Multivariate Control Charts

September 8, 2020

Outline

Introduction

When the quality of a product or service is defined by more than one property, all the properties should be studied simultaneously to control and improve quality.

Examples of process outputs that have multiple quality characteristics are as follows:

- Several impurities in a chemical product can be identified as peaks on a chromatography report, and all should be reduced to improve quality.
- In printed circuit board manufacturing, there are many variables measured on each board that defines its quality.
- In healthcare, many simultaneously measured outcomes contribute to the quality of a service to patients.
- A manufactured part may have several physical dimensions that can be measured, and they must all be within a joint region to insure the part fitness for use.

Introduction

If there are p quality characteristics and separate Shewhart control charts that are maintained on each with $\pm 3\sigma$ control limits:

- The probability of a false signal from any one control chart is 0.0027 and the $ARL_0 = 370$.
- The probability of a false signal from at least one of the p control charts is increased to $1 - (1 - .0027)^p$, and the overall ARL_0 will be greatly decreased.
- This can result in in frequent false-positive signals if the p quality characteristics being monitored are independent.
- One way to compensate for this problem is to widen the control limits on each of the separate control charts, thus increasing each of their individual ARL_0's to the point that the overall ARL_0 for obtaining a false signal on at least one of the charts is still equal to 370. Again, that can be done if the quality characteristics are jointly independent or uncorrelated.

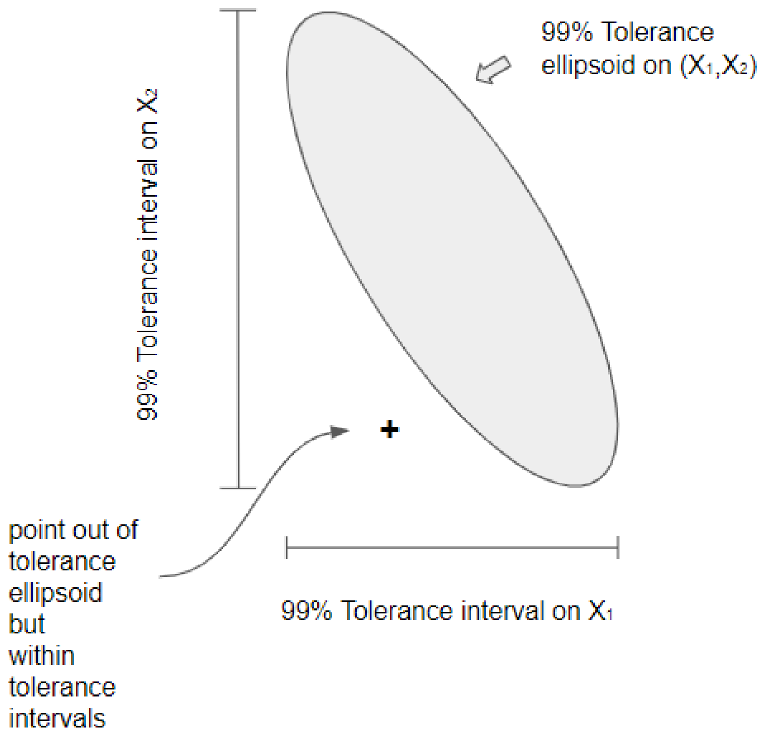
Introduction

If the several properties or quality characteristics measured on one product or service are interrelated or correlated, the opposite problem may occur.

- The chance of missing a real change in the process is increased.
- Although several charts are maintained, together they may be insensitive to detecting real changes in the process.
- Widening the limits of each individual control chart in this situation makes the charts even less sensitive to detecting real changes.

Elliptical Control Region

The figure below illustrates two interrelated quality characteristics X_1 and X_2. It can be seen that an observation on (X_1 and X_2) could be very unusual and outside the tolerance ellipse, yet well within the independent tolerance regions for X_1 and X_2.



Control Charts for Many Variables

When several quality characteristics are used in either a Phase I study or during Phase II monitoring with a control chart, the following considerations are important:

1. Reduce the chance of false-positive and false-negative signals when determining whether the process is in control.
2. Take into account any correlation among the quality characteristics studied.
3. If the process is out of control, be able to identify the nature of the problem.

Multivariate Control Charts

There are four distinct situations where multivariate control charts are used:

1. Phase I with data in rational subgroups
2. Phase I with individual observations
3. Phase II with data in rational subgroups
4. Phase II with individual observations.

Test of Multivariate Mean

When there are p correlated quality characteristics that follow a multivariate normal distribution, the statistic for testing the hypothesis

$H_0 : \mu = \mu_{\mathbf{0}}$, where $\mu_{\mathbf{0}} = \begin{pmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{pmatrix}$ is a hypothesized mean vector, and

$\mathbf{\Sigma}$ is the known covariance matrix, is: $T^2 = n(\bar{\mathbf{x}} - \mu_{\mathbf{0}})'\mathbf{\Sigma}^{\mathbf{-1}}(\bar{\mathbf{x}} - \mu_{\mathbf{0}})$,

and $\bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix}$ is a sample mean vector of n vectors of p quality characteristics. The distribution of this statistic is χ^2 with p degrees of freedom.

Multivariate T^2 Control Chart

Therefore, if rational subgroups of size n are collected, a standardized control chart for the mean vector is made by charting the quantities

$$T_i^2 = n(\bar{\mathbf{x}}_\mathbf{i} - \mu_\mathbf{0})'\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}_\mathbf{i} - \mu_\mathbf{0}) \tag{1}$$

where $\bar{\mathbf{x}}_\mathbf{i} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix}$ is the ith subgroup mean vector, $\mu_\mathbf{0} = \begin{pmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{pmatrix}$ is the known in-control process mean vector, and $\mathbf{\Sigma}$ is the known in-control process covariance matrix. The lower control limit for this chart is zero, and the upper control limit is $\chi^2_{\alpha/2,p}$.

