively.

Figure 2 illustrates the four possible states of the system. In case 1, points should rarely fall outside the control limits for the $\bar{X}$ chart or the R chart. In Case 2, a shift in the mean has occurred while the variance has remained constant. A shift of this nature should be detected by a point outside the control limits of the $\bar{X}$ chart. Case 3 depicts a change in the variance with no change in process average. In this situation, the shift should be recognized by a signal on the range chart. And finally, Case 4 represents a shift in both process average and process variability. Out of control signals may occur on either chart.

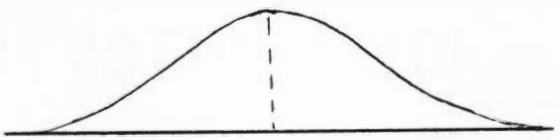
The user should realize that the information received from a sample may be misleading. Points could fall within the control limits of both charts even though the process mean, standard deviation or both have actually shifted (Cases 2, 3, and 4). An approximation to the expected number of samples that must be taken before a shift of a known size can be detected is computed as the expected



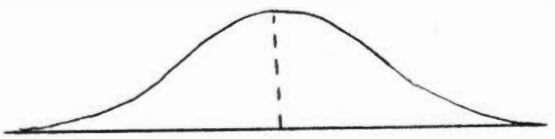
Case 1: Process in control
$X \sim N(\mu, \sigma^2)$



Case 2: Shift in the process average
$X \sim N(\mu+\delta, \sigma^2)$



Case 3: Shift in the process variance
$X \sim N(\mu, (\sigma+\Theta)^2)$



Case 4: Shift in average and variance
$X \sim N(\mu+\delta, (\sigma+\Theta)^2)$

FIGURE 2. Distribution of X

value of a random variable that follows a geometric distribution. A failure to detect a shift in the process is called a Type II error. On the other hand, the point could fall outside the control limits by chance alone. The user would be sent looking for a special cause of variability when none exists (Case 1). This is called a Type I error.

In addition to examining the position of plotted points relative to control limits, time-ordered data are examined for non-random patterns. Patterns (runs or clusters) can suggest a shift in the process even if all points are inside the control limits.

Also, the method of subgrouping can affect the interpretation of the charts. It will be assumed in the following that subgroups are formed in such a way that within-group variability is the common cause variation present in the system (including measurement variability).

Theoretical Implications

Control limits should be set so that the probability of a Type I error for any given point is known (and small). The probability of a Type I error for a single point on a range chart--or for any other chart used to track variability--is generally not known by users of control charts. When three standard deviation limits are

set for the $\bar{X}$ chart, the probability of a Type I error for a single point is taken to be 0.0027. To arrive at this value, it has been assumed that the X's are normally distributed and that enough subgroups were taken to assure "good" estimates of μ_x and σ_x. The Central Limit Theorem assures us that as the sample size (n) increases, the sampling distribution of $\bar{X}$ approaches the normal distribution regardless of the distribution of the individual measurements. Even with small sample sizes, the distribution of $\bar{X}$ will be nearly normal if the distribution of individual measurements is not markedly non-normal. Since many distributions encountered in industrial applications are approximately normal, the $\bar{X}$ chart is considered robust to its distributional assumptions.

The distribution of the range is non-normal even when the individual measurements are from a normal population. In fact, the distribution of the range can be shown to be skewed to the left for small samples from some distributions ($n > 3$ from the uniform distribution) and skewed to the right for other distributions (the normal and exponential). Traditionally, control limits for the R chart have been set as though the distribution of the range were symmetric.

By analogy with the sample variance, a preliminary idea of the effect of non-normality on the probability of a Type I error on charts for variability may be gained. If we use $s^2 = \Sigma (x - \bar{x})^2/(n - 1)$, we have an unbiased estimate for σ^2 no matter how X is distributed. But, the variance of s^2, $V(s^2) = \sigma^4[2/(n - 1) + \gamma_2/n]$, depends on the standardized kurtosis (γ_2) of the underlying distribution. When the distribution of individual measurements is normal, or assumed normal, $\gamma_2 = 0$. If the true distribution is platykurtic, $\gamma_2 < 0$, the normal theory estimate for $V(s^2)$ will be too large. Such estimates produce three-sigma limits that are too large so that fewer than the expected number of s^2 values will lie outside of those limits. The opposite is true from leptokurtic distributions, $\gamma_2 > 0$--the normal theory estimate of $V(s^2)$ is too small, and more than the expected number of s^2 values will end up outside the control limits. This relationship is shown in Figure 3 where $E[\hat{V}(s^2)]$ is the expected value of the normal theory estimate ($\gamma_2 = 0$), and $V(s^2)$ is the true value of the variance of the sampling distribution of s^2. Unfortunately, the user of charts for variability rarely knows the underlying distribution and cannot use the true $V(s^2)$.

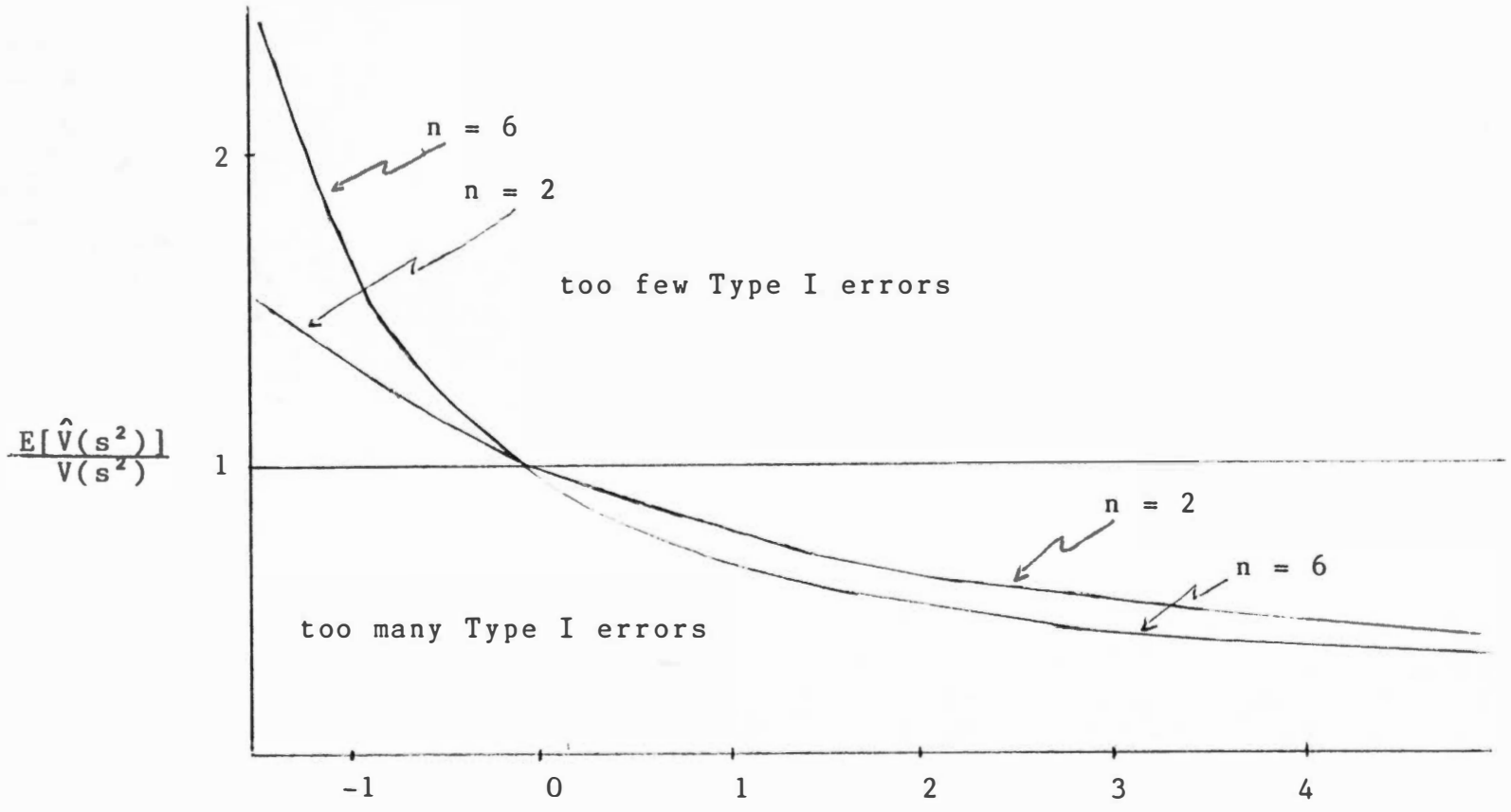


FIGURE 3. Ratio of the Expected Value of the Normal Theory Estimate of Variance of s^2 to the True Variance of s^2 as a Function of Kurtosis of Parent Distribution

The effect of using normal theory limits on the Range Chart is almost as dramatic. Even though a general expression for the variance of the range is not easily comprehended, examination of a few specific distributions can adequately illustrate the effect of deviations from normality. For samples drawn from a uniform distribution on the interval (a,b), the true variance of the range can be expressed as

$$V(R) = \frac{2(b - a)^2(n - 1)}{(n + 1)^2(n + 2)} .$$

The normal theory estimator for V(R) can be obtained by using the knowledge that the upper control limit for the R chart is set $3\sigma_R$ above the central line. Then for samples of size 2 the ratio $E[\hat{V}(R)]/V(R)$ has a value of 1.142. With samples of size 6 this ratio climbs to 2.240. In other words, what we thought were 3 sigma control limits are really 4.43 sigma control limits. A similar comparison can be made in the case of the exponential distribution. In this case the variance of R can be written

$$V(R) = \frac{n-1}{\lambda^2}\sum_{i=0}^{n-2}(-1)^i\binom{n-2}{i}\frac{2}{(i+1)^3} - \left[\frac{n-1}{\lambda}\sum_{i=0}^{n-2}(-1)^i\binom{n-2}{i}\frac{1}{(i+1)^2}\right]^2 .$$

For n = 2, $E[\hat{V}(R)]/V(R) = 0.571$; and for n = 6, the ratio drops to 0.399. Therefore, for samples of size 6 from the

exponential distribution, 3 sigma control limits are actually only 1.89 sigma control limits.

Another rarely considered fact is that the number of subgroups used to set up the control limits does affect the probability that a given point will lie outside the limits. As the number of subgroups increases, the estimate of the population parameter should improve. Based on this knowledge control limits established for a given nominal value of Type I error should be closer to the true limits when more subgroups are used.

An Example

In the manufacture of explosives, one of the quality characteristics of interest is moisture content. The desired amount of moisture is zero. If X represents the moisture content of an item, we can study the distribution of X. When the process is in control the density function of X could resemble that shown in Figure 4.

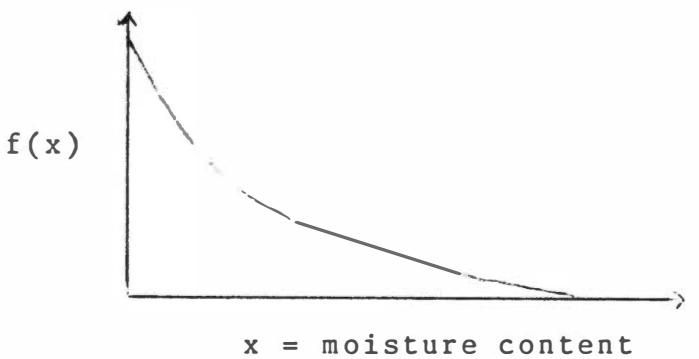


FIGURE 4. Distribution of Moisture Content

We might classify this distribution as "exponential-like." If the usual normal theory approach were used to analyze the process, the results could be misleading. For the sake of this example, let us assume that moisture content follows an exponential distribution with parameter 1 and that we are observing a process that is in control. The effects of using normal theory to analyze data collected from this process can be seen in Table 1. An assumption made to produce the table is that there is no sampling variation in the centerline for the R chart. The centerline is taken to be E(R), where

$$E(R) = \frac{n-1}{\lambda} \sum_{i=0}^{n-2} (-1)^i \binom{n-2}{i} \frac{1}{(i+1)^2} .$$

For samples of size 2 to 6 there is no lower control limit, and the normal theory upper control limits are given by $D_4 E(R)$.

