Improved and robust monitoring in statistical process control

Riaz, M.

Citation for published version (APA):
CHAPTER 2

IMPROVED MONITORING OF THE PROCESS
LOCATION PARAMETER

In this chapter, use of auxiliary information is introduced for improved monitoring of the process location parameter with respect to a single quality characteristic of interest, say $Y$. A proposal is made for the location parameter in the form of a Shewhart type control chart which uses information of an auxiliary characteristic, say $X$, on the regression pattern. Assuming bivariate normality of $(Y, X)$, the design structure of the proposed control chart is developed for Phase-I quality control, targeting moderate and large shifts which is of major concern of Shewhart type control charts. Using power curves (which are performance measures for Phase-I quality control) as a performance measure, the comparisons of the proposed charts are made with some existing control charts used for the same purposes. Finally, an illustrative example is added to explain the procedure of the proposed control chart. This chapter is based on Riaz (2008a).

2.1 Introduction

The monitoring of any process output demands an early detection of shifts in the process parameters. The shift may be in the process variability or the process mean level or both. The variability of any process is controlled first followed by controlling of the mean level. In the 1920s, Walter A. Shewhart introduced the idea of control charts to monitor any process for variability or process mean level. The
commonly used control charts for monitoring process variability are the $R$ chart, the $S$ chart, the $S^3$ chart and for the process mean level are the $\bar{X}$ chart, the median chart, the trimmed mean chart, and the mid-range chart.

For an improved monitoring of a quality characteristic of interest, the idea of exploiting correlation of the characteristic of interest with some other associated quality characteristic(s) has been used in the form of cause-selecting and regression-adjusted control charts, see e.g. Mandel (1969), Zhang (1984, 1985), Hawkins (1991, 1993), Wade and Woodall (1993), and Shu et al. (2005). The cause-selecting and regression-adjusted control charts are constructed for a quality characteristic of interest after adjusting for the effect of some associated characteristic(s) (i.e. the residuals are obtained and used for monitoring the quality characteristic(s) of interest).

The concept of using information on some auxiliary characteristics is quite popular in different fields. The dictionary meaning of the word ‘auxiliary’ is: “An individual or group that assists or functions in a supporting capacity”. By identifying and observing some auxiliary variables along with the variable of interest, the information on the relationship between auxiliary variables and the variable of interest can be used to improve the precision with which parameters are estimated. According to Singh and Mangat (1996), the prior information on an auxiliary variable can be used to enhance the precision of an estimator.

There are different objectives for which researchers have made use of information on auxiliary characteristics. To refer a few of these: Olkin (1958), Raj (1965), Srivastava (1965, 66), Rao and Mudholkar (1967), and Adhvaryu (1975) used multi-auxiliary supplementary information for estimating the population mean. Tikkiwal (1960) used the information on a single auxiliary variable for the estimation of the mean and also considered the case when the variable of interest and the

There are different ways of taking benefit from information on auxiliary variables, out of which the Ratio and Regression methods are the most popular. A variety of literature is available where researchers have made use of auxiliary information on Ratio and Regression patterns. To refer a few of these: Chand (1975) discussed some ratio type estimators using two or more auxiliary characteristics. Fuller (1980) comments on a ratio estimator with respect to its smaller approximate variance. Royall and Cumberland (1981) provided some empirical results for a ratio estimator. Kiregyera (1984) used two auxiliary variables for regression type estimators. Mukerjee et al. (1987) and Ahmed (1998) discussed regression type estimators using multiple auxiliary information. Prasad (1989) introduced some improved ratio type estimators for the population mean. Singh and Gangele (1989-95) discussed an improved estimator with known coefficient of variation using two auxiliary variables. Reddy (1993) discussed product and ratio estimation procedures. Sahoo and Sahoo (1999) compared some regression type estimators in double sampling procedures. Singh (2001) used transformed auxiliary variables for estimating the population mean in two-phase sampling. Magnus (2002) introduced an estimator for the mean of a normal distribution in a regression context.
Hawkins (1993) concludes in his paper that i) Multivariate control charts capitalize on the correlation between different correlated characteristics at the cost of losing simplicity, ii) Individual univariate control charts provide better visual picture and clues for process improvement at the cost of missing out the possibility of capitalizing on the correlation between different correlated characteristics, iii) Quadratic form based control charts exploit the correlation between different correlated quality characteristics at the cost of losing the benefit of interpretability and performance loss. He proposed an alternative approach which overcomes the aforementioned three problems keeping their respective benefits. He argued about his approach as: “In this approach, information on the dynamics of the process itself is used to uncover the directions of causality giving rise to the correlations in the data. This diagnosis in turn leads to prescriptions for the regression adjustment of different variables”. He classified the processes into two main categories i) Cascade: “where each variable that changes the distribution affects those below but not above it”, ii) Without Cascade: “where each variable may undergo a distributional change without affecting any others”.