Multivariate T^2 Control Chart

When the in-control mean $\mu_{\mathbf{0}} = \begin{pmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{pmatrix}$ and in-control covariance matrix $\mathbf{\Sigma}$ are unknown, they must be estimated by $\bar{\bar{\mathbf{x}}}$ and $\mathbf{S}$ from a series of in-control points on a Phase I control chart. If a Phase I study used m subgroups of size n, $\bar{\bar{\mathbf{x}}} = \begin{pmatrix} \bar{\bar{\mathbf{x}}}_{\mathbf{1}} \\ \bar{\bar{\mathbf{x}}}_{\mathbf{2}} \\ \vdots \\ \bar{\bar{\mathbf{x}}}_{\mathbf{p}} \end{pmatrix}$, and

$$\mathbf{S} = \begin{bmatrix} \bar{s}_j^2 & \dots & \bar{s}_{1p} \\ & \ddots & \\ \bar{s}_{p1} & \dots & \bar{s}_p^2 \end{bmatrix},$$

Multivariate T^2 Control Chart

where $\bar{\bar{\mathbf{x}}}_{\mathbf{j}}$ is the mean over the m subgroups of the subgroup sample means ($\bar{x}_j = \sum_{l=1}^{n} x_{jl}/n$)) for the jth quality characteristic, $\bar{s}_j^2$ is the mean over the m subgroups of the sample variances ($s_j^2 = \sum_{l=1}^{n} (x_{jl} - \bar{x}_l)^2/(n-1)$ for quality characteristic j within each subgroup, and $\bar{s}_{jk}$ is the mean over the m subgroups of the sample covariances $s_{jk} = \sum_{l=1}^{n} (x_{jl} - \bar{x}_j)(x_{kl} - \bar{x}_k)/(n-1)$ between quality characteristics j and k within each subgroup.

If the Phase I control chart used individual samples, then $\bar{\mathbf{x}}$ is the mean vector of the p quality characteristics over the m individual values, and $\mathbf{S}$ is the sample covariance matrix over the m individual pairs of the p quality characteristics.

Multivariate T^2 Control Chart

When the data on p correlated quality characteristics is gathered in m subgroups of size n, the ith T^2 statistic plotted on the control chart is:

$$T_i^2 = n(\bar{\mathbf{x}}_{\mathbf{i}} - \bar{\bar{\mathbf{x}}})' \mathbf{S}^{-\mathbf{1}} (\bar{\mathbf{x}}_{\mathbf{i}} - \bar{\bar{\mathbf{x}}}), \tag{2}$$

with upper control limit for a Phase I control chart is:

$$UCL = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}. \tag{3}$$

For Phase II monitoring the T^2 statistic is

$$T_f^2 = n(\bar{\mathbf{x}}_{\mathbf{f}} - \bar{\bar{\mathbf{x}}})' \mathbf{S}^{-\mathbf{1}} (\bar{\mathbf{x}}_{\mathbf{f}} - \bar{\bar{\mathbf{x}}}), \tag{4}$$

where $\bar{\mathbf{x}}_{\mathbf{f}}$ is a future subgroup mean to be observed, and the control limit changes to:

$$UCL = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}. \tag{5}$$

Multivariate Phase I T^2 Control Chart with individual data

When the data on p correlated quality characteristics is gathered in a Phase I study with m individual observations, the ith T^2 statistic plotted on the control chart is:

$$T_i^2 = (\mathbf{x_i} - \bar{\mathbf{x}})'\mathbf{S^{-1}}(\mathbf{x_i} - \bar{\mathbf{x}}), \tag{6}$$

and the upper control limit for a Phase I control chart is:

$$UCL = \frac{(m-1)^2}{m}\beta_{\alpha,p/2,(m-p-1)/2}, \tag{7}$$

where $\beta_{\alpha,p/2,(m-p-1)/2}$ is the α percentile of the beta distribution with parameters $p/2$ and $(m-p-1)/2$.The upper control limit for the Phase II control limit is:

$$UCL = \frac{p(m+1)(m-1)}{m^2 - mp}F_{\alpha,p,mn-p}, \tag{8}$$

Multivariate Phase I T^2 Control Chart with sub-grouped data

Table 9.2 in Ryan(2011) presents data for m=20 subgroups of n=4 observations on p=2 quality characteristics from a Phase I study. That book also illustrates the use of Minitab®, a commercial program, to construct a Phase I T^2 control chart. The data is included as the dataframe `RyanMultivar` in the R package `qcc`.

That dataframe is in the format for the `mqcc()` function in the `qcc` package that makes, T^2 control charts. However, it is not in the format that would normally be used to store multivariate data. The `IAcsSPCR` package, that contains the dataframes from this book, also contains the same data (`Ryan92`) in a more familiar format. This format includes one line for each observation and one column for each quality characteristic.