If the parent distribution were normal, the true single point α value (probability of a Type I error) would be approximately 0.005. Using n = 5, the moisture content data would produce about 10 times as many "false signals" as would be expected.

Even though α is the probability that a single point will be outside the control limits, it may be more interesting to look at the probability that the chart will

TABLE 1

EFFECTS OF USING NORMAL THEORY CONTROL LIMITS ON P(R > UCL)--EXPONENTIAL PARENT DISTRIBUTION

n	E(R)	D_4	Normal theory UCL	Single point α P(R > UCL)
2	1.00	3.267	3.267	0.0381
3	1.50	2.575	3.8625	0.0416
4	1.83	2.282	4.1829	0.0451
5	2.083	2.115	4.4055	0.0480
6	2.283	2.004	4.5751	0.0505

contain one or more points outside the control limits. Let p be the probability that a single point is outside the control limits by chance alone, and let k be the number of points, assumed to be independent, on the chart. Then, assuming no sampling variability in ¢, $1 - (1 - p)^k$ is the probability that at least one point will fall outside the control limits. If we have a chart for samples of size 5 with 20 points drawn from a normal population, the probability of at least one false out-of-control signal is 0.095. If the underlying population had been exponential and the same chart were used, the probability would increase to 0.626. Since control limits are not placed on the $\bar{X}$ chart until process variability can be estimated (the process is in control with respect to variability), the chances of setting up charts for continued sampling are low. We need new ways to initiate process control.

Overview of Remaining Chapters

In the remaining chapters we deal with the development of a new procedure for initiating process control. This method would be useful in situations where monitoring is resumed after adjustments have been made as well as where SPC techniques are being used for the first time.

Since one of the primary roadblocks to establishing both $\bar{X}$ and R charts seems to be the sensitivity of the test for variability, two avenues of study are available. The most desirable approach would be to develop robust tests for variability. Many researchers, both inside and outside the quality control setting, have attempted to develop such procedures. Their work is reviewed in Chapter 2.

A second approach would be to develop new methods for the initial testing of the process average that are less dependent on the equality of variances. Since traditional tests on means have been shown to be fairly robust to distributional assumptions and since we can rely on the Central Limit Theorem (at least asymptotically), this approach seems to be worth considering. Chapter 3 reviews methods that have been suggested for initiating control charts for means and tests for equality of means when variances are not assumed to be equal.

Chapter 4 will present a sequential procedure for assessing process average. The method of evaluation of this procedure will also be presented.

In Chapter 5, the results of an empirical study to evaluate the new procedure will be presented. The effect

of unequal variances will be evaluated. In addition, the robustness of this new procedure will also be studied.

Chapter 6 concludes with a discussion of related future research. This includes suggestions for extensions of this method to estimating other population parameters. Uses for this procedure outside the SPC setting will also be suggested.

CHAPTER 2

REVIEW OF VARIANCE TESTING

A Historical Overview of Applications in SPC

Because it is so simple to compute and intuitively easy to understand, the range has been a popular measure of variability for use in quality control applications. Since Tippett (1925) published his table of the mean range in random samples from a normal population, the range has been the most widely recognized measure of dispersion for use in the process control setting. It has become the cornerstone of variance testing methods in the industrial setting.

As early as 1928 Pearson and Adyanthaya questioned the use of the range to estimate variability in non-normal populations. Cox (1954) stressed that in large samples the distribution of the range is determined by the tails of the population and therefore is very sensitive to non-normality. In the case of small samples, Pearson (1950) concluded that the ratio of mean range to population standard deviation is not much affected by the form of the population, but the coefficient of variation of the range depends on the form of the parent distribution. David (1954) stated, without proof, that the standard deviation

of the range is roughly a function of $\sqrt{\gamma_2 + 2}$. The implications to SPC are direct--even in the presence of non-normality, the use of $\bar{R}/d_2$ to estimate σ_x may be appropriate, but the control limits on the R chart are suspect.

Burr (1967) addressed the problem of setting control limits when he published modified control chart constants for use with non-normal data. Burr concluded that the change in constants was so small over a wide range of values for skewness and kurtosis, that the absence of normality did not have a marked effect on the control charts. The example in Chapter 1 seems to contradict this statement. When Burr's constants would be used is also questionable--in order to estimate the mean and variance (first and second moments) of the parent distribution, the user is required to know the third and fourth moments of that distribution.

Even though statisticians are in agreement that the range is an inefficient estimator of the variance in large samples and the distribution of the range is so dependent on the population, other measures of spread have been extremely slow to appear in the quality control setting. Duncan (1974) suggested the use of s^2 or s, but noted that

the traditional tests on these variables are not robust to non-normality.

More recently Kander and Zacks (1966), Lorden (1971), and Pollak (1985) suggested approaches to detecting changes in parameters of a distribution that are not based on the concept of the control chart. None of these authors gave specific tests for variability based on the range or any other measures of dispersion. Their general approach was to observe X_1, X_2, . . . independent random variables such that X_1, . . ., X_{m-1} each had distribution function F_0; and X_m, X_{m+1}, . . . each had distribution function F_1, where $1 \leq m < \infty$ was unknown. The task was to detect the shift from F_0 to F_1 as soon as possible after its occurrence. Kander and Zacks and Lorden employed a Bayesian approach along with a sequential analysis test that resembled a CUSUM procedure. Pollak claimed to set stopping rules in a non-Bayesian setting. Even if these approaches are appropriate in a process control setting, which they seem to be, their mathematical complexity will jeopardize the possibility of their implementation.

Using Control Charts for Variance Testing

If the range is the cornerstone of variance testing in the quality control setting, the control chart itself must be considered the foundation of testing in the

process control area. A general approach to setting up control charts was given in Chapter 1. This procedure involved taking k subgroups of size n from a process that was assumed to be in control and computing, from these nk observations, values for the central line and control limits. In order to use the control chart constants A_2, D_3, D_4, and d_2, it was assumed that k (the number of subgroups) was large and the individual measurements followed a normal distribution.

Once initial charts are established, the traditional approach to continued testing is very simple and straightforward. When the initial control limits are established and there is no indication that the process is out of control, extend the center line and the control limits. As new samples are taken, plot the sample statistic on the appropriate chart, observe whether the point is inside the control limits, and continue to watch for nonrandom patterns in the points. Continue to plot points on this chart until there is some indication that the process has shifted.

Hillier (1967) derived control chart constants for the range chart when a small number of subgroups of size 5 were used. With the use of his modified constants, D_3^* and D_4^*, valid control limits (the probability of a Type I

error on future points is the stated value) can be established after as few as two subgroups. He pointed out that the power of the test increases as more subgroups are used. For this reason, he suggested updating control limits several times as more subgroups are obtained. From the standpoint of effort involved and the psychological impact on the workers, he suggested that the updating be done after 5, 10, 25, and 100 subgroups. Interpretation of this chart is the same as that for the traditional chart.

Additional work (Hillier, 1969 and Yang and Hillier, 1970) extended this approach to handle retrospective testing and tests based on the sample variance rather than the range. Retrospective testing attempts to assess if the process was in control when the original subgroups were drawn. Process control then becomes a two stage procedure--stage 1 sets control limits for retrospectively testing whether the process was in control while the initial groups were drawn, and stage 2 sets control limits for testing whether the process remains in control when future subgroups are drawn. Determination of the number of subgroups to include in stage 1 is based on the need for early control as well as the desired power of the control limits for detecting an out of control process.

If power is more important than computational ease, charts based on the sample variance or standard deviation may be used instead of the R chart. Modified control chart constants for these charts are derived in the 1970 paper.

Variance Testing in a General Setting

Tests for variability are not unique to SPC. Until recently techniques developed for use in the process control setting emphasized easy interpretation and easy computation. Because of the computational aspect, industrial users have been reluctant to leave tests based on the range. As SPC techniques are applied in a wider variety of situations, where the assumption of normality is likely to be violated, other tests must be considered. With the rapidly expanding availability of computers, techniques that were previously considered inappropriate in the process control setting deserve another look.

Currently existing methods for testing variability may be classified in three broad areas: 1) methods based on the assumption of normality, including methods for transforming data to meet this assumption, 2) methods that rely on the jackknife technique, and 3) other methods that use traditional ANOVA approaches. Rather than one test that works in most situations, a variety of variance tests

are available. Statisticians have developed tests based on s^2, s, ln s, ln s^2, the range, and a host of other spread statistics.

Tests Based on Normal Theory. Introductory applied statistics texts teach the use of a χ^2 or F test as a test for dispersion. To test whether the variance of a normal distribution, σ^2, equals some specified value, say σ_o^2, the test statistic $\chi_o^2 = (n-1)s^2/\sigma_o^2$ is computed. Likewise, to test that two independent normal populations have equal variances, $F_o = s_1^2/s_2^2$ is calculated. Comparing either χ_o^2 or F_o to their appropriate sampling distribution leads to a decision to accept or reject H_o: $\sigma^2 = \sigma_o^2$ or H_o: $\sigma_1^2 = \sigma_2^2$. Most texts contain warnings similar to that found in Hines and Montgomery (1980): "Unlike the t test . . ., tests on variances are rather sensitive to the normality assumption."

Box (1953) went so far as to suggest that even though the fairly robust procedure for testing equality of means depends on the assumption of equal variances, the researcher may be better advised to skip the step of testing for homogeneity of variance. He pointed out that the sensitivity to non-normality increased as the number of variances to be compared increased. Without offering an actual test, he concluded that in order for the

decision criterion to be less dependent of the population kurtosis, information on the variation to be expected in the sample variances should be gathered from internal evidence. Then a test of variances should be "studentized" for the fourth moment just as a test for means is studentized for the second moment.

Bartlett (1937) generalized the F_o test to test the equality of more than two variances. If there are k random samples from independent normal populations, the sampling distribution of the statistic is approximately χ^2 with k-1 degrees of freedom. The test statistic is calculated as

$$\chi_o^2 = 2.3026\ q/c$$

where

$$q = (N-k)\log s_p^2 - \sum_{i=1}^{k}(n_i-1)\log s_i^2$$

$$c = 1 + \frac{1}{3(k-1)}\left[\sum_{i=1}^{k}(n_i-1)^{-1} - (N-k)^{-1}\right]$$

$$s_p^2 = \sum_{i=1}^{k}(n_i-1)s_i^2/(N-k)$$

and s_i^2 is the sample variance for the i^{th} subgroup, n_i is the number of observations in subgroup i, and N is the total number of observations. Bartlett's test is very sensitive to the normality assumption.

Often, when the normality assumption is violated, transformations can be used. Possible transformations include 1) the square root transformation, $x^*_{ij} = \sqrt{x_{ij}}$ or $x^*_{ij} = \sqrt{1 + x_{ij}}$, for observations that follow a Poisson distribution, 2) the log transformation, $x^*_{ij} = \log x_{ij}$, for data that follows a lognormal distribution or where the mean is proportional to the standard deviation, 3) the arcsine transformation, $x^*_{ij} = \arcsin x_{ij}$, for observations from a binomial distribution, and 4) the reciprocal transformation, $x^*_{ij} = 1/x_{ij}$, for situations where the standard deviation is proportional to the square of the mean.