The inspirations and directions of this chapter and the next chapter (i.e. Chapter 3) are taken from Hawkins (1993). For processes without cascade property, information about an auxiliary characteristic $X$ is introduced for improved monitoring of process location and variability parameters with respect to a quality characteristic of interest $Y$.

In this chapter, the information about a single auxiliary characteristic $X$ is introduced for improved monitoring of a process location parameter of a quality characteristic of interest $Y$. Assuming bivariate normality of $(Y, X)$, a Shewhart type
process location control chart (which is a threshold control chart) is proposed for Phase-I quality control.

In the following sections, i) the design structure of the proposed chart is developed for improved process monitoring following the pioneering work of Shewhart (1931), Pearson (1932), Pappanastos and Adams (1996), Ramalhoto and Morais (1999), Gonzalez and Viles (2000 and 2001), ii) the power curves are constructed as a performance measure of the proposed chart following Scheffe (1949), Duncan (1951), Nelson (1985), iii) the comparisons of the design structure of the proposed chart is made with some existing control charts (like Shewhart’s well-known $\bar{X}$ chart, cause-selecting and regression-adjusted charts) used for the same purposes following Tuprah and Ncube (1987), Acosta-Mejia et al. (1999), Ding et al. (2005) and iv) an illustrative example is also given to explain the working of the proposed control chart.

2.2 The Proposed Control Chart (the $M_r$ Chart)

In this section, the information about a single auxiliary characteristic $X$ is introduced for the improved monitoring of the process mean level of a quality characteristic of interest $Y$. Assuming bivariate normality of $(Y, X)$, a Shewhart type process mean control chart, namely the $M_r$ chart (a threshold control chart), is proposed which is based on the regression estimator of the process mean level. The focus of the proposed chart is on Phase-I quality control. The regression estimator for the mean of $Y$, using a single auxiliary variable $X$, is defined for a bivariate random sample $(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)$ of size $n$ as:

$$M_r = \bar{y} + b(\mu_x - \bar{x}),$$

(2.1)
where $\bar{Y}$ is the sample mean of $Y$, $\bar{X}$ is the sample mean of $X$, $\mu$ is the population mean of $X$ (assumed to be known) and $b$ is defined as:

$$b = r_{yx}(s_x/s_y),$$

(2.2)

where $r_{yx}$ is the sample correlation coefficient between $Y$ and $X$, $s_x$ is the sample standard deviation of $X$ and $s_y$ is the sample standard deviation of $Y$.

Assuming bivariate normality of $(Y, X)$ a relationship between $\mu_y$ (the unknown process mean of the quality characteristic of interest $Y$ which has to be monitored) and $M_r$ (the regression estimator of $\mu_y$ defined in (2.1)) is required to develop the structure of the $M_r$ chart. Let $(y_1, x_1), (y_2, x_2), ..., (y_n, x_n)$ be a bivariate random sample of size $n$ from a bivariate normal distribution, and let $C$ be the random variable that defines a relationship between $\mu_y$ and $M_r$ as:

$$C = \sqrt{n(M_r - \mu_y)} / \sigma_y,$$

(2.3)

where $\sigma_y$ is the standard deviation of $Y$, which has already been controlled using some variability control chart (e.g. the $R$ chart, or the $S$ chart). The relationship defined in (2.3) helps determining the parameters (i.e. centerline, lower and upper control limits) of the proposed $M_r$ chart.