Multivariate Phase I T^2 Control Chart with sub-grouped data

The R code below illustrates retrieving the dataframe and reformatting it so that the `mqcc` function in the `qcc` package can be used to create a T^2 chart.

```
> library(IAcsSPCR)
> data(Ryan92)
> head(Ryan92,10)
   subgroup x1 x2
1         1 72 23
2         1 84 30
3         1 79 28
4         1 49 10
5         2 56 14
6         2 87 31
7         2 33  8
8         2 42  9
9         3 55 13
10        3 73 22
# reformat as a list of matricies required by the mqcc function
> X1<-matrix(Ryan92$x1,nrow=20,byrow=TRUE)
> X2<-matrix(Ryan92$x2,nrow=20,byrow=TRUE)
> XR = list(X1 = X1, X2 = X2) # a list of matrices, one for each variable
```

Multivariate Phase I T^2 Control Chart with sub-grouped data

The `mqcc` function requires the data be formatted as a list of m by n matrices one for each quality characteristic. The next two statements in the code on the last slide extract the quality characteristic columns from the data frame, and then reformat each of them into a $m = 20 \times 4 = n$ matrices. Finally these matrices are combined in a list named `XR`. This list is shown side by side below.

```
XR
```

```
## $X1                                ## $X2
##       [,1] [,2] [,3] [,4]          ##       [,1] [,2] [,3] [,4]
##  [1,]   72   84   79   49          ##  [1,]   23   30   28   10
##  [2,]   56   87   33   42          ##  [2,]   14   31    8    9
##  [3,]   55   73   22   60          ##  [3,]   13   22    6   16
##  [4,]   44   80   54   74          ##  [4,]    9   28   15   25
##  [5,]   97   26   48   58          ##  [5,]   36   10   14   15
##  [6,]   83   89   91   62          ##  [6,]   30   35   36   18
##  [7,]   47   66   53   58          ##  [7,]   12   18   14   16
##  [8,]   88   50   84   69          ##  [8,]   31   11   30   19
##  [9,]   57   47   41   46          ##  [9,]   14   10    8   10
## [10,]   26   39   52   48          ## [10,]    7   11   35   30
## [11,]   46   27   63   34          ## [11,]   10    8   19    9
## [12,]   49   62   78   87          ## [12,]   11   20   27   31
## [13,]   71   63   82   55          ## [13,]   22   16   31   15
## [14,]   71   58   69   70          ## [14,]   21   19   17   20
## [15,]   67   69   70   94          ## [15,]   18   19   18   35
## [16,]   55   63   72   49          ## [16,]   15   16   20   12
## [17,]   49   51   55   76          ## [17,]   13   14   16   26
## [18,]   72   80   61   59          ## [18,]   22   28   18   17
## [19,]   61   74   62   57          ## [19,]   19   20   16   14
## [20,]   35   38   41   46          ## [20,]   10   11   13   16
```

Multivariate Phase I T^2 Control Chart with sub-grouped data

The next block of code illustrates the use of the `mqcc()` function to create the T^2 chart shown in Figure 7.2. The summary of the object `q` created by the `mqcc()` function is shown below the code.

```
> library(qcc)
> q = mqcc(XR, type = "T2",add.stats=TRUE,title="T2 chart for Data in Ryan's Table 9.2")
> summary(q)

-- Multivariate Quality Control Chart ------------

Chart type                 = T2
Data (phase I)             = X
Number of groups           = 20
Group sample size          = 4
Center =
     X1      X2
60.3750 18.4875
Covariance matrix =
         X1        X2
X1 222.0333 103.11667
X2 103.1167  56.57917
|S| = 1929.414

Control limits:
 LCL      UCL
   0 11.03976
```

Multivariate Phase I T^2 Control Chart with sub-grouped data

The `mqcc()` function uses $\alpha = 1 - (1 - .0027)^p$ in order to reduce the chance of false positive results. Knowing this value of α the UCL can be verified using Equation 7.3

UCL= $\frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}$ as shown in the code below.

```
> p<-2
> n<-4
> num<-m*n*p-m*p-n*p+p
> dfd<-m*n-m-p+1
> alpha<-1-(1-.0027)^p
> UCL<-(num/dfd)*qf(1-alpha,p,dfd)
> UCL

[1] 11.03976
```

Multivariate Phase I T^2 Control Chart with sub-grouped data

The T^2 chart in Figure 7.2 shows subgroups 10 and 20 to be out of control. Since the sample covariance matrix $\mathbf{S}$ is not diagonal, the two quality characteristics `x1` and `x2` are not independent and we would not expect $\bar{X}$-charts constructed for each quality characteristic to be as sensitive to detecting changes as the T^2 chart.

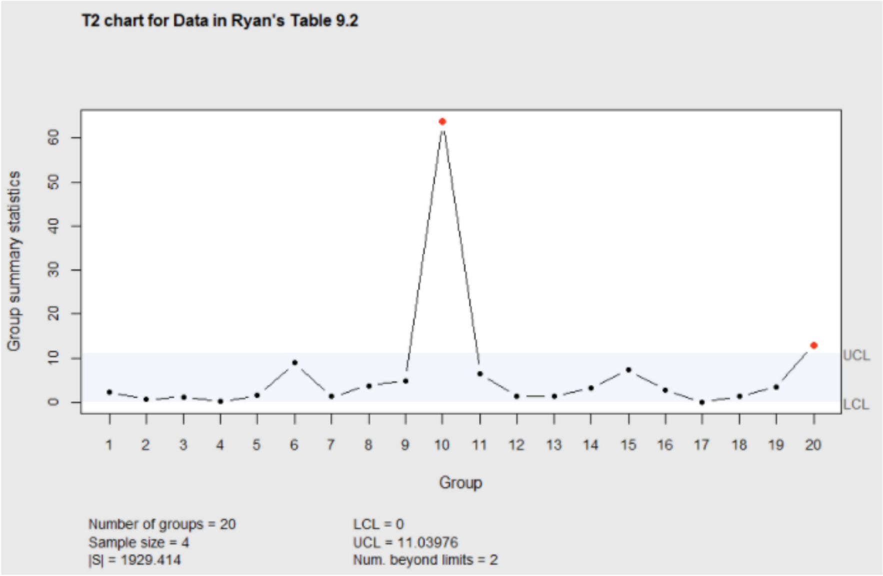


Figure 7.2 Phase I T^2 chart

Multivariate Phase I T^2 Control Chart

In Figure 7.3, it can be seen that although the `x1` coordinate and the `x2` coordinate for subgroups 10 and 20 are not unusual on their own, the two pairs of coordinates lie outside the control ellipse.

```
> qcc(Ryan92$x1,type="xbar",sizes=4)
> qcc(Ryan92$x2,type="xbar",sizes=4)
> ellipseChart(q, show.id = TRUE)
```