Tests Based on the Jackknife Technique. No discussion of variance testing would be complete without reference to the jackknife. The jackknife is a statistical tool that can be used to reduce bias and develop approximate confidence intervals. Although the jackknife is not appropriate in all situations, it does seem to provide a robust and reasonably powerful tool for testing variances. A basic description of the jackknife was given by Miller (1968). Let Θ be an unknown parameter, and let $X_1, \ldots, X_N$ be a sample of N independent, identically distributed observations with cumulative distribution function F_Θ. Let $\hat{\Theta}$ denote an

estimate of Θ based on all N observations. Divide the data into n groups of size k (N=nk). This division may be determined by the structure of the experiment or arbitrarily imposed by the statistician. Now, let $\hat{\Theta}_{-i}$, i=1, . . ., n, denote the estimate of Θ obtained by deleting the i^{th} group and estimating Θ from the remaining (n-1)k observations. Form the pseudovalues

$$\tilde{\Theta}_i = n\hat{\Theta} - (n-1)\hat{\Theta}_{-i} , \quad i = 1, \ldots n.$$

The jackknife estimate of Θ is the average of the $\tilde{\Theta}_i$:

$$\tilde{\Theta}_\cdot = \sum \tilde{\Theta}_i/n.$$

Based on Tukey's proposal that in many instances the $\tilde{\Theta}_i$ are approximately independently, identically distributed, the variance of $\tilde{\Theta}_\cdot$ can be expressed approximately as

$$V(\tilde{\Theta}_\cdot) = (n(n-1))^{-1} \sum (\tilde{\Theta}_i - \tilde{\Theta}_\cdot)^2 .$$

Also, $(\tilde{\Theta}_\cdot - \Theta)[n(n-1))^{-1} \sum (\tilde{\Theta}_i - \tilde{\Theta}_\cdot)^2]^{-1/2}$ "should" follow a t distribution with n-1 degrees of freedom. Based on this information, approximate confidence intervals and significance tests on Θ can be performed.

Arvesen and Schmitz (1970) suggested using the log transformation as a variance stabilizing transformation in conjunction with the jackknife. They claimed that if no transformation were made, in moderate sample sizes random

variables from skewed distributions would be forced to behave as though they came from symmetric distributions.

Other Tests Based on ANOVA Methods. One way to view Miller's jackknife technique is to note that the original data is being transformed (to pseudovalues) prior to the use of ordinary ANOVA-type inference methods. Other transformations have also been suggested. Data is transformed so that robust normal theory tests on means can be applied to test dispersion.

One of the most widely recognized of these tests was suggested by Bartlett and Kendall (1946). This test depends on the arbitrary subdividing of samples drawn from two populations, calculation of log s^2 from each of these subsamples, and comparison of the log s^2 values from the two populations by use of a two sample t test for location. Since the subdivision of samples is arbitrary, the outcome of the test can vary with different groupings. This method can also be employed in cases with several populations using ANOVA on the subsample log s^2 values.

More recently Levene (1960), Miller (1968), O'Brien (1979, 1981), and others have suggested ANOVA based methods that do not depend on grouping the data in any arbitrary fashion. Levene proposed the spread variable $z_{ij}^2 = (x_{ij}-\bar{x}_j)^2$ where x_{ij} is the i^{th} observation in the

j^{th} group and x_j is the sample mean for the j^{th} group. Miller suggested jackknife pseudovalues based on $\Theta = \sigma^2$,

$$q_{ij} = n_j s_j^2 - (n_j - 1) s_{j-i}^2$$

where n_j is the sample size of the j^{th} group, s_j^2 is the sample variance of the j^{th} group, and s_{j-i}^2 is the sample variance of the j^{th} group if the i^{th} observation is deleted.

O'Brien (1979) suggested the use of the rather complicated looking transformation

$$r_{ij}(w) = [(w+n_j-2)n_j(x_{ij}-x_j)^2 - ws_j^2(n_j-1)]/[(n_j-1)(n_j-2)] .$$

Both Levene's z_{ij}^2 and Miller's q_{ij} are special cases of $r_{ij}(w)$--$z_{ij}^2 = (n_j-1)r_{ij}(0)/n_j$ and $q_{ij} = r_{ij}(1)$. The advantage of this measure is that the mean of $r_{ij}(w)$, $\bar{r}_j(w)$, is equal to the sample variance for the j^{th} group (s_j^2). The possibility then exists to use standard robust tests for equality of means to test these measures of variability. For balanced designs this test performs favorably. O'Brien advised against further transformation of the $r_{ij}(w)$ variable and against this procedure for one group tests.

Brown and Forsythe (1974) compared several of the ANOVA related methods. Included in their study were

Levene's $z_{ij} = |x_{ij} - \bar{x}_j|$ and several variations such as replacing $\bar{x}_j$ with the median or the trimmed mean; a χ^2 procedure given by Layard (1973); and the ln σ^2 version of Miller's jackknife procedure. Their research pointed to the need to consider robust estimates of location when developing robust tests for homogeneity of variances.

Implementation. Why have none of these methods made their way into the foreground of SPC work? There are several reasons. The first reason has already been discussed--there is a strong effort to avoid computationally complex approaches.

A second, and more difficult to grasp, reason may be based on a fundamental difference in the goal of the procedures. In traditional hypothesis testing, one usually works to show that there is sufficient evidence to conclude that the null hypothesis is not true, or the alternative is more likely the correct choice. In SPC, rejection of the null hypothesis is the exception rather than the rule. (A signal on a control chart is reason for action.)

Last, a more theoretical and philosophical question must be raised. The statistician must decide how to handle extreme values. Since samples are being drawn from an unknown population, do these values represent outliers

from a normal population, or are they reasonable values from a non-normal distribution. If the latter is true, can this be consistent with a concept of control?

CHAPTER 3

A REVIEW OF MEANS TESTING

Measures of Location Used in SPC

As in any area of application, in SPC central tendency can be measured in a number of ways. The most popular of these ways is 1) to look at the individual observations (X's), 2) to look at the median of a group of observations ($\tilde{X}$), or 3) to look at the mean of a group of observations ($\bar{X}$). The control chart for individuals, the X chart (Nelson, 1982), and the control chart for medians (Nelson, 1982b) have received less attention than the traditionally favored $\bar{X}$ chart. The X chart is used when it is not possible, or desirable, to form subgroups. Since individual measurements are plotted, the underlying distribution must be known if accurate probability statements are desired. Median charts are advocated when it is desirable to keep computation to a minimum. Today, the subgroup mean is very easily computed on a hand held calculator. The loss in efficiency (0.743 for samples of size 3 and lower for larger samples) that results from using the sample median instead of the sample mean should dictate the use of the $\bar{X}$ chart over the median chart for samples from a normal population.

Probability statements on the $\bar{X}$ chart are made on the basis of $\bar{X}$ being normally distributed. Asymptotically, this is true regardless of how X is distributed, but in small samples care needs to be exercised. Unless the deviations from normality are severe, probability statements can be taken as approximately correct. Since $\bar{X}$ is the most common estimate of location and it is reasonably robust, the remainder of this chapter will focus on the sample mean.

Means Testing in SPC

In process control there is a two-fold question to be answered--do the sample means indicate that their respective populations are centered on the same value, and what is that value. Most approaches to answering these questions, including traditional $\bar{X}$ control charts, start by assuming that the process is centered at one specific value; then they attempt to estimate that value. Once an estimate is available, the question asked in future samples becomes, "Can I accept that the sample mean of this new group is a reasonable value for a sample drawn from a population with the mean already estimated?"

Hillier (1969) and Yang and Hillier (1970) developed control charts to test the assumption of control on the initial samples as well as on future samples. In the

earlier work Hillier based his estimate of variability on the range, while the latter used the sample variance. When small numbers of subgroups are used, their control limits tend to be wider than limits on traditional charts. The authors claimed to set control limits so that the probability that a point will fall outside the control limits is a specified value.

King (1954) also suggested a method for retrospective testing. He noted that the estimate of variability tends to be inflated when a small number of subgroups is used. To compensate for this, he suggested new factors to replace A_2. His values, referred to as C, are smaller than the corresponding A_2 values. The C values are computed so that the probability that one or more points in 25 will be outside the control limits will be approximately 0.05.

The work of Hillier, Yang and King appears to stand alone in the area of initiating process control. These authors all suggest reformulating control limits after a large number of samples (25 or more). Their methods then coincide with traditional control charts.

Traditional control charts are designed so that the probability that a single point falls outside the control limits is a specified small value, regardless of the

sample size. Weiler (1954) suggested an alternative that set control limits so that the expected number of pieces inspected between false signals would be known. To illustrate how Weiler's method differs from the traditional approach, consider that the average run length--time between signals--for a traditional chart is approximately 1/0.0027 or 370. If samples are of size 4, an average of 1480 pieces will be inspected between false signals; for samples of size 6, the average number of pieces inspected between false signals jumps to 2220. Weiler would adjust the control limits to correspond to the sample size--control limits for samples of size 6 would be tighter than the limits for smaller samples for a fixed average number inspected between signals. This approach does not seen to have gained much attention.

Another rather unique approach to establishing an $\bar{X}$ chart can be attributed to Rahim (1985). He attempted to optimally determine sample size, sampling interval, and control limit coefficients. Rahim divided the production cycle into four periods--a period when the process is operating in control, a period when the process is operating out of control, a search period due to a false alarm, and a period that involves search and correction for a true signal. Optimization involves a search

technique over a function that includes estimated costs for the various parts of the production cycle. He also recognized measurement error, and did not require that the individual measurements follow a normal distribution.

The most popular approach to means testing other than the control chart is a procedure referred to as CUSUM (cumulative sum) (Lucas, 1976). In order to use a CUSUM procedure, estimates must be available for μ, σ, and the size shift that is to be recognized quickly. This procedure can be viewed as a special case of Wald's (1947) sequential analysis. Supporters of this method claim that small shifts in a process may be detected faster by a CUSUM procedure than by traditional control charts.

Means Testing in a General Setting

Means testing can be divided into three problems--the one group problem, the two group problem, and the multiple group problem (the first two actually being special cases of the last). Many times, tests for the one group hypothesis, H_o: $\mu = \mu_o$, can be modified to help estimate the unknown mean of a population.

The One Group Problem and Estimation.

The t test has served as the primary method for testing whether the mean of a population is equal to a specified value. To conduct a t test, a sample of size n is taken from a normal

population, the mean ($\bar{X}$) and the sample variance (s^2) are calculated, a test statistic $t_c = (\bar{x} - \mu_o)/\sqrt{s^2/n}$ is calculated, and t_c is compared to critical values from a t distribution with n-1 degrees of freedom. When the variance is known, s^2 is replaced with σ^2 and critical values are obtained from the standard normal table.

Wald (1947) suggested a sequential approach to testing whether the mean of a normal population is equal to a specified value. His procedure can be used with known or unknown variance. Wald suggested three decision alternatives at each stage of the procedure--accept H_o, reject H_o, or continue sampling. The expected number of observations required to reach a decision should be lower under the sequential analysis approach than under the corresponding single sample approach.

Another way to test the hypothesis is by inspection of the confidence interval. If the hypothesized value μ_o is not contained in the interval $\bar{x} - t_{n-1,\alpha/2}s/\sqrt{n} < \mu < \bar{x} + t_{n-1,\alpha/2}s/\sqrt{n}$, the null hypothesis is rejected.

Many times the confidence interval is also used to estimate the mean of a population. Ghosh (1975) described two measures of accuracy for estimating parameters by use of confidence intervals. Neyman accuracy refers to the likelihood that the interval actually contains the

parameter, and Wolfowitz accuracy refers to the width of the interval. A short confidence interval with high probability that the parameter is contained in the interval would provide a reliable estimate of the mean. One way to decrease the width of a confidence interval is to increase the sample size. Dantzig (1940) proved that the problem of finding a confidence interval of a preassigned length and a given confidence coefficient for an unknown mean of a normal distribution with unknown variance is insoluble if the sample size is fixed before sampling.

In an attempt to determine the number of observations that would be necessary to establish a confidence interval of a known length with a given confidence coefficient, Stein (1945), Weiss (1955) and Graybill (1958) introduced two stage procedures where the number of observations needed in the second stage is estimated at the first stage. Stein limited his work to normal populations. Weiss extended the procedure to show that the mean of all observations (stage 1 and stage 2) should be used for the center of the final confidence interval. Graybill dropped the requirement that the samples were drawn from a normal population.