Now if the distributional behavior of $C$ is known then the sample statistic $M_r$ can easily be used for the testing of hypotheses about shifts in $\mu_y$. When $(Y, X)$ follow bivariate normal distribution, the distributional behavior of $C$ depends only on $\rho_{yx}$ (the correlation between $Y$ and $X$) and $n$. The distributional behavior of $C$, in terms of its mean, standard error and quantile points, is required for the development
of the $M_r$ chart, and is explored in the following paragraphs when $(Y, X)$ follow a
bivariate normal distribution.

First for the mean, applying expectations to (3) gives:
\[
E(C) = E\left(\sqrt{n}(M_r - \mu_y) / \sigma_y\right) = \sqrt{n}E(M_r - \mu_y) / \sigma_y.
\] (2.4)

Here $E(M_r)$ can safely be replaced by its estimate $\bar{M}_r$ (the mean of the sample
$M_r$’s) using an appropriate number of random samples, as discussed in Hillier
(1969) and Yang and Hillier (1970), from the process under study when the process is
in the state of statistical control (just like $\bar{R}$ replaces the $E(R)$ for the $R$-Chart). Thus
from (2.4) an estimate of $\mu_y$, after rearranging and simplifying the terms, is given as:
\[
\hat{\mu}_y = \bar{M}_r - \hat{\sigma}_y E(C) / \sqrt{n}.
\] (2.5)

The regression estimator $M_r$ is generally a biased estimator of the population
mean $\mu_y$, but the bias vanishes when the relationship between $Y$ and $X$ is linear (see
Sukhatme and Sukhatme (1984, p. 238)). For the case of bivariate normal $(Y, X)$,
$M_r$ is unbiased for $\mu_y$ and hence $E(C) = 0$. Thus (2.5) results into the following:
\[
\hat{\mu}_y = \bar{M}_r.
\] (2.6)

Also from (2.4) we have:
\[
E(M_r) = \mu_y.
\] (2.7)

Replacing the estimate of $\mu_y$ (given in (2.6)) in (2.7) gives:
\[
E(M_r) \approx \bar{M}_r.
\] (2.8)

Secondly, for the standard error, let the standard deviation of $C$ (i.e. $\sigma_c$) be
\[
\sigma_c = k_2.
\] (2.9)
It is not easy to get the analytical results for $k_2$ because $E(M_r^2)$ is difficult to obtain analytically. Note that approximations for $k_2$ exist (see Sukhatme and Sukhatme (1984, p. 267) and (2.14)). So simulation results are obtained for $k_2$ (in practice, simulation methods are often used to evaluate the expectation of a statistic, see Ross (1990)). The coefficient $k_2$ depends entirely on $\rho_{yx}$ and $n$, in the case of a bivariate normal distribution. Using 10,000 random samples generated from a standard bivariate normal distribution, without loss of generality, the results of $k_2$ have been obtained, for different combinations of $\rho_{yx}$ and $n$, 1000 times each (for a detailed discussion regarding the number of simulations needed in control chart studies, see Schaffer and Kim (2007)). Based on these results, the mean values of $k_2$, along with their respective standard errors, are provided in Table 1, given in Riaz (2008a), for $n = 5, 6, \ldots, 15, 20, 25, 30, 50, 100$ at some representative values of $\rho_{yx}$. Similar results can easily be obtained for any combination of $\rho_{yx}$ and $n$.

Also taking the variance of $C$ and then simplification gives the expression for $\sigma_c$ as:

$$\sigma_c = \sqrt{n} \sigma_{M_r} / \sigma_y,$$

(2.10)

where $\sigma_{M_r}$ represents the standard deviation of the distribution of the sample statistic $M_r$.

Using (2.9) in (2.10) and rearranging yields the following result for $\sigma_{M_r}$:

$$\sigma_{M_r} = k_2 \sigma_y / \sqrt{n}.$$

(2.11)

Substituting the estimate for $\sigma_y$ in (2.11), the estimate for $\sigma_{M_r}$ is given as:

$$\hat{\sigma}_{M_r} = k_2 \hat{\sigma}_y / \sqrt{n}.$$

(2.12)
where \( \hat{\sigma}_y \) is the controlled estimate of the process standard deviation \( \sigma_y \) which can be obtained from some variability control chart.