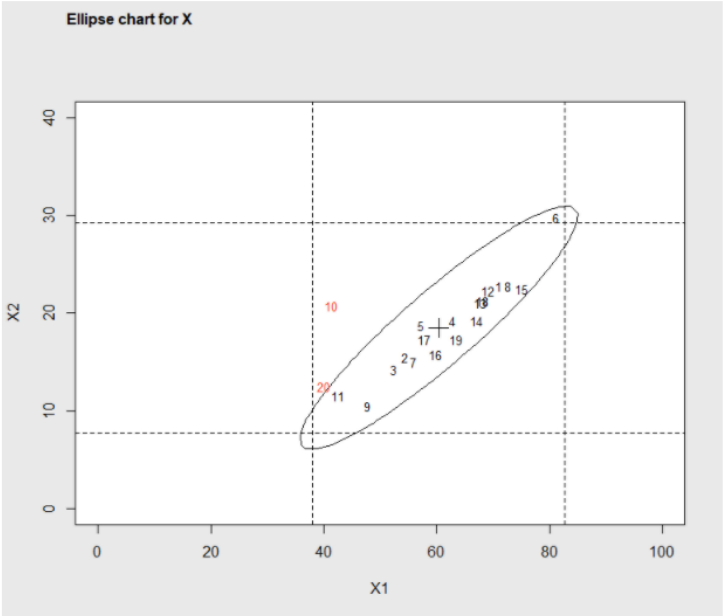


Figure 7.3 Ellipse Plot identifying out-of-control subgroups

Multivariate Phase I T^2 Control Chart with sub-grouped data

The control chart should be reconstructed without the out-of-control points to get better estimates of the in-control mean vector and covariance matrix. This is illustrated in the R code below.

```
> library(IAcsSPCR)
> data(Ryan92)
> # the following statement eliminates subgroup 10
> Ryan92s<-subset(Ryan92, subgroup != 10)
> # the following statement eliminates subgroup 20
> Ryan92s<-subset(Ryan92s, subgroup != 20)
> X1<-matrix(Ryan92s$x1,nrow=18,byrow=TRUE)
> X2<-matrix(Ryan92s$x2,nrow=18,byrow=TRUE)
> XR2 = list(X1 = X1, X2 = X2) # a list of matrices, one for each variable
> library(qcc)
> q2 = mqcc(X, type = "T2",add.stats=TRUE,
       title="T2 chart for Data in Ryan's Table 9.2 eliminating Subgroups 10 and 20")
> summary(q2)
```

Multivariate Phase I T^2 Control Chart with sub-grouped data

The resulting T^2 chart is shown in Figure 7.4.

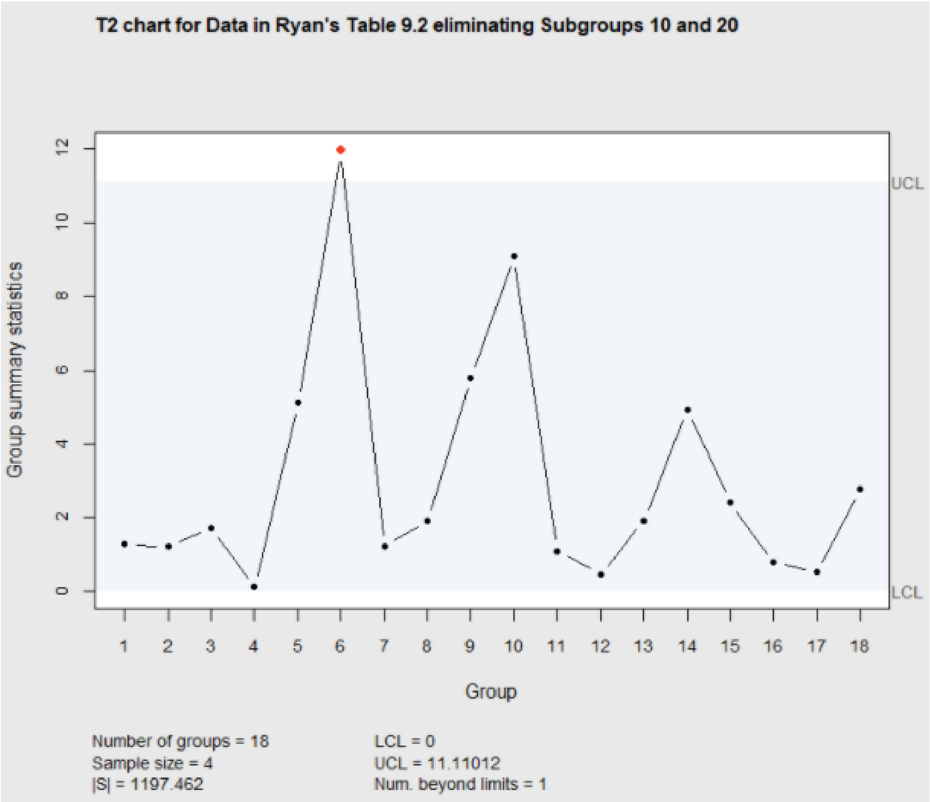
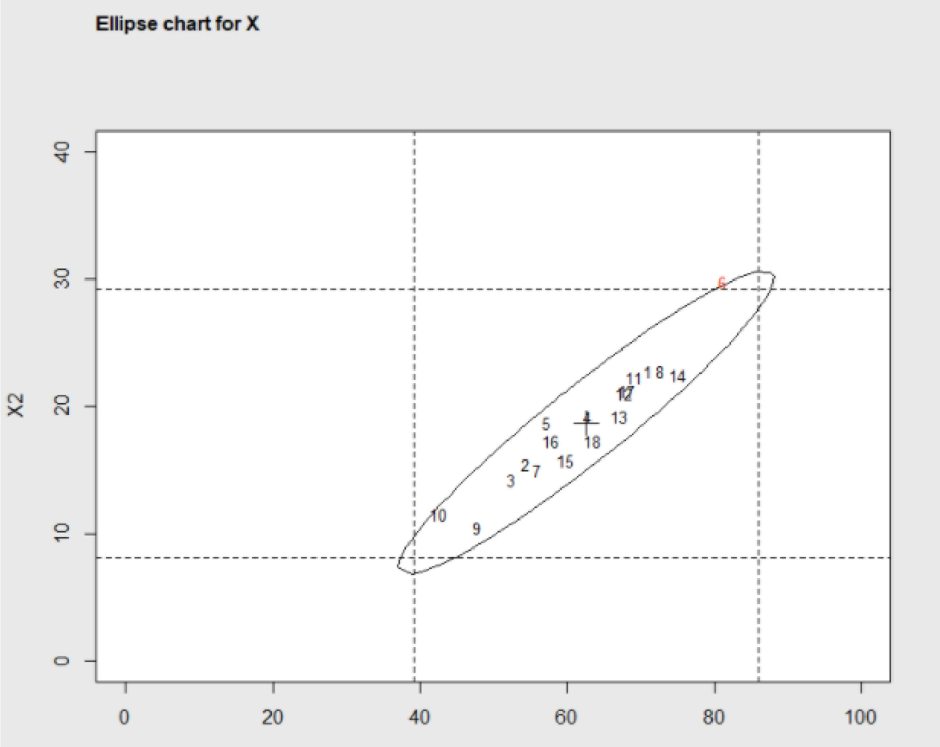