The Two Group Problem. To test the hypothesis H_o: $\mu_1 = \mu_2$, we usually look at the contrast $\bar{x}_1 - \bar{x}_2$ and ask if this quantity is significantly different from zero. When the variances of the two populations are known, the test statistic $z_c = (\bar{x}_1 - \bar{x}_2)/\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ is compared to critical values obtained from the standard normal table. If the variances are not known, two possibilities exist. If these variances can be presumed to be equal then the sample variances may be pooled and the test statistic $t_c = (\bar{x}_1 - \bar{x}_2)/s_p\sqrt{1/n_1 + 1/n_2}$ can be compared to values from a t distribution with n_1+n_2-2 degrees of freedom. When the variances cannot be expected to be equal, an approximate t test may be employed. This is referred to as the Behrens-Fisher problem. Welch (1947) proposed an approximate t test for this problem. The test statistic is $t_c = (\bar{x}_1 - \bar{x}_2)/\sqrt{s_1^2/n_1 + s_2^2/n_2}$. He compared t_c to a t distribution where degrees of freedom were estimated by

$$\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2 \Bigg/ \left[\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}\right]$$ (Satterwaite, 1941).

Wang (1971) provided evidence that this approximation provides accurate results and that there is little loss of power when the variances are equal. Both SPSS and Minitab now provide the Welch t test as the default (SPSS also

prints out the more traditional t) when testing the difference in means from two independent samples.

The Multiple Group Problem. The traditional ANOVA F test is used to test H_o: $\mu_1 = \mu_2 = \ldots = \mu_k$ against an alternative H_a: at least one μ_j is different. When sample sizes are equal and the within group variances are equal, this test is robust to distributional assumptions. Brown and Forsythe (1974b) showed that when these assumptions were violated, the F test is no longer accurate. They observed that for a nominal α level of 0.05 the observed levels ranged from 0.03 to 0.17. The higher levels were observed when large variances were associated with small samples.

Welch's t test could be modified so that complex contrasts involving several group means could be tested. Kohr and Games (1977) provided support that the test statistic $t'' = \sum c_j \bar{x}_j / \sqrt{\sum c_j^2 s_j^2 / n_j}$ (c_j is the contrast coefficient for the j^{th} group) is approximately distributed as a t with degrees of freedom estimated by

$$df'' = \frac{(\sum c_j^2 s_j^2 / n_j)^2 - 2[\sum c_j^2 s_j^4 / n_j^2 (n_j+1)]}{\sum c_j^4 s_j^4 / n_j^2 (n_j+1)} .$$

Unfortunately, in small samples the estimate of degrees of freedom can be negative if the estimate of variance is small for groups that have large c_j's.

In 1951 Welch modified his two group test to form a test on the equality of several means when the population variances are not equal. His test statistic

$$W = \frac{w_j(\bar{x}_j - \tilde{x}_{..})^2/(k-1)}{1 + \frac{2(k-2)}{(k^2-1)}\sum(1 - w_j/u)^2/(n_j - 1)}$$

is approximately distributed as an F with k-1 and f degrees of freedom, where

$\bar{x}_j$ = the mean of group j

n_j = the number of observations in group j

$w_j = n_j/s_j$

$u = \sum w_j$

$\tilde{x}_{..} = \sum w_j\bar{x}_j/u$

k = the number of groups

$f = [(3/(g^2 - 1))\sum(1 - w_i/u)^2/(n_i - 1)]^{-1}$.

Brown and Forsythe (1974b) studied the behavior of this test statistic in small samples with unequal variances. They reported acceptable results with sample sizes as small as five. They actually recommended the use of this test statistic over the traditional ANOVA F test because of protection against possible distortions in reported significance levels.

Implementation. In the past little emphasis has been placed on modifying tests for the means in the SPC setting. Researchers have used the Central Limit Theorem

as a crutch to support their hypothesis that the mean is a robust estimate of location. Unfortunately, most of the tests that are used depend on being able to successfully conduct tests on variability. As SPC techniques make their way into more non-industrial settings, the likelihood of encountering non-normal populations will increase. As the distribution deviates more and more from the normal distribution, testing variability becomes more difficult. Robust tests for means are not useful unless they can be conducted. The groundwork has been provided in the preceding section to modify general procedures for testing equality of means, without first knowing that the variances are equal. Any modification should be compared to current procedures when the appropriate assumptions are met. Also the robustness of these procedures must be studied.

CHAPTER 4

A NEW PROCEDURE FOR MEANS TESTING

Initiating a program in process control is not a trivial procedure. Many times a company's first attempts to use traditional control charts result in confusion and chaos. Processes are not in control with respect to variability (or at least that is what the charts indicate); limits on charts for the mean are not appropriate; and the user is not sure where to look for the cause. The trouble may be limited to the variability in the process, the variance and the mean may both be varying, or the test for variability may be misleading because the distribution of the characteristic is not normal.

Variance tests that have been developed in an attempt to eliminate the normality assumption are not effective with sample sizes as small as those used in SPC applications. Tests for means that do not depend on equality of variances exist, but they also suffer with the small sample sizes. Since the assumption of normality is usually less binding in means tests, modification of the current means tests so that they are more efficient for small samples from unknown populations is worthwhile.

Several criteria would be used to judge the efficiency of a new procedure. First, does this procedure provide an accurate (on target) estimate for the mean? In other words, is the estimate of the mean unbiased? Second, is the variability in the estimate of the mean reasonable (in a practical sense)? This is not a statistical question even though the estimate of the variability in the sampling distribution of the mean would be used to help answer this question. Practical significance will vary from user to user, but statistics can provide some insight into the effects of some choices. And finally, since this is a procedure for initiating process control, how fast can we achieve an estimate for the mean?

A Proposed Procedure

A General Overview. The procedure suggested in this section can be viewed as a modification of the methods suggested by Yang and Hillier (1970). This procedure is sequential in nature. The testing and estimation procedure continues until the process is classified as 1) out of control or 2) in control at a predictable level.

In contrast to Yang and Hillier's approach, this procedure only modifies control limits until the user concludes that the process is in control at a predictable

level. After that control limits are not updated as a routine procedure--signals from the chart are used to determine when updating is necessary.

The procedure that I am recommending consists of a series of two group t tests with a slightly different denominator from what is traditionally used. At each iteration, there are three decision alternatives. We could accept the alternative hypothesis--conclude that the current sample is significantly different from what we have seen (conclude that the process is not in control with respect to the mean); we could accept the null hypothesis--conclude that the process is operating at predictable level; or we could conclude that we need more information before we will conclude that the process is operating at a consistent and predictable level. As we progress through iterations, the test to accept the alternative hypothesis becomes more like a one group test with a given hypothesized value for the mean.

To start the process, draw two samples of size n. Compute $\bar{x}_1$, $\bar{x}_2$, s_1^2, s_2^2. Then, without testing, compute $s_p^2 = (s_1^2 + s_2^2)/2$. Form the test statistic $t_{r_2} = (\bar{x}_1 - \bar{x}_2)/s_p\sqrt{(2/n)}$. If, based on the test statistic t_r, the alternative hypothesis is not accepted, the iterative process can begin.

Each time a new sample is taken $\bar{x}$ and s^2 are calculated for that sample. A test statistic for accepting the alternative hypothesis, t_{r_k}, is calculated and the test performed. If the alternative hypothesis is not accepted, a test can then be made to determine if we can accept the null hypothesis. For this test t_{a_k} will be calculated and a test performed. If the null hypothesis is accepted, the mean will be estimated; if the null hypothesis is not accepted another sample will be taken. Figure 5 shows the flow of this procedure.

Test to Accept the Alternative Hypothesis. Two test statistics for concluding that the process is out of control will be considered. Each attempts to reduce the effect of unequal variances on the outcome of the test. Refer to the test statistics as t_{r_k} and t'_{r_k}.

First, consider

$$t_{r_k} = \frac{\bar{x}_k - \bar{\bar{x}}_{k-1}}{\sqrt{s_p^2 \left(\frac{1}{(k-1)n} + \frac{1}{n}\right)}}$$

where

$\bar{x}_k$ = the sample mean of the new (k^{th}) sample

$\bar{\bar{x}}_{k-1}$ = the mean of the first k-1 samples

s_p^2 = the mean of the first k-1 sample variances.

Since k-1 variances with n-1 degrees of freedom each were "pooled" to form s_p^2, this test statistic is compared to

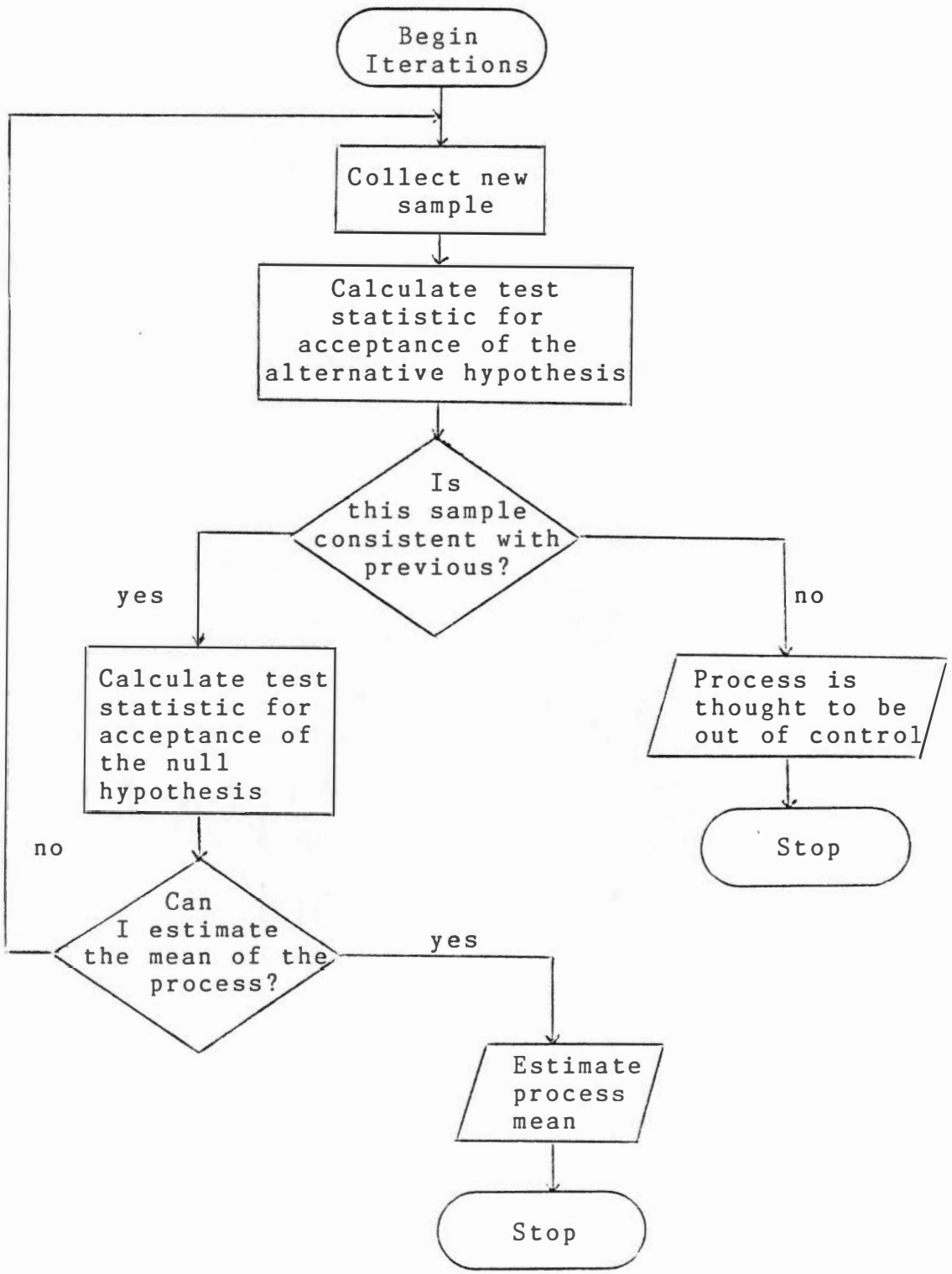


FIGURE 5. Flow Chart for the Iterative Procedure

critical values from a t distribution with (k-1)(n-1) degrees of freedom. Unlike any approach that estimates degrees of freedom, this method would always use the same reference value for any sample of size n after k iterations.