An approximation for \( \sigma_{M_r} \), when \((Y, X)\) follow a bivariate normal distribution, is given as (see Sukhatme and Sukhatme (1984, p. 267)):

\[
\sigma_{M_r} \approx \sqrt{\sigma_y^2(1-\rho_{yx})^2(1+1/(n-3))}/n .
\] (2.13)

Consequently

\[
k_2 \approx \sqrt{1-\rho_{yx}^2}(1+1/(n-3)) .
\] (2.14)

Asymptotically, and even for small values of \( n \), the approximation (2.14) works well, as can be concluded from Table 1, given in Riaz (2008a).

Lastly, for the quantile points of the distribution of \( C \), let \( C_a \) represent the \( a^{th} \) quantile point of the distribution of \( C \) (i.e. the point where \( C \) completes \( a \% \) area). Analytical results for \( C_a \) are difficult to obtain so simulation results are obtained for \( C_a \). For a bivariate normal distribution of \((Y, X)\), the quantile points of the distribution of \( C \) depend entirely on \( \rho_{yx} \) and \( n \). Using the same 10,000 simulated random samples, the results of \( C_a \) have been obtained (like the quantile points of \( W = R/\sigma \) that determine the values of control limits of the \( R \) chart and the power of the chart) for different combinations of \( \rho_{yx} \) and \( n \), 1000 times each. Based on these results, the mean values of some commonly used quantile points, along with their respective standard errors, are provided for \( n = 5, 6, \ldots, 15, 20, 25, 30, 50, 100 \) in Tables 2−11, given in Riaz (2008a), at some representative values of \( \rho_{yx} \). Similar results can easily be obtained for any combination of \( \rho_{yx} \) and \( n \). These quantile points help determining the control limits and the power of the proposed \( M_r \) chart to detect shifts in the process mean level for the quality characteristic of interest \( Y \). The distributional
behavior of $C$ is not symmetrical, at least for small values of $n$, as is obvious from Tables 2–11, given in Riaz (2008a). Asymptotically, $C$ is normally distributed as $N(0,(1 - \rho^2_{yx})(1 + 1/(n - 3)))$.

2.2.1 Parameters of the Proposed $M_r$ Chart

The Central Line ($CL$), Lower Control Limit ($LCL$) and Upper Control Limit ($UCL$) are the three parameters of any Shewhart type control chart. There are two approaches to express these parameters namely the probability limits approach and the 3-sigma limits approach. In case of an asymmetric distributional behavior of the relevant estimator, the probability limits approach is preferred. If the distributional behavior of the relevant estimator is nearly symmetric then the 3-sigma limits approach is a good alternative. The parameters of the proposed $M_r$ chart using both approaches are expressed in the following two subsections.

2.2.1.1 Probability Limits Approach

The value $\bar{M}_r$ corresponds to $CL$ of the proposed $M_r$ chart, just like $\bar{R}$ for the $R$ chart provided in Alwan (2000, p. 347) and $\bar{S}$ for the $S$ chart provided in Alwan (2000, p. 362). Assuming the probability of making a Type-I error to be less than a specified value say $\alpha$, the control limits (which are actually the true probability limits) for the proposed $M_r$ chart are defined as:

$$LCL = M_{\eta_i} \text{ with } P_n \left( M_r = M_{\eta_i} \right) \leq \alpha,$$

$$UCL = M_{\eta_u} \text{ with } P_n \left( M_r = M_{\eta_u} \right) \geq 1 - \alpha,$$

(2.15)

where $M_{\eta_i}$ and $M_{\eta_u}$ are the quantile points of the distribution of $M_r$ below which the
areas are $\alpha_i$ and $1-\alpha_u$ respectively, $\alpha = \alpha_i + \alpha_u$ and $P_n(X = x)$ represents the cumulative distribution function of $X$ at point $x$, for a given value of $n$.