Figure 7.4 Phase I T^2 chart eliminating subgroups 10 and 20

Multivariate Phase I T^2 Control Chart with sub-grouped data

Now it can be seen that subgroup 6 is beyond the UCL. The process of refining the control limits and adding items to the OCAP is often an iterative process as described in Chapter 4.



Multivariate Phase I T^2 Control Chart

The code below was used to re-create the T^2 chart eliminating subgroup 6. The output of the summary, below the code, shows the revised estimates of the mean vector, covariance matrix, and the revised UCL.

```
> # the next statment eliminates subgroup 6
> Ryan92s<-subset(Ryan92s, subgroup != 6)
> X1<-matrix(Ryan92s$x1,nrow=17,byrow=TRUE)
> X2<-matrix(Ryan92s$x2,nrow=17,byrow=TRUE)
> XR3 = list(X1 = X1, X2 = X2) # a list of matrices, one for each variable
> library(qcc)
> q3 = mqcc(X, type = "T2",add.stats=TRUE,
>    title="T2 chart for Data in Ryan's Table 9.2 Eliminating Subgroups, 6, 10, and 20")
summary(q3)
-- Multivariate Quality Control Chart ------------

Chart type                 = T2
Data (phase I)             = X
Number of groups           = 17
Group sample size          = 4
Center =
      X1       X2
61.47059 18.04412
Covariance matrix =
         X1        X2
X1 241.7353 105.38235
X2 105.3824  50.90686
|S| = 1200.545

Control limits:
 LCL      UCL
   0 11.15193
```

Multivariate Phase I T^2 Control Chart with sub-grouped data

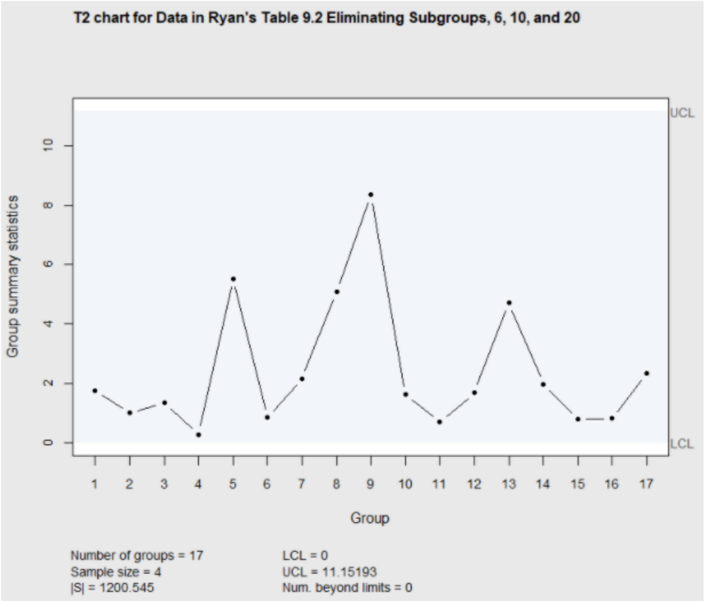


Figure 7.6 Phase I T^2 chart eliminating subgroups 6, 10 and 20

Assuming out-of-control points like subgroups 6, 10, and 20 can be avoided in the future, the process should remain in control with the mean vector and covariance matrix approximately equal to the results shown in the summary output on the last slide.