The decision to average the previous k-1 sample variances into s_p^2 is a two-fold attempt to reduce the effect of unequal variances. First, averaging provides a protection. Shifts in the mean will have less effect on s_p^2 than on a sample variance calculated about a grand mean. Any source of variability that can be eliminated should be considered. The decision to include only the first k-1 sample variances is a more direct attack on the possibility of unequal variances. If X_1 and X_2 are independent random variables from distributions with means μ_i and variances σ_i^2, then the distribution of $\bar{X}_1 - \bar{X}_2$ has a mean $\mu_1 - \mu_2$ and variance $\sigma_1^2/n_1 + \sigma_2^2/n_2$. If $\sigma_1 = \sigma_2$, we can write the variance of the difference as $\sigma^2(1/n_1 + 1/n_2)$. Traditionally, when σ^2 is unknown, it is replaced by the sample variance. For this iterative procedure, in order to reach the k^{th} iteration, we must have failed to accept the null and the alternative hypothesis at each of the previous iterations. Failing to accept the alternative hypothesis could be restated as not having

sufficient evidence to conclude that the process is out of control. Following this approach, the best estimate of σ^2 should be obtained from the first k-1 samples. Omission of the k^{th} sample variance provides some protection against sudden large shifts in the variance.

Now, consider

$$t'_{r_k} = \frac{\bar{x}_k - \bar{\bar{x}}_{k-1}}{\sqrt{\frac{s_p^2}{n(k-1)} + \frac{s_k^2}{n}}}$$

where $\bar{x}_k$, $\bar{\bar{x}}_{k-1}$, s_p^2 are the same as before and s_k^2 is the sample variance for the k^{th} sample. The test statistic will be compared to critical values from a t distribution with degrees of freedom estimated by the Satterwaite approximation:

$$df' = \frac{\left(\frac{s_p^2}{n(k-1)} + \frac{s_k^2}{n}\right)^2}{\frac{\left(\frac{s_p^2}{n(k-1)}\right)^2}{n(k-1)-1} + \frac{\left(\frac{s_k^2}{n}\right)^2}{n-1}} .$$

This would be a variation of Welch's t test that has been used for the two group means test when variances are not equal.

Test to Accept the Null Hypothesis. Each test for accepting the alternative hypothesis has a corresponding test for accepting the null hypothesis. These will be referred to as t_{a_k} and t'_{a_k}. The concept is the same for

both values. This test is only performed if the alternative hypothesis is not accepted. Rejecting the alternative allows the "pooling" of information. For this test, all k samples are divided into two equal (or nearly equal) groups of approximately k/2 samples in each group; a two group t test is performed; and a decision made.

The two test statistics are

$$t_{a_k} = \frac{\bar{\bar{x}}_{g_1} - \bar{\bar{x}}_{g_2}}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

and

$$t'_{a_k} = \frac{\bar{\bar{x}}_{g_1} - \bar{\bar{x}}_{g_2}}{\sqrt{\frac{s_{g_1}^2}{n_1} + \frac{s_{g_2}^2}{n_2}}} .$$

where

$\bar{\bar{x}}_{g_1}$ = grand mean of the first [(k+1)/2] sample means
[·] represents the integer portion of ·

$\bar{\bar{x}}_{g_2}$ = grand mean of the remaining sample means

s_p^2 = mean of the k sample variances

$s_{g_1}^2$ = mean of the first [(k+1)/2] sample variances

$s_{g_2}^2$ = mean of the remaining sample variances

n_1 = [(k+1)/2]n

n_2 = nk - n_1.

For t_{a_k} the critical values are obtained from a t distribution with k(n-1) degrees of freedom. In the case

of t'_{a_k}, degrees of freedom are estimated by the Satterwaite approximation:

$$df = \frac{\left(\frac{s_{g_1}^2}{n_1} + \frac{s_{g_2}^2}{n_2}\right)^2}{\frac{\left(\frac{s_{g_1}^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_{g_2}^2}{n_2}\right)^2}{n_2-1}} .$$

Choice of Critical Values. All four test statistics that have been suggested depend on critical values from a t distribution. The choice of significance levels will affect the ability of the procedure to recognize shifts.

The test for accepting the alternative hypothesis will be performed at a constant level of $\alpha = 0.003$. Therefore, when the process is in control with respect to the mean, at each iteration the probability that we will conclude that the process is out of control should be 0.003. This is approximately the same as the value claimed by traditional control charts.

The significance level for the test for accepting the null hypothesis will change at each iteration. The reason for this is simple--the more samples we take that are not significantly different, the more willing we are to conclude that they came from a distribution with a constant mean. Since this procedure is intended to

provide an initial estimate of the mean, the speed with which we can make the estimate is important. Unfortunately, there is a trade-off. A decision based on a small number of observations may be influenced by chance similarities in the data. As the number of observations increases, these chance similarities have less effect on the test. For this reason, we start with α set so that acceptance of the null hypothesis is difficult. As the number of iterations increases, the "acceptance region" also increases. Traditional control charts generally estimate the mean after 20 samples. In order to quarantee that the new procedure will produce estimates within this time period, the α for this new procedure is set to decrease linearly until a decision is forced by the twentieth sample. Since the first test for acceptance of the null hypothesis occurs after the third sample, $\alpha_k = (20-k)/18$. If the procedure does not terminate prior to the twentieth sample, the traditional control chart and this procedure should provide the same estimate of the mean.

Testing the New Procedure

Evaluation of the new procedure needs to include three areas: 1) how the procedure works when traditional assumptions of normality and homogeneity of variance are

met, 2) how the procedure works when the homogeneity of variance assumption is violated, and 3) how the procedure works when the normality assumption is violated. In each case, comparisons need to be made between this procedure and the traditional $\bar{X}$ and R charts or $\bar{X}$ and s charts. These comparisons will be based on the results of an empirical study.

CHAPTER 5

AN EMPIRICAL STUDY TO EVALUATE THE PROCEDURE

The Mechanics of the Empirical Study

If this procedure is going to be considered valid, it must work reasonably well when the traditional assumptions of normality and equality of variances are met, and it must perform better than the traditional $\bar{X}$ chart when these assumptions are violated. As a reference, traditional $\bar{X}$ charts were constructed. Using 10,000 replicates, $\bar{X}$ and R charts and $\bar{X}$ and s charts were constructed with 20 samples used in the initial phase. Charts were constructed for sample sizes of 4, 5, 6, and 7 from a normal distribution with $\mu = 0$ and $\sigma^2 = 1$. Charts were also constructed for samples from a uniform distribution, for samples from standard gamma distributions with $\alpha = 1, 2, 5, 8,$ and 15 and for samples from normal distributions with $\mu = 0$ and σ^2 ranging from 1 to 1.44, 1.96, 4, and 9.

Observations were generated using the random number generators available in Release 82.4 of SAS installed for use on the IBM 3081 at Virginia Commonwealth University. To generate observations from a standard normal distribution the RANNOR function was used; observations

from a uniform distribution were generated through the use of RANUNI; and observations from a standard gamma distribution used RANGAM. For each sample from a normal distribution with σ varying, a value for σ was generated using RANUNI. Then, observations were generated using RANNOR. Note that all observations within a sample come from the same normal distribution--the variance within a sample is not changing, but the variance is changing from sample to sample.

Using samples from the same distributions, both t (the combination of t_r and t_a) and t' (the combination of t_r' and t_a') were evaluated. Due to a limitation on CPU time, only 8000 replicates were used. First, the test to accept the alternative hypothesis was checked. At this stage, the stated and observed α levels were compared. Then the test to accept the null hypothesis was studied. The questions to be answered were 1) how fast can we estimate the mean and 2) is that estimate on target? Finally, the two tests were combined to test the proposed procedure.

Criterion for Evaluation

Since traditional $\bar{X}$ charts are only constructed when the accompanying chart for variability shows no points out of control, we first estimated the probability that a test

on the mean would be considered. The results are reported in Table A-2. These values could be interpreted as providing an estimate of the upper limit on the proportion of the time that a user will be able to assess the mean if traditional control chart procedures are used. As expected, lower values are associated with charts for observations drawn from distributions with positive kurtosis, and higher values are associated with charts for data from populations with negative kurtosis. For this study, X charts were constructed only when the chart for variability showed no points outside the control limits.

The first actual comparison of the procedures involves the accuracy of the test used to conclude that the process is out of control with respect to the mean. If the stated α value is 0.003 (the probability that, for any one point, we will conclude that the mean has shifted when, in reality, it has not shifted), how close is the observed (empirical) α value? The results are given in Table A-3.

The remainder of the comparisons deal with the acceptance of the null hypothesis--the estimation of the mean. Table A-4 reports the probability that the mean will actually be estimated. This represents the proportion of time that we would conclude that the process

is operating at a predictable level. The probabilities reported for traditional charts can be interpreted as the probability that both the $\bar{X}$ chart and the chart for variability will show no points outside the control limits. Table A-5 reports the standardized estimate of that level (the difference in the observed value and the true mean divided by the expected value of the standard deviation). Since the average of the observations is used to estimate the mean, the estimate for the mean should be unbiased if the testing procedure is symmetric. Table A-5 indicates that this is a reasonable conclusion. Table A-6 reports the ratio of the standard error of the estimate to the expected value of the standard deviation. For traditional $\bar{X}$ charts, it has been assumed that 20 samples were used to construct the charts. Therefore, the mean is estimated after 20 samples (provided the $\bar{X}$ chart shows no points outside the control limits). The median and expected number of iterations required by the new procedure are reported in Table A-7.

Evaluation of the Results

The initial comparisons are between traditional methods and the new procedure when the assumptions of normality and homogeneity of variance are met. Then, the

effect of unequal variances is studied. Finally, the effect of non-normality is studied. In all cases, the study is based on samples taken from populations that are in control with respect to the mean. The power of the test is not explored in this research.

Normal Population with Variance in Control. From Table A-4, we can see that traditional control charts provide an estimate of the mean approximately 86 to 90 percent of the time. In fact, since the chart for variability must be constructed and the variance estimated before the $\bar{X}$ chart is constructed, the traditional chart methods only begin to provide an estimate of the mean less than 95 percent of the time (Table A-2). In contrast, t and t′ provide estimates at least 97 percent of the time.

In terms of false signals (Type I errors), traditional methods and t perform equally. On the other hand, the observed α level for t′ tends to increase as the number of samples increases and is more sensitive at smaller sizes.

All of the procedures provide reasonable estimates of the mean (Table A-5), but the standard errors for t and t′ are almost double those of traditional methods (Table A-6). This is not really surprising. All four approaches use the mean of all observations to estimate the mean.

The standard error for the sample mean is given by σ divided by the square root of the number of observations. When traditional methods are used with samples from a standard normal distribution, the standard error of the mean is 0.112 for samples of size 4, 0.1 for samples of size 5, 0.091 for samples of size 6, and 0.085 for samples of size 7--note that the observed values are very close to the theoretical values. The new procedure makes an estimate based on fewer observations so the variability of that estimate is greater. Since at least 50 percent of the time estimates are made with seven or fewer samples (less than half as many observations), a corresponding increase in the standard error results. Consider the case of n=4. If estimates of the mean were always made after three samples, the standard error would be 0.289. This value decreases to 0.189 if estimates were always made after seven samples. In both the traditional methods and the new procedures, the standard error decreases as the sample size increases.

Normal Populations with Variance Out of Control. As the variability in the variance increases, the possibility of constructing $\bar{X}$ charts decreases. This speaks favorably for the charts for variability, but makes problem identification difficult. Even in cases where the $\bar{X}$ chart

is constructed, the probability that the mean will be estimated is lower than the corresponding probabilities for either t or t′.