Now using (2.3) and (2.6) in (2.15) and simplification gives the following:

$$\begin{align*}
LCL &= \bar{M}_o + C_i \hat{\sigma}_y / \sqrt{n} \quad \text{with} \quad P_n(C = C_i) \leq \alpha_i, \\
UCL &= \bar{M}_u + C_u \hat{\sigma}_y / \sqrt{n} \quad \text{with} \quad P_n(C = C_u) \geq 1 - \alpha_u,
\end{align*}$$

(2.16)

Thus the quantile points of the distribution of $C$, the average of sample $M_r$’s (i.e. $\bar{M}_r$) and $\hat{\sigma}_y$ (a controlled estimate of the process standard deviation $\sigma_y$, available from some process variability control chart like the $R$ chart, or the $S$ chart) allow setting the true probability limits for the proposed $M_r$ chart. These results are based on the observed data set of say $m$ subgroups each of a fixed size say $n$.

### 2.2.1.2 Three-Sigma Limits Approach

If the normal approximation to the distribution of $C$ is used then the parameters of $M_r$ chart with the usual 3-sigma control limits are given as:

$$\begin{align*}
LCL &= \bar{M}_r - 3 \sigma_{M_r}, \\
CL &= \bar{M}_r, \\
UCL &= \bar{M}_r + 3 \sigma_{M_r},
\end{align*}$$

(2.17)

Using (2.12) in (2.17) gives the following result:

$$\begin{align*}
LCL &= \bar{M}_r - 3k_2 \hat{\sigma}_y / \sqrt{n} \\
CL &= \bar{M}_r \\
UCL &= \bar{M}_r + 3k_2 \hat{\sigma}_y / \sqrt{n}
\end{align*}$$

(2.18)

where the values of $k_2$ are provided in Table 1, given in Riaz (2008a).

The validity of these 3-sigma limits based parameters of the proposed $M_r$ chart depends on how close the normal approximation is to the true distribution of $C$.
After deciding the control structure, for given significance level, by either the probability limit approach or the 3-sigma limit approach, the sample statistic $M_r$ is plotted against time order of the samples. If all the sample $M_r$’s lie within control limits, there is reasonable evidence to conclude that there is no shift in the process mean level and the process is stable at $M_r$. Otherwise, some assignable cause(s) are at work, causing a shift in the process mean level.

To address small and moderate shifts, using the developed structure of the $M_r$ chart, the runs rules (as discussed by Nelson (1984), Wheeler (1995) and Does and Schriever (1992)) may be supplemented to its basic structure. As a result, the risk of false alarms is increased.

2.2.2 Comparisons of the Proposed $M_r$ Chart

In this section, some comparisons of the proposed $M_r$ chart are made with the well-known Shewhart $\bar{X}$ chart (as the characteristic of interest is denoted by $Y$ in this study so the name $\bar{Y}$ chart will be used instead of $\bar{X}$ chart later in this chapter) and the cause-selecting charts of Zhang (1984) and Wade and Woodall (1993). As the focus of the proposal is on Phase-I quality control, so power curves have been used as a performance measure (in contrast to Phase-II quality control where the Average Run Length (ARL) is used as a performance measure) of the control charts following Albers and Kallenberg (2006).

2.2.2.1 $M_r$ Chart vs. $\bar{Y}$ Chart

Using the parameters for the $M_r$ and $\bar{Y}$ charts, as given in (2.16) and Alwan (2000, p.394) respectively, the performance of the $M_r$ chart has been compared with
that of the $\bar{Y}$ chart in terms of discriminatory power. As the distributional behavior of $C$ is not symmetrical, at least for small values of $n$, so we have preferred to use the probability limits approach for the two charts to set control limits for a given significance level ($\alpha$). Using the respective control structures of the two charts, the control limits of the $M_r$ and $\bar{Y}$ charts have been obtained for different combinations of $\rho_{yx}$ and $n$, using different significance levels, and power curves for the two charts have been constructed. The power curves, for $n = 15$, are produced here for some values of $\rho_{yx}$ in Figure 2.1 (using $\alpha = 0.01$). We have equally divided the $\alpha$ on both the tails i.e. half of the $\alpha$ on the left tail and half on the right tail.