Monitoring Variability with Multivariate Control Charts

- When multivariate data is collected in subgroups, it is possible to monitor the process variability as well as the process mean vector.
- High values for some quality characteristics may be an indicator of a deterioration in quality
- Monitoring the mean vector of quality characteristics is not the only way to detect this possible problem
- Increases in the variability correspond to short term spikes that could again indicate sporadic reductions in quality
- For that reason the process variability, indicated by $\mathbf{S}$, should be monitored in addition to the mean vector

Monitoring Variability with Multivariate Control Charts

Alt(1985) proposed two different control charts for monitoring the covariance matrix $\mathbf{S}$

- The simplest of the two is based on the *generalized variance*, $|\mathbf{S}|$ the determinant of the covariance matrix
- This is a univariate statistic and Alt constructed a control chart based on the interval $E(|\mathbf{S}|) \pm 3\sqrt{V(|\mathbf{S}|)}$

The expected value and variance of $|\mathbf{S}|$ are:

$$E(|\mathbf{S}|) = b_1|\boldsymbol{\Sigma}|,$$

$$V(|\mathbf{S}|) = b_2|\boldsymbol{\Sigma}|.$$

$$b_1 = \tfrac{1}{(n-1)^p}\prod_{i=1}^{p}(n-i),$$

and $$b_2 = \tfrac{1}{(n-1)^{2p}}\prod_{i=1}^{p}(n-i)\left[\prod_{j=1}^{p}(n-j+2) - \prod_{j=1}^{p}(n-j)\right].$$

Monitoring Variability with Multivariate Control Charts

The R package `IAcsSPCR` contains a function `GVcontrol()` for constructing a control chart of the generalized variances. The example code below shows how it can be used to create a control chart of the generalized variances for each subgroup in the data from Ryan's table 9.2 that was described in the last section. Part of the output is shown below the function call and the rest is on the next slide.

```
> data(Ryan92)
> GVcontrol(Ryan92,20,4,2)
$name
[1] "UCL="

$value
[1] 10771.1

$name
[1] "Covariance matrix="

$value
         x1        x2
x1 222.0333 103.11667
x2 103.1167  56.57917
```

Monitoring Variability with Multivariate Control Charts

```
$name
[1] "Generalized Variance |S|"

$value
[1] 1929.414

$name
[1] "mean vector="

$value
     x1      x2
60.3750 18.4875

$name
[1] "Subgroup Generalized Variances="

$value
 [1]   45.0555556 2035.6666667 1195.0555556   30.8888889 9445.5000000   57.0555556
 [7]    4.0000000  452.8333333    1.1111111 3150.1666667  798.7777778  286.6111111
[13]  453.5000000  101.5000000  120.5555556   47.0555556    0.3888889   72.5000000
[19]  156.2777778    1.8888889
```

Monitoring Variability with Multivariate Control Charts

The control chart produced by the `GVControl()` function is shown in Figure 7.7

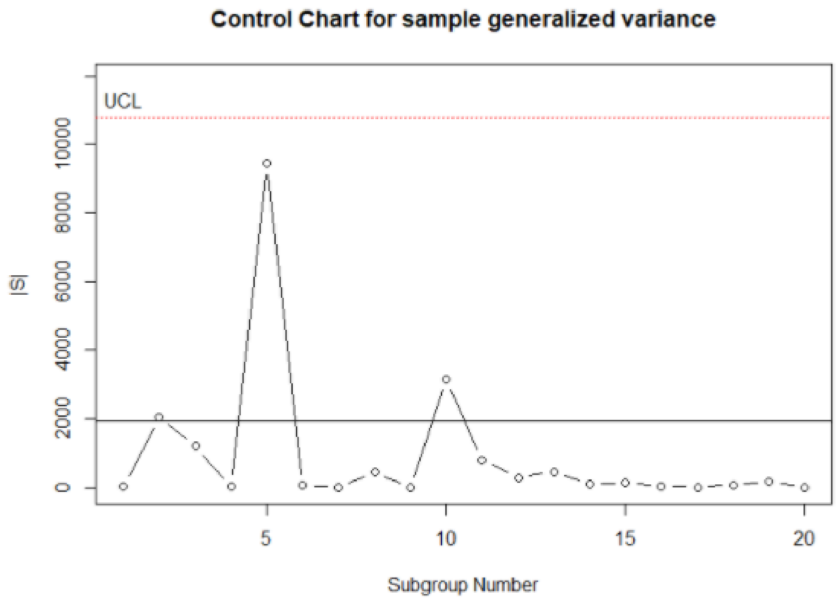


Figure 7.7 Control chart of Generalized Variances |S| for Ryan's Table 9.2

In this control chart, it can be seen that the generalized variances from each subgroup fall within the control limits, and process variability appears to be in control.

Monitoring Variability with Multivariate Control Charts

- In order to have accurate estimates of the covariance matrices within each subgroup, the subgroup size n should be substantially larger than the number of quality characteristics p
- In Phase I studies with sub-grouped data, the variance estimates within each subgroup are pooled, or averaged, to get an estimate of the process σ or Σ (for multivariate charts) that becomes the standard for Phase II monitoring
- If there are assignable causes in some subgroups that indicate an increase in the covariance matrix, Σ, those subgroups should be removed before finalizing the estimate of Σ and checking for out of control signals on the T^2 chart

Monitoring Variability with Multivariate Control Charts

Consider the data in the dataframe **`Frame`** that is included in the **`IAcsSPCR`** package. This dataframe contains $m = 10$ subgroups of data with $p = 3$ quality characteristics. The subgroup size is $n = 10$. This data is in the familiar format with one line for each observation, a column indicating the subgroup number, and additional columns for each of the 3 quality characteristics.