In terms of false alarms, t′ appears to be more predictable as the variability and sample size increase. For small samples where the variability in the variance is small, t and traditional $\bar{X}$ charts are preferable.

The inevitable trade-off between when an estimate is made and the reliability of that estimate appears again. The new procedure, with both t and t′, continues to provide estimates of the mean with high probability, and these estimates still average requiring only about eight samples. Unfortunately, the estimates still exhibit about four times as much variability (a standard error twice as large) as exhibited by traditional methods.

Non-normal Populations. Even though the new procedure was not specifically designed to be robust to distributional assumptions, the effects will still be explored. Traditionally, means tests have been considered robust to distributional assumptions.

Traditional $\bar{X}$ charts have little chance with populations that have positive kurtosis. As kurtosis increases, the probability that a traditional chart for variability will not show points outside the control

limits decreases. For populations with negative kurtosis, charts for variability will rarely have points outside the control limits. Even when $\bar{X}$ charts are constructed, the difference in the observed α level and the stated level tends to increase as the kurtosis deviates from zero. This is logical since the normal theory limits on charts for variability are too wide for populations with negative kurtosis. This would tend to lead to inflated estimates of σ^2 which, in turn, would lead to larger than expected control limits on the $\bar{X}$ chart and fewer than expected signals on the $\bar{X}$ chart. Likewise, the normal theory limits on the $\bar{X}$ chart will be too narrow for populations with negative kurtosis. This helps explain the low observed α's for the samples from a uniform distribution and the high observed α's for the samples from gamma distributions.

For the uniform distribution, the new procedure using t seems to give predictable α values. The use of t' appears to produce extremely high observed α values. For gamma distributions, lower values for the shape parameter α correspond to more highly skewed distributions. In terms of false signals, none of the procedures behave predictably with the extreme deviations from normality.

Even in the absence of normality, t and t' have a high probability of providing an estimate of the mean. Again, the rapidity of the estimate is at the expense of the variability in that estimate.

Conclusions

In all cases studied the new procedure outperformed the traditional methods in terms of the number of iterations required to produce an estimate for the mean and in terms of the probability that an estimate would be made. This result was partially forced on the procedure through the choice of an "acceptance region" for the null hypothesis.

As a result of the decreased number of samples used to produce the estimate, the precision of that estimate was brought into question. Assuming that acceptance of the null hypothesis means establishing a value for the central line for continued monitoring, the increased variability in the estimate is likely to cause difficulty after the initial phase. This suggests that other criteria for accepting the null hypothesis should be considered.

CHAPTER 6

EXTENSIONS OF THE NEW PROCEDURE

Uses Outside of SPC

As we have seen in Chapter 3, means testing and estimation are not unique to the SPC field. Traditional tests have asked one of two questions--1) is the population mean equal to some specified value, or 2) are the population means of several populations equal to each other? This new procedure allows us to extend the last question. Now, we can ask, "Are the population means equal to each other; and if so, what is the value of that mean?"

When estimation is the goal, rather than testing, the new procedure can be used without the test to accept the alternative hypothesis. This approach could be compared to the sequential analysis methods of means testing suggested by Wald. Immediate applications of such a procedure should be available in the medical field and the chemical industry. In both cases, a forced decision by the twentieth sample may be too strict. This is easily modified by changing the critical region for the test.

Similar Procedures for Estimating and Testing Dispersion

In Chapter 2, several methods for testing variability were suggested. Some of these recommended using estimators other than s or s^2 since tests on these statistics are not robust to distributional assumptions. Several approaches based on transformations of data and subsequent use of ANOVA procedures were suggested.

Since the procedure introduced here is similar to the basic ANOVA procedures for two groups, it seemed reasonable to try it out on some of the spread statistics. Attempts were made to extend the sequential estimation and testing procedure to O'Brien's $r_{ij}(w)$ (p. 28), Levene's $z_{ij} = |x_{ij} - \bar{x}_j|$ and $z_{ij} = |x_{ij} - \tilde{x}_j|$ where $\tilde{x}_j$ is the median of the j^{th} group. In all cases, the attempt failed. Observed α levels were up to four times has high as the stated α levels. Even with moderate deviations from normality the test was unpredictable.

Possible Extension of this Research

There are two areas of study that require more attention before this procedure could be suggested for general use. First, the effect of changing the critical values for accepting the alternative or null hypothesis needs to be explored. I have suggested that the variability in the estimate could be decreased by

including more samples. Of course, this goes against the desire to obtain a "quick" estimate. One approach that would not rule out an early decision, would be to increase the number of samples that could be taken before a decision would be required. A procedure that averages a decision by the twentieth sample should perform comparably with traditional charts in terms of variability in the estimate. A second possibility would be to let the criterion for accepting the null hypothesis be determined by the sample size and the desired accuracy. Another approach would be to hold the probability of accepting the null hypothesis constant at each iteration and not fix a limit on when a decision must be reached. The behavior of this procedure could be studied through a Markov Chain approach.

A second area to be studied concerns the power of the test. How long does it take this procedure to recognize a shift? There are several difficulties with a study of this nature. First, a power study cannot be viewed as simply the development of an OC curve--we must now talk about a multidimensional figure. Any measure of power will be a function of the definition of power. Is power a measure of the ability of the procedure to recognize a shift on the first sample after the shift occurs or at

some other point in time? Second, since the sequential procedure pools information when the process is not classified as out of control, the ability to recognize shifts on future samples is reduced. Not only is the number of samples taken since the shift important, but the number of samples taken prior to the shift also affects the ability of the procedure to recognize the shift. Next, the size of the shift will affect the power of the test. Larger shifts are more likely to be recognized. And finally, what is meant by a shift must be considered. A shift may be a jump from one level to another, it may be a steady increase or decrease in the mean, or it may be a mean that behaves erratically.

The power of traditional control charts will also be affected by when the shift occurs, how large the shift is, and what is meant by a shift. Since traditional charts are not attempted until the initial group of samples is collected, the ability to recognize shifts is diluted because the information collected from samples from more than one distribution is included in the calculation of the control limits. Even if power studies could be conducted and interpreted for each procedure, comparisons would be difficult.

Conclusions

Without better reliability (less variability in the estimates of the mean), I would be skeptical about using estimates obtained by this method for continued monitoring. Both t and t′ require more computation than traditional control chart procedures with little benefit. The need for better variance testing methods has not been eliminated. Even though the new procedure overcomes the problem of making an estimate of the mean when the variance is out of control, the accuracy of the estimate obtained by this new procedure is still dependent on the variance or expected value of the variance (as any estimate based on the sample mean will be).

Until the power of the test is studied, I would recommend the new procedure be used only in the initial phase of variance testing. Referring back to Figure 2 (p. 5), this approach would allow the user to distinguish among Cases 1, 3, and 4. If the new method provided an estimate for the mean and the variance appeared to be in control after the initial phase of testing, we would assume that the proper classification would be Case 1. Since the variability in the sample mean is a function of the population variance, I would suggest that new limits,

including the central line, be set when the variance is estimated.

If the initial phase of testing indicates that the variance is out of control, the user can use the new procedure to distinguish between Cases 3 and 4. In either case the user would be looking for a special cause of variability, but the search for the possible cause(s) can be simplified by knowing if the trouble is only in the process variance or if the problem is affecting both the mean and the variance.

LIST OF REFERENCES

REFERENCES

Arvesen, J.N. and Schmitz, T.H. (1970), "Robust Procedures for Variance Component Problems Using the Jackknife," Biometrics, (26), 677-686.

Bartlett, M.S. (1937), "Properties of Sufficiency and Statistical Tests," Proceedings of the Royal Society of London, Series A, 160, 268-282.

_____. and Kendall, D.G. (1946), "The Statistical Analysis of Variance--Heterogeneity and the Logarithmic Transformation," Supplement to the Journal of the Royal Statistical Society, 8, 128-138.

Box, G.E.P. (1953), "Non-Normality and Tests on Variances," Biometrika, 40, 318-335.

Brown, M.B. and Forsythe, A.B. (1974), "Robust Tests for the Equality of Variance," Journal of the American Statistical Association, 69, 364-367.

Brown, M.B. and Forsythe, A.B. (1974b), "The Small Sample Behavior of Some Statistics Which Test the Equality of Several Means," Technometrics, 16, 129-132.

Burr, I.W. (1967), "The Effect of Non-Normality on Constants for $\bar{X}$ and R Charts," Industrial Quality Control, 23, 563-569.

_____. (1976), Statistical Quality Control Methods, New York: Marcel Dekker, Inc.

Cox, D.R. (1954), "The Mean and Coefficient of Variation of Range in Small Samples from Non-Normal Populations," Biometrika, 41, 469-481.

Dantzig, G.B. (1940), "On the Non-existence of Tests of "Student's" Hypothesis Having Power Functions Independent of σ," Annals of Mathematical Statistics, 11, 186-192.

David, H.A. (1954), "The Distribution of Range in Certain Non-Normal Populations," Biometrika, 41, 463-468.

Duncan, A.J. (1974), Quality Control and Industrial Statistics, Homewood, Illinois: Richard D. Irwin, Inc.

Ghosh, B.K. (1975), "A Two-Stage Procedure for the Behrens-Fisher Problem," Journal of the American Statistical Association, 70, 457-462.

Graybill, F.A. (1958), "Determining Sample Size for a Specified Width Confidence Interval," Annals of Mathematical Statistics, 29, 282-287.

Hillier, F.S. (1967), "Small Sample Probability Limits for the Range Chart," Journal of the American Statistical Association, 62, 1488-1493.

_____. (1969), "$\bar{X}$ and R Chart Control Limits Based on a Small Number of Subgroups," Journal of Quality Technology, 1, 17-26.

Hines, W.W. and Montgomery, D.C. (1980), Probability and Statistics in Engineering and Management Science, New York: John Wiley and Sons, Inc.

Kander, Z. and Zacks, S. (1966), "Test Procedures for Possible Changes in Parameters of Statistical Distributions Occurring at Unknown Time Points," The Annals of Mathematical Statistics, 37, 1196-1210.

King, E.P. (1954), "Probability Limits for the Average Chart When Process Standards Are Unspecified," Industrial Quality Control, 10, 62-64.

Kohr, R.L. and Games, P.A. (1977), "Testing Complex a Priori Contrasts on Means from Independent Samples," Journal of Educational Statistics, 2, 207-216.

Layard, W.M.J. (1973), "Robust Large Sample Tests for Homogeneity of Variances," Journal of the American Statistical Association, 68, 196-198.

Levene, H. (1960), "Robust Tests for the Equality of Variances," Contribution to Probability and Statistics, Palo Alto: Stanford University Press.

Lorden, G. (1971), "Procedures for Reacting to a Change in Distribution," The Annals of Mathematical Statistics, 42, 1897-1908.

Lucas, J.M. (1976), "The Design and Use of V-Mask Control Schemes," Journal of Quality Technology, 8, 1-12.

Miller, R.G. (1968), "Jackknifing Variances," The Annals of Mathematical Statistics, 39, 567-582.

Nelson, L.S. (1982), "Control Charts for Individual Measurements," Journal of Quality Technology, 14, 172-173.

_____. (1982b), "Control Charts for Medians," Journal of Quality Technology, 14, 226-227.

O'Brien, R.G. (1979), "A General ANOVA Method for Robust Tests of Additive Models for Variances," Journal of the American Statistical Association, 74, 877-880.

_____. (1981), "A Simple Test for Variance Effects in Experimental Designs," Psychological Bulletin, 89, 570-574.

Pearson, E.S. (1932), "The Percentage Limits for the Distribution of Range in Samples from a Normal Population," Biometrika, 24, 404-417.

_____. (1950), "Some Notes on the Use of Range," Biometrika, 37, 88-92.

_____. and Adyanthaya, N.K. (1928), "The Distribution of Frequency Constants in Small Samples from Symmetrical Populations," Biometrika, 20A, 356-360.

Pollak, M. (1985), "Optimal Detection of a Change in Distribution," The Annals of Statistics, 13, 206-227.