**Fig. 2.1:** Power curves of the $\bar{Y}$ chart and the $M_r$ chart with $|\rho_{yx}| = 0.10, 0.30, 0.50, 0.70$ using $\alpha = 0.01$

![Power curves of Y and M charts](image.png)

In Figure 2.1, the curve referred to as $\bar{Y}$ represents the power curve of the $\bar{Y}$ chart while the curves referred to as $M_{r_{0.10}}, M_{r_{0.30}}, M_{r_{0.50}}$ and $M_{r_{0.70}}$ represent the power curves of the $M_r$ chart when $|\rho_{yx}|$ is 0.10, 0.30, 0.50 and 0.70 respectively. Similar
behavior is observed for other values of $n$. We may conclude that the discriminatory power of the $M_r$ chart is higher than that of the $\bar{Y}$ chart and the gain in terms of discriminatory power for the $M_r$ chart keeps increasing with an increase in $|\rho_{xy}|$, as can be seen in Figure 2.1.

### 2.2.2.2 $M_r$ Chart vs. Cause-Selecting & Regression-Adjusted Charts

The cause-selecting and regression-adjusted control charts exploit the correlation between $Y$ and $X$ in the same manner (i.e. both types of charts are constructed for a quality characteristic of interest $Y$ after adjusting for the effect of some correlated characteristic $X$) so the comparisons of the proposed $M_r$ chart are made only with the cause-selecting charts. In this chapter, the performance of the $M_r$ chart is compared with those of cause-selecting approaches proposed by Zhang (1984) and Wade and Woodall (1993).

Wade and Woodall (1993) proposed prediction limits by modifying Zhang (1984) limits. Particularly, they considered $n = 50$, $\alpha = 0.006$, $\sigma_y^2 = \sigma_x^2 = 1$ and computed the powers for detecting different shifts in $\mu_y$ for given $x$, using Zhang’s limits and their proposed prediction limits. They claimed slightly better performance of their proposed prediction limits than Zhang’s limits, for detecting shifts in $\mu_y$ for given $x$. For the same situation (i.e. $n = 50$, $\alpha = 0.006$, $\sigma_y^2 = \sigma_x^2 = 1$), the powers are computed using the control limits of the proposed $M_r$ chart of this chapter. The same shifts in $\mu_y$ for given $x$ have been considered, as were considered by Wade and Woodall (1993) for Zhang’s limits and their proposed prediction limits. The powers are then plotted against different shifts in $\mu_y$ given $x$ for the Zhang (1984) limits, the
Wade and Woodall (1993) prediction limits and the limits based on the proposed $M_r$ chart of this chapter. The resulting power curves are shown in Figure 2.2 using $\rho_{yx} = 0.90$ for comparison purposes. A similar pattern of power curves has been observed for $\rho_{yx} = 0.50$ as well.

In Figure 2.2, the curves referred to as ‘1’, ‘2’ and ‘3’ are of Zhang’s limits, Wade and Woodall’s prediction limits and $M_r$ chart based limits respectively.

We may conclude that in terms of the discriminatory power, i) the prediction limits of Wade and Woodall (1993) are slightly better than Zhang’s (1984) (as claimed by Wade and Woodall (1993)) ii) the proposed $M_r$ chart based limits of this chapter are slightly better than those of Wade and Woodall (1993), as is obvious from Figure 2.2. This improved performance of the proposed $M_r$ chart based limits may be due to the exploitation of the correlation between $Y$ and $X$ (once in estimating $b$ of (2.1) and then in computing $M_r$ defined in (2.1)). These results may easily be claimed
for the $M_r$ chart relative to Hawkins (1993) regression-adjusted charts at least for the processes without cascade property (see Hawkins (1993)) due to the similar structure of these charts.