The code below retrieves the dataframe and calls the **`GVcontrol()`** function to make a control chart of the generalized variances. The control chart is shown in Figure 7.8.

```
> library(IAcsSPCR)
> data(Frame)
> GVcontrol(Frame,10,10,3)
$name
[1] "UCL="

$value
[1] 67.46235
```

Monitoring Variability with Multivariate Control Charts

```
$name
[1] "Covariance matrix="

$value
         V2       V3       V4
V2 3.477476 2.623105 1.147732
V3 2.623105 4.625853 1.234318
V4 1.147732 1.234318 2.287319

$name
[1] "Generalized Variance |S|"

$value
[1] 17.09664

$name
[1] "mean vector="

$value
     V2      V3      V4
10.4409 17.3352  9.2669
```

Monitoring Variability with Multivariate Control Charts

```
$name
[1] "Subgroup Generalized Variances="


$value
 [1]   6.710274   3.124969   2.327437   6.346057  16.971770   6.491100   1.070446
 [8]  14.251040   4.186463 112.781321
```

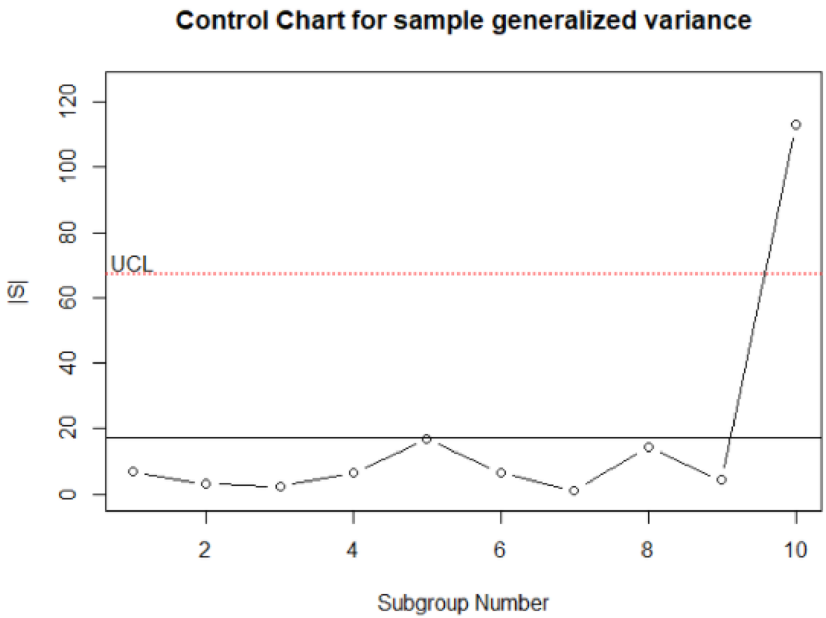


Figure 7.8 Control chart of Generalized Variances |S| for dataframe Frame.

Monitoring the Mean Vector with Multivariate Control Charts

The code below illustrates reformatting the data in **`Frame`** so that the `mqcc()` function can be used to produce a T^2 chart of the data. The T^2 control chart of this data is shown in Figure 7.9

```
> # reformat as a list of matricies required by the mqcc function
> X1<-matrix(Frame$V2,nrow=10,byrow=TRUE)
> X2<-matrix(Frame$V3,nrow=10,byrow=TRUE)
> X3<-matrix(Frame$V4,nrow=10,byrow=TRUE)
> X = list(X1 = X1, X2 = X2, X3=X3) # a list of matrices, one for each variable
> > q<-mqcc(X,type="T2",add.stats=TRUE,title="T2 Generated data")
> summary(q)
```

Monitoring the Mean Vector with Multivariate Control Charts

```
-- Multivariate Quality Control Chart ------------

Chart type                = T2
Data (phase I)            = X
Number of groups          = 10
Group sample size         = 10
Center =
     X1      X2      X3
10.4409 17.3352  9.2669
Covariance matrix =
         X1       X2       X3
X1 3.477476 2.623105 1.147732
X2 2.623105 4.625853 1.234318
X3 1.147732 1.234318 2.287319
|S| = 17.09664

Control limits:
 LCL      UCL
   0 11.56047
```

Monitoring the Mean Vector with Multivariate Control Charts

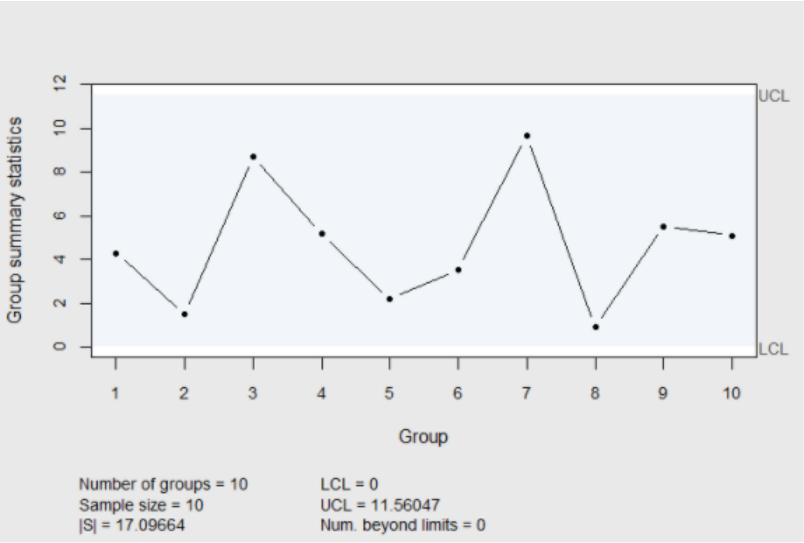


Figure 7.9 T^2 Control chart of subgroup mean vectors in dataframe Frame

The T^2 control chart in Figure 7.9 appears to indicate that there are no assignable causes of differences in the subgroup mean vectors. However, assuming that whatever caused the increased variances for subgroup 10 can be eliminated in the future, this subgroup should be removed and the two control charts reconstructed.