Rahim, M.A. (1985), "Economic Model of $\bar{X}$-Chart Under Non-Normality and Measurement Errors," Computers and Operations Research, 12, 291-299.

Stein, C. (1945), "A Two-Sample Test For a Linear Hypothesis Whose Power is Independent of the Variance," Annals of Mathematical Statistics, 16, 243-258.

Tippett, L.H.C. (1925), "On the Extreme Individuals and the Range of Samples Taken from a Normal Population," Biometrika, 17, 364-387.

Wald, A. (1947), Sequential Analysis, New York: John Wiley and Sons, Inc.

Wang, Y.Y. (1971), "Probabilities of the Type I Errors of the Welch Tests for the Behrens-Fisher Problem," Journal of the American Statistical Association, 66, 605-608.

Weiler, H. (1954), "A New Type of Control Chart Limits for Means, Ranges, and Sequential Runs," Journal of the American Statistical Association, 49, 298-314.

Weiss, L. (1955), "On Confidence Intervals of Given Length of the Mean of a Normal Distribution with Unknown Variance," Annals of Mathematical Statistics, 26, 348-352.

Welch, B.L. (1947), "The Generalization of Students' Problem When Several Different Population Variances Are Involved," Biometrika, 34, 28-35.

_____. (1951), "On the Comparison of Several Mean Values: An Alternative Approach," Biometrika, 39, 330-336.

Yang, C.-H. and Hillier, F.S. (1970), "Mean and Variance Control Chart Limits Based on a Small Number of Subgroups," Journal of Quality Technology, 2, 9-16.

APPENDIX

TABLE A-1

CONTROL CHART CONSTANTS FOR SAMPLES DRAWN FROM NORMAL POPULATIONS USED WHEN COMPUTING THREE SIGMA CONTROL LIMITS*

n	A_2	A_3	c_4	d_2	B_3	B_4	D_3	D_4
2	1.880	2.659	.798	1.128	--	3.267	--	3.267
3	1.023	1.954	.886	1.693	--	2.568	--	2.575
4	.729	1.628	.921	2.059	--	2.266	--	2.282
5	.577	1.427	.940	2.326	--	2.089	--	2.115
6	.483	1.287	.952	2.534	.030	1.970	--	2.004
7	.419	1.182	.959	2.704	.118	1.882	.076	1.924
8	.373	1.099	.965	2.847	.185	1.815	.136	1.864
9	.337	1.032	.969	2.970	.239	1.761	.184	1.816
10	.308	.975	.973	3.078	.284	1.716	.223	1.777

* Taken from Burr, I.W. (1976), Statistical Quality Control Methods, New York: Marcel Dekker, Inc.

TABLE A-2

ESTIMATES* FOR THE PROBABILITY THAT TRADITIONAL CHARTS FOR VARIABILITY WILL SHOW NO POINTS OUTSIDE CONTROL LIMITS

Distribution	Sample Size 4	5	6	7
Normal N(0,1)	.922 .929	.926 .936	.931 .942	.932 .946
Normal $1<\sigma<1.2$	.914 .921	.911 .922	.920 .930	.920 .931
Normal $1<\sigma<1.4$	.889 .896	.886 .892	.888 .899	.888 .895
Normal $1<\sigma<2$	.788 .796	.773 .777	.755 .759	.742 .736
Normal $1<\sigma<3$	.638 .637	.599 .596	.559 .545	.520 .502
Uniform (0,1)	1.000 1.000	1.000 1.000	1.000 1.000	1.000 1.000
Gamma $\alpha = 1$	.364 .328	.331 .305	.315 .291	.294 .267
Gamma $\alpha = 2$	.594 .557	.570 .535	.557 .527	.541 .518
Gamma $\alpha = 5$	.772 .752	.780 .764	.765 .763	.752 .753
Gamma $\alpha = 8$	.825 .813	.835 .832	.831 .826	.824 .829
Gamma $\alpha = 15$	.870 .865	.870 .875	.874 .878	.873 .878

* Standard error of the estimate < .005

x.xxx ⟵ When used with R Chart
o.ooo ⟵ When used with s Chart

TABLE A-3

OBSERVED SINGLE POINT α VALUES* WHEN NOMINAL α = .003

n = 4

Distribution	$\bar{X}(R)$	$\bar{X}(s)$	t k=2	t k=3	t k=4	t k=5	t' k=2	t' k=3	t' k=4	t' k=5
Normal N(0,1)	.003	.003	.003	.003	.003	.003	.002	.003	.007	.007
Normal $1<\sigma<1.2$	.003	.003	.003	.004	.002	.003	.002	.003	.005	.008
Normal $1<\sigma<1.4$	.004	.004	.004	.004	.003	.003	.002	.003	.005	.008
Normal $1<\sigma<2$	.005	.005	.003	.004	.005	.004	.002	.003	.006	.007
Normal $1<\sigma<3$	.008	.008	.003	.006	.007	.007	.002	.003	.005	.006
Uniform (0,1)	.001	.001	.005	.003	.002	.002	.005	.009	.014	.019
Gamma α = 1	.009	.008	.002	.020	.020	.016	.001	.003	.005	.014
Gamma α = 2	.007	.006	.003	.011	.008	.009	.002	.004	.007	.010
Gamma α = 5	.005	.005	.004	.004	.005	.006	.003	.004	.005	.009
Gamma α = 8	.004	.004	.003	.004	.005	.005	.001	.004	.006	.009
Gamma α = 15	.004	.004	.003	.003	.004	.003	.002	.004	.006	.006

TABLE A-3 (continued)

n = 5

Distribution	$\bar{X}(R)$	$\bar{X}(s)$	t k=2	t k=3	t k=4	t k=5	t′ k=2	t′ k=3	t′ k=4	t′ k=5
Normal N(0,1)	.003	.003	.003	.003	.002	.003	.002	.004	.006	.007
Normal $1<\sigma<1.2$	.003	.003	.003	.003	.003	.003	.003	.004	.005	.007
Normal $1<\sigma<1.4$	.003	.003	.002	.003	.004	.003	.002	.005	.005	.006
Normal $1<\sigma<2$	.005	.005	.004	.006	.004	.006	.002	.004	.005	.006
Normal $1<\sigma<3$	.008	.008	.005	.008	.007	.007	.003	.004	.005	.005
Uniform (0,1)	.002	.001	.005	.002	.002	.001	.004	.008	.012	.014
Gamma $\alpha = 1$	.007	.006	.001	.022	.021	.019	.001	.003	.006	.014
Gamma $\alpha = 2$	.006	.005	.002	.012	.011	.009	.002	.004	.007	.012
Gamma $\alpha = 5$	.004	.004	.002	.007	.006	.008	.001	.005	.008	.008
Gamma $\alpha = 8$	.004	.004	.003	.007	.006	.005	.002	.004	.005	.008
Gamma $\alpha = 15$	.003	.003	.003	.004	.005	.004	.002	.003	.005	.008

TABLE A-3 (continued)

n = 6

Distribution	$\bar{X}(R)$	$\bar{X}(s)$	t k=2	t k=3	t k=4	t k=5	t' k=2	t' k=3	t' k=4	t' k=5
Normal N(0,1)	.003	.003	.003	.002	.002	.003	.002	.003	.005	.006
Normal $1<\sigma<1.2$	.003	.003	.003	.003	.003	.004	.002	.002	.005	.007
Normal $1<\sigma<1.4$	.003	.003	.004	.003	.003	.003	.002	.002	.005	.006
Normal $1<\sigma<2$	.005	.005	.003	.004	.004	.006	.003	.003	.005	.006
Normal $1<\sigma<3$	.008	.008	.004	.007	.009	.007	.003	.003	.005	.005
Uniform (0,1)	.002	.001	.005	.001	.001	.001	.004	.006	.010	.011
Gamma $\alpha = 1$	.007	.005	.002	.024	.018	.017	.001	.002	.009	.013
Gamma $\alpha = 2$	.005	.005	.002	.011	.010	.010	.001	.002	.007	.014
Gamma $\alpha = 5$	.004	.004	.002	.006	.005	.005	.002	.003	.005	.007
Gamma $\alpha = 8$	.003	.003	.003	.004	.006	.005	.003	.005	.006	.007
Gamma $\alpha = 15$	.003	.003	.004	.003	.003	.004	.003	.004	.005	.006

TABLE A-3 (continued)

n = 7

Distribution	$\bar{X}(R)$	$\bar{X}(s)$	t k=2	t k=3	t k=4	t k=5	t' k=2	t' k=3	t' k=4	t' k=5
Normal N(0,1)	.003	.003	.003	.003	.003	.002	.002	.003	.003	.005
Normal $1<\sigma<1.2$	.003	.003	.003	.003	.003	.004	.002	.002	.004	.005
Normal $1<\sigma<1.4$	.003	.003	.004	.003	.003	.004	.002	.003	.004	.005
Normal $1<\sigma<2$	.005	.005	.003	.005	.005	.006	.003	.004	.004	.004
Normal $1<\sigma<3$	.008	.008	.003	.008	.008	.011	.002	.004	.004	.004
Uniform (0,1)	.003	.002	.005	.002	.002	.002	.004	.007	.009	.009
Gamma $\alpha = 1$	.006	.005	.002	.020	.017	.016	.001	.003	.008	.013
Gamma $\alpha = 2$	.005	.004	.002	.011	.009	.007	.002	.003	.007	.009
Gamma $\alpha = 5$	.004	.004	.002	.007	.007	.006	.002	.003	.006	.006
Gamma $\alpha = 8$	.003	.003	.003	.005	.004	.005	.002	.004	.006	.007
Gamma $\alpha = 15$	.003	.003	.003	.005	.004	.002	.003	.003	.005	.005

* Standard error of estimate < .002

TABLE A-4

ESTIMATES* FOR THE PROBABILITY THAT THE MEAN WILL BE ESTIMATED

n = 4

Distribution	Testing Procedure $\bar{X}(R)$	$\bar{X}(s)$	t	t'
Normal N(0,1)	.869	.876	.982	.981
Normal $1<\sigma<1.2$	.858	.865	.981	.976
Normal $1<\sigma<1.4$	.827	.834	.973	.976
Normal $1<\sigma<2$	.709	.714	.948	.969
Normal $1<\sigma<3$	.540	.539	.958	.958
Uniform (0,1)	.976	.980	.989	.989
Gamma $\alpha = 1$	.307	.281	.925	.924
Gamma $\alpha = 2$	.520	.490	.952	.950
Gamma $\alpha = 5$	.689	.684	.971	.967
Gamma $\alpha = 8$	.761	.752	.970	.969
Gamma $\alpha = 15$	.810	.807	.971	.974

TABLE A-4 (continued)

n = 5

Distribution	Testing Procedure $\bar{X}(R)$	$\bar{X}(s)$	t	t′
Normal N(0,1)	.880	.889	.976	.979
Normal $1<\sigma<1.2$	.862	.873	.971	.976
Normal $1<\sigma<1.4$	.831	.838	.975	.971
Normal $1<\sigma<2$	.695	.697	.962	.967
Normal $1<\sigma<3$	.509	.507	.950	.951
Uniform (0,1)	.968	.977	.989	.992
Gamma $\alpha = 1$	.287	.271	.930	.928
Gamma $\alpha = 2$	.510	.483	.951	.954
Gamma $\alpha = 5$	.716	.703	.969	.968
Gamma $\alpha = 8$	.776	.775	.972	.968
Gamma $\alpha = 15$	.817	.816	.974	.971

TABLE A-4 (continued)

n = 6

Distribution	Testing Procedure $\bar{X}(R)$	$\bar{X}(s)$	t	t′
Normal N(0,1)	.886	.895	.979	.979
Normal $1<\sigma<1.2$	.871	.883	.979	.976
Normal $1<\sigma<1.4$	.833	.844	.972	.970
Normal $1<\sigma<2$	.678	.683	.961	.965
Normal $1<\sigma<3$	.479	.461	.948	.946
Uniform (0,1)	.960	.973	.986	.987
Gamma $\alpha = 1$	.277	.261	.933	.935
Gamma $\alpha = 2$	.503	.481	.942	.945
Gamma $\alpha = 5$	.705	.709	.961	.963
Gamma $\alpha = 8$	.776	.773	.965	.964
Gamma $\alpha = 15$	.823	.828	.971	.972