### 2.2.3 An Illustrative Example of the Proposed $M_r$ Chart

To illustrate the application of the proposed $M_r$ chart, let us consider the data considered first by Constable et al. (1988) and then by Wade and Woodall (1993) on $X = \text{ROLLWT}$ and $Y = \text{BAKEWT}$. They considered the paired information on $(Y, X)$ and used $X$ to find the fitted values $\hat{Y}$ and then the residuals $Y - \hat{Y}$ by fitting the least square regression line of $Y$ on $X$ when the process is assumed to be stable (in-control). They used the first 45 data points (which are from a stable state of the process) to find the fitted values $\hat{Y}$. After capitalizing on the relationship between $Y$ and $X$ this way, the residuals (i.e. $Y - \hat{Y}$) were computed for the improved process monitoring for $Y$. We consider here the same data set by giving the role of auxiliary variable to $X$ and the role of characteristic of interest to $Y$. The same first 45 stable data points on $(Y, X)$ are used and the means, standard deviations and correlation between $Y$ and $X$ are computed. These are: Mean $(Y) = 201.18$, Mean $(X) = 210.24$, SD $(Y) = 1.17$, SD $(X) = 1.23$ and Corr $(Y, X) = 0.54$. As these results are obtained from stable points of the process, we assume here these estimates as the true parameter values just for illustration purposes (i.e. $\mu_y = 201.18$, $\mu_x = 210.24$, $\sigma_y = 1.17$, $\sigma_x = 1.23$ and $\rho_{yx} = 0.54$). Assuming bivariate normality of $(Y, X)$ and considering $\mu_x = 210.24$ to be the known mean value of auxiliary variable $X$, ten bivariate random samples (ideally these should be 20-30 initial random samples as recommended by Shewhart but for convenience we have considered only ten here)
each of size ten are simulated from a bivariate normal distribution i.e. $N_2(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho_{xy})$. The inspiration for this approach of simulation is taken from Singh and Mangat (1996, p.221). The simulated data contains 90% values from $N_1(201.18, 210.24, 1.17, 1.23, 0.54)$ and 10% values from $N_1(203.18, 210.24, 1.17, 1.23, 0.54)$ (i.e. the data is contaminated for $\mu_y$). The resulting data set of ten bivariate random samples $(Y, X)$, each of size ten, is given in Table A2.1, given in the Appendix of this chapter. Now the objective is to see whether the $M_r$ chart is able to detect the contamination in the data or not? For this purpose, the sample statistics $M_r$ and $R_y$ (i.e. the sample ranges for $Y$) are computed for these ten samples and are given in the following table.

The Sample Statistics $M_r$ and $R_y$ for the Data Set of Table A2.1

<table>
<thead>
<tr>
<th>Sample Number</th>
<th>$M_r$</th>
<th>$R_y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>201.286</td>
<td>4.321</td>
</tr>
<tr>
<td>2</td>
<td>201.593</td>
<td>4.417</td>
</tr>
<tr>
<td>3</td>
<td>201.691</td>
<td>3.845</td>
</tr>
<tr>
<td>4</td>
<td>201.278</td>
<td>4.568</td>
</tr>
<tr>
<td>5</td>
<td>201.863</td>
<td>3.504</td>
</tr>
<tr>
<td>6</td>
<td>201.226</td>
<td>5.106</td>
</tr>
<tr>
<td>7</td>
<td>201.240</td>
<td>3.470</td>
</tr>
<tr>
<td>8</td>
<td>201.653</td>
<td>3.737</td>
</tr>
<tr>
<td>9</td>
<td>200.364</td>
<td>2.603</td>
</tr>
<tr>
<td>10</td>
<td>201.408</td>
<td>3.813</td>
</tr>
</tbody>
</table>
Based on the results of the above table, the control limits for the $M_r$ chart using its control structure given in (2.16) are given as (using $\alpha = 0.02$):

$$
\begin{align*}
LCL &= \bar{M}_r + C_{0.01} \frac{\hat{\sigma}_y}{\sqrt{n}} \\
CL &= \bar{M}_r \\
UCL &= \bar{M}_r + C_{0.99} \frac{\hat{\sigma}_y}{\sqrt{n}}
\end{align*}
$$

(2.19)

where $\hat{\sigma}_y = \frac{\bar{R}_y}{d_2} = \frac{3.938}{3.078} = 1.2794$, $C_{0.01} = -2.1391$.

The $M_r$ chart using the control limits given in (2.19), for the simulated data set of Table A2.1, is given in Figure 2.3.