TABLE A-4 (continued)

n = 7

Distribution	Testing Procedure $\bar{X}(R)$	$\bar{X}(s)$	t	t'
Normal N(0,1)	.884	.899	.981	.970
Normal $1<\sigma<1.2$	.867	.883	.977	.977
Normal $1<\sigma<1.4$	.833	.842	.974	.980
Normal $1<\sigma<2$	.669	.665	.963	.966
Normal $1<\sigma<3$	.444	.429	.934	.947
Uniform (0,1)	.949	.971	.980	.984
Gamma $\alpha = 1$	.260	.246	.928	.928
Gamma $\alpha = 2$	.491	.477	.956	.952
Gamma $\alpha = 5$	.697	.702	.968	.961
Gamma $\alpha = 8$	.769	.779	.972	.976
Gamma $\alpha = 15$	.825	.833	.969	.972

* Standard error of the estimate < .005

TABLE A-5

STANDARDIZED ESTIMATES FOR THE MEAN

n = 4

Distribution	$\bar{X}(R)$	$\bar{X}(s)$	t	t′
Normal N(0,1)	.000	.001	-.004	-.001
Normal $1<\sigma<1.2$	.000	.001	.003	-.006
Normal $1<\sigma<1.4$	.001	.001	.001	.003
Normal $1<\sigma<2$	.001	.001	.004	.003
Normal $1<\sigma<3$	-.001	.000	-.002	-.001
Uniform (0,1)	.003	.003	.007	.003
Gamma $\alpha = 1$	.000	.001	.013	.014
Gamma $\alpha = 2$	-.001	.000	.009	.010
Gamma $\alpha = 5$	.000	.000	.011	.007
Gamma $\alpha = 8$	.001	.001	.003	.001
Gamma $\alpha = 15$	.000	-.002	-.003	-.002

TABLE A-5 (continued)

n = 5

Distribution	$\bar{X}(R)$	$\bar{X}(s)$	t	t′
Normal N(0,1)	-.002	-.002	.000	-.002
Normal $1<\sigma<1.2$	-.002	-.002	-.004	-.007
Normal $1<\sigma<1.4$	-.002	-.001	-.006	-.006
Normal $1<\sigma<2$	-.001	-.002	-.001	-.004
Normal $1<\sigma<3$	-.001	.000	-.007	-.009
Uniform (0,1)	.000	.000	-.003	.003
Gamma $\alpha = 1$	-.003	-.004	.027	.025
Gamma $\alpha = 2$	.001	.001	.018	.015
Gamma $\alpha = 5$	-.001	-.001	.013	.010
Gamma $\alpha = 8$	.000	.000	.007	.009
Gamma $\alpha = 15$	-.001	.002	.009	.016

TABLE A-5 (continued)

n = 6

Distribution	$\bar{X}(R)$	$\bar{X}(s)$	t	t′
Normal N(0,1)	-.001	-.001	-.005	-.003
Normal $1<\sigma<1.2$	.001	-.001	.002	-.004
Normal $1<\sigma<1.4$	-.001	-.001	-.002	-.015
Normal $1<\sigma<2$	.000	.000	-.003	.003
Normal $1<\sigma<3$	-.001	.001	-.002	-.002
Uniform (0,1)	.000	-.003	-.007	-.003
Gamma $\alpha = 1$	.000	.000	.025	.027
Gamma $\alpha = 2$	.000	-.003	.019	.015
Gamma $\alpha = 5$	.001	-.001	.017	.013
Gamma $\alpha = 8$	.001	-.001	-.001	-.001
Gamma $\alpha = 15$	.002	-.001	.004	.007

TABLE A-5 (continued)

n = 7

Distribution	$\bar{X}(R)$	$\bar{X}(s)$	t	t′
Normal N(0,1)	.000	-.002	-.001	-.003
Normal $1<\sigma<1.2$	-.001	-.002	.005	.009
Normal $1<\sigma<1.4$	-.002	-.001	.003	.011
Normal $1<\sigma<2$	-.001	-.001	.002	.001
Normal $1<\sigma<3$	-.002	-.001	.003	.004
Uniform (0,1)	.000	.000	.000	.003
Gamma $\alpha = 1$	-.002	-.003	.024	.026
Gamma $\alpha = 2$	-.001	-.001	.024	.023
Gamma $\alpha = 5$	.000	-.001	.020	.012
Gamma $\alpha = 8$	.000	-.001	.007	.008
Gamma $\alpha = 15$	-.002	-.002	.004	.004

TABLE A-6

RATIO OF THE OBSERVED STANDARD ERROR TO THE EXPECTED STANDARD DEVIATION FOR THE POPULATION

n = 4

Distribution	$\bar{X}(R)$	$\bar{X}(s)$	t	t′
Normal N(0,1)	.111	.110	.206	.209
Normal $1<\sigma<1.2$	.112	.111	.207	.211
Normal $1<\sigma<1.4$	.112	.112	.209	.212
Normal $1<\sigma<2$	.113	.113	.209	.215
Normal $1<\sigma<3$	.116	.116	.217	.221
Uniform (0,1)	.111	.111	.215	.211
Gamma $\alpha = 1$	.111	.112	.217	.221
Gamma $\alpha = 2$	.111	.111	.210	.209
Gamma $\alpha = 5$	.112	.112	.211	.214
Gamma $\alpha = 8$	.112	.113	.212	.211
Gamma $\alpha = 15$	.112	.111	.206	.206

TABLE A-6 (continued)

n = 5

Distribution	$\bar{X}(R)$	$\bar{X}(s)$	t	t′
Normal N(0,1)	.100	.100	.197	.188
Normal $1<\sigma<1.2$	.100	.100	.193	.192
Normal $1<\sigma<1.4$	.100	.100	.200	.200
Normal $1<\sigma<2$	.101	.102	.200	.197
Normal $1<\sigma<3$	.103	.104	.205	.198
Uniform (0,1)	.100	.100	.191	.191
Gamma $\alpha = 1$	.100	.100	.204	.202
Gamma $\alpha = 2$	.100	.100	.202	.202
Gamma $\alpha = 5$	.100	.100	.196	.197
Gamma $\alpha = 8$	.099	.099	.188	.196
Gamma $\alpha = 15$	.100	.100	.195	.193

TABLE A-6 (continued)

n = 6

Distribution	$\bar{X}(R)$	$\bar{X}(s)$	t	t′
Normal N(0,1)	.092	.092	.177	.181
Normal $1<\sigma<1.2$	.092	.093	.181	.182
Normal $1<\sigma<1.4$	.092	.093	.178	.176
Normal $1<\sigma<2$	.095	.094	.176	.183
Normal $1<\sigma<3$	.098	.098	.183	.183
Uniform (0,1)	.090	.090	.180	.180
Gamma $\alpha = 1$	.092	.092	.190	.186
Gamma $\alpha = 2$	.091	.092	.185	.187
Gamma $\alpha = 5$	.091	.090	.180	.180
Gamma $\alpha = 8$	.091	.091	.179	.181
Gamma $\alpha = 15$	.091	.092	.179	.179

TABLE A-6 (continued)

n = 7

Distribution	$\bar{X}(R)$	$\bar{X}(s)$	t	t′
Normal N(0,1)	.085	.084	.167	.172
Normal $1<\sigma<1.2$	.085	.084	.170	.166
Normal $1<\sigma<1.4$	.085	.084	.169	.163
Normal $1<\sigma<2$	.087	.086	.170	.175
Normal $1<\sigma<3$	.088	.088	.169	.171
Uniform (0,1)	.083	.083	.166	.166
Gamma $\alpha = 1$	.085	.086	.167	.165
Gamma $\alpha = 2$	.086	.085	.166	.166
Gamma $\alpha = 5$	.085	.084	.170	.170
Gamma $\alpha = 8$	.083	.083	.164	.164
Gamma $\alpha = 15$	.085	.085	.166	.168

TABLE A-7

MEDIAN AND EXPECTED NUMBER OF ITERATIONS
TO ESTIMATE THE MEAN USING THE PROPOSED PROCEDURE

	t			
	Sample Size			
Distribution	4	5	6	7
Normal N(0,1)	7.000 8.098	7.000 8.011	7.000 8.149	7.000 8.160
Normal $1<\sigma<1.2$	7.000 8.137	7.000 8.109	7.000 8.080	7.000 8.054
Normal $1<\sigma<1.4$	7.000 8.081	7.000 8.054	7.000 8.166	7.000 8.024
Normal $1<\sigma<2$	7.000 8.049	7.000 8.199	7.000 8.241	7.000 8.111
Normal $1<\sigma<3$	7.000 8.142	7.000 7.963	7.000 8.182	7.000 8.221
Uniform (0,1)	7.000 8.037	7.000 7.855	7.000 8.126	7.000 8.253
Gamma $\alpha = 1$	7.000 8.104	7.000 8.171	7.000 8.266	7.000 8.202
Gamma $\alpha = 2$	7.000 7.840	7.000 8.044	7.000 8.227	7.000 8.091
Gamma $\alpha = 5$	7.000 7.848	7.000 8.084	7.000 8.080	7.000 8.133
Gamma $\alpha = 8$	7.000 8.057	7.000 8.170	7.000 8.116	7.000 8.082
Gamma $\alpha = 15$	7.000 8.194	7.000 8.012	7.000 8.045	7.000 8.177

TABLE A-7 (continued)

t'

Distribution	Sample Size 4	5	6	7
Normal N(0,1)	7.000	7.000	7.000	7.000
	8.091	8.132	8.120	8.028
Normal $1<\sigma<1.2$	7.000	7.000	7.000	7.000
	7.975	8.218	8.199	8.163
Normal $1<\sigma<1.4$	7.000	7.000	7.000	7.000
	8.073	8.076	8.122	8.023
Normal $1<\sigma<2$	7.000	7.000	7.000	7.000
	7.978	8.116	8.236	7.922
Normal $1<\sigma<3$	7.000	7.000	7.000	7.000
	8.023	8.248	8.229	8.251
Uniform (0,1)	7.000	7.000	7.000	7.000
	8.045	7.864	8.163	8.247
Gamma $\alpha = 1$	7.000	7.000	7.000	7.000
	8.109	8.228	8.224	8.133
Gamma $\alpha = 2$	7.000	7.000	7.000	7.000
	7.884	8.076	8.125	8.043
Gamma $\alpha = 5$	7.000	7.000	7.000	7.000
	7.874	8.080	8.046	8.058
Gamma $\alpha = 8$	7.000	7.000	7.000	7.000
	8.018	8.052	8.179	8.127
Gamma $\alpha = 15$	7.000	7.000	7.000	7.000
	8.097	8.136	8.018	8.180

x.xxx ⟵ Median
o.ooo ⟵ Expected number of iterations

VITA

Kim Ingrid Melton was born in Macon, Georgia on April 19, 1956. She attended elementary school in that city and was graduated from Central High School in June 1974. The following September she entered Florida State University, and she received an Associate of Arts degree in March 1976. In spring of 1976 she entered The University of Georgia. In June 1978 she graduated (magna cum laude) with a Bachelor of Science in Education degree with a major in Parks and Recreation Administration.

After two years as a field director with Tanasi Girl Scout Council, Inc. in Knoxville, Tennessee, she entered The Graduate School of The University of Tennessee, Knoxville, in September 1980. During her graduate studies whe was employed as a graduate teaching assistant and later as an instructor in the Management Science Program. She was awarded the Master of Science degree in Management Science in March 1982. She received the Doctor of Philosophy degree with a major in Management Science in June 1986.

The author is a member of Phi Kappa Phi, the Operations Research Society of America, the American Statistical Association, the Decision Sciences Institute,

and the American Camping Association. She is currently employed in the Department of Business Administration and Management, School of Business, Virginia Commonwealth University.