**Fig. 2.3: The $M_r$ Chart**

![Chart](image)

The sample statistic $M_r$, on its control chart, falling outside lower and upper control limits refers to some assignable cause(s) in the process at that time point. The $M_r$ chart gives a signal of some assignable cause(s) for the sample at time point 9 in the simulated data set.
2.2.4 Conclusions for the Proposed $M_r$ Chart

The proposal of this chapter is a Shewhart type control chart for Phase-I quality control. The proposed $M_r$ chart uses the information on a single auxiliary variable and capitalizes on its correlation with a quality characteristic of interest for the improved monitoring of the mean level of the quality characteristic of interest. It has been observed that the performance of the $M_r$ chart, in terms of the discriminatory power, keeps improving with an increase in $\rho_{yx}$. Also comparisons with the conventional, cause-selecting and regression-adjusted control charts have proven a better performance of the proposed $M_r$ chart, in terms of the discriminatory power, for detecting shifts in the process mean level for $Y$ (i.e. $\mu_y$).
### Appendix to Chapter 2

#### Table A2.1: Simulated Data Set

<table>
<thead>
<tr>
<th>Sample Number</th>
<th>Sample Values in Pairs (Y,X) of size Ten</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Y 203.32 200.65 204.06 202.01 200.36 200.58 199.73 200.68 200.81 200.92</td>
</tr>
<tr>
<td></td>
<td>X 212.04 210.38 209.50 210.66 210.12 212.05 209.91 209.14 210.30 212.90</td>
</tr>
<tr>
<td>2</td>
<td>Y 202.88 201.15 200.54 199.11 201.50 200.79 201.72 203.25 201.77 203.52</td>
</tr>
<tr>
<td></td>
<td>X 211.23 209.11 210.18 210.92 209.62 209.86 213.26 211.31 211.36 210.90</td>
</tr>
<tr>
<td>3</td>
<td>Y 201.34 201.41 203.76 200.57 200.76 203.44 201.31 202.47 201.77 199.91</td>
</tr>
<tr>
<td></td>
<td>X 209.86 207.93 209.26 208.50 210.21 211.61 210.99 210.08 210.94 209.50</td>
</tr>
<tr>
<td>4</td>
<td>Y 199.51 199.39 202.62 201.88 200.48 199.29 202.93 203.86 200.43 202.61</td>
</tr>
<tr>
<td></td>
<td>X 208.94 210.75 210.71 211.01 212.09 207.76 210.77 211.87 210.40 211.52</td>
</tr>
<tr>
<td>5</td>
<td>Y 200.86 200.42 201.27 201.70 203.92 202.37 202.45 202.16 201.04 202.65</td>
</tr>
<tr>
<td></td>
<td>X 211.20 208.89 210.53 209.16 211.59 211.06 211.72 209.34 211.77 212.02</td>
</tr>
<tr>
<td>6</td>
<td>Y 201.11 201.48 203.09 201.27 199.65 201.42 203.84 201.19 198.73 200.52</td>
</tr>
<tr>
<td></td>
<td>X 209.99 212.08 209.67 211.64 208.98 212.15 211.39 209.78 207.87 209.84</td>
</tr>
<tr>
<td>7</td>
<td>Y 200.49 201.38 202.63 201.02 202.04 199.39 199.33 202.80 201.29 201.88</td>
</tr>
<tr>
<td></td>
<td>X 209.35 211.11 209.27 210.23 209.74 210.75 208.71 210.73 209.88 210.55</td>
</tr>
<tr>
<td>8</td>
<td>Y 202.74 201.59 203.83 202.69 201.89 200.72 200.30 200.10 202.38 200.27</td>
</tr>
<tr>
<td></td>
<td>X 211.43 208.45 211.29 212.20 208.75 208.55 211.41 212.91 208.98 207.56</td>
</tr>
<tr>
<td>9</td>
<td>Y 200.95 201.65 201.76 200.14 199.36 199.55 199.75 201.20 199.16 199.91</td>
</tr>
<tr>
<td></td>
<td>X 212.17 210.69 209.30 209.60 206.72 209.72 209.94 210.41 208.13 208.53</td>
</tr>
<tr>
<td>10</td>
<td>Y 199.92 200.85 201.38 201.92 201.29 201.11 199.78 203.12 203.60 201.04</td>
</tr>
<tr>
<td></td>
<td>X 208.93 210.28 210.57 210.31 209.22 210.51 209.29 210.02 212.12 210.14</td>
</tr>
</tbody>
</table>