Monitoring Variability with Multivariate Control Charts

The code below removes subgroup 10 and calls the `GVcontrol()` function to make a control chart of the generalized variances for the first 9 subgroups in Figure 7.10.

```
> sFrame<-subset(Frame,subgroup!=10)
> GVcontrol(sFrame,9,10,3)
$name
[1] "UCL="

$value
[1] 35.73629

$name
[1] "Covariance matrix="

$value
         V2       V3       V4
V2 3.003067 2.794676 1.293888
V3 2.794676 4.337586 1.382692
V4 1.293888 1.382692 2.312179

$name
[1] "Generalized Variance |S|"

$value
[1] 9.056467
```

Monitoring Variability with Multivariate Control Charts

```
$name
[1] "mean vector="

$value
       V2        V3        V4
10.492556 17.338444  9.182111

$name
[1] "Subgroup Generalized Variances="

$value
[1]  6.710274  3.124969  2.327437  6.346057 16.971770  6.491100  1.070446 14.251040  4.186463
```

Figure 7.10 on the next slide shows the variability now to be in a state of control, and the printed results in the previous slide shows that the generalized variance $|\mathbf{S}|$ has decreased from 17.10 to 9.06.

Monitoring Variability with Multivariate Control Charts

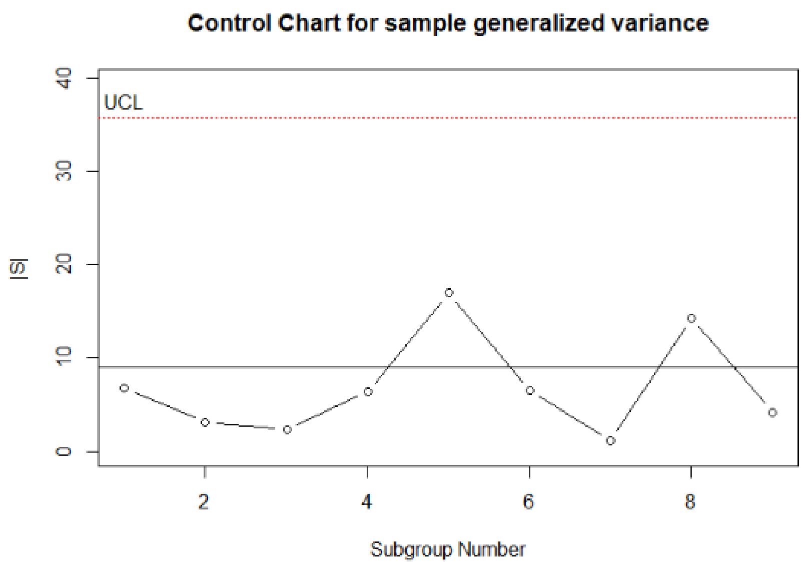


Figure 7.10 Control chart of Generalized Variances |S| for dataframe Frame eliminating subgroup 10.

Monitoring the Mean with Multivariate Control Charts

The code below re-formats the data in the reduced dataframe **sFrame** and calls the **mqcc()** function to produce a T^2 control chart. The resulting chart is shown in Figure 7.11 on the next slide.

```
> # reformat as a list of matricies required by the mqcc function
> X1<-matrix(sFrame$V2,nrow=9,byrow=TRUE)
> X2<-matrix(sFrame$V3,nrow=9,byrow=TRUE)
> X3<-matrix(sFrame$V4,nrow=9,byrow=TRUE)
> X = list(X1 = X1, X2 = X2, X3=X3) # a list of matrices, one for each variable
> q<-mqcc(X,type="T2",add.stats=TRUE,title="T2 Generated data")
> summary(q)
-- Multivariate Quality Control Chart -----------

Chart type                 = T2
Data (phase I)             = X
Number of groups           = 9
Group sample size          = 10
Center =
       X1        X2        X3
10.492556 17.338444  9.182111
Covariance matrix =
         X1       X2       X3
X1 3.003067 2.794676 1.293888
X2 2.794676 4.337586 1.382692
X3 1.293888 1.382692 2.312179
|S| = 9.056467

Control limits:
 LCL      UCL
   0 11.52832
```

Monitoring the Mean Vector with Multivariate Control Charts

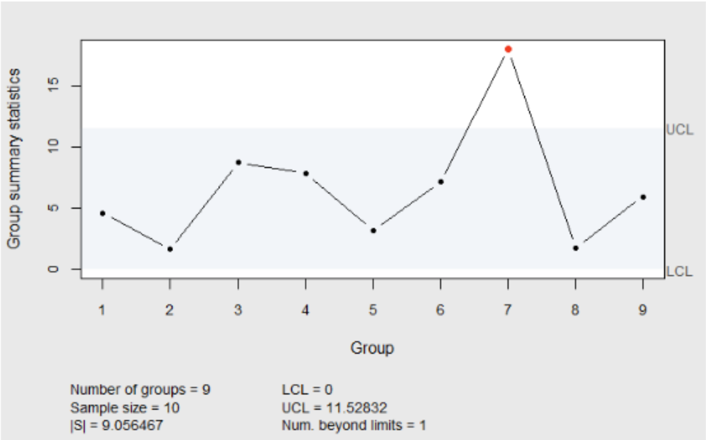


Figure 7.11 T^2 Control chart of subgroup mean vectors in dataframe Frame after eliminating subgroup 10

Figure 7.11 shows that the apparent in-control situation shown in Figure 7.9 was caused by the inflated generalized variance estimate $|\mathbf{S}| = 17.10$ that was made prior to eliminating subgroup 10. If this were a Phase I study, the next step would be to investigate the cause for the out-of-control subgroup 7.

Phase II T^2 Control Chart with subgrouped data

Continuing with the example using the dataframe `Frame`, suppose new process data was obtained in real-time monitoring. The dataframe `Xnew` in the R package `IAcsSPCR` contains 20 new subgroups of 4. The R code below creates a control chart using the Phase II control limits based on the mean and covariance matrix estimates from the Phase I study after eliminating subgroups 6, 10 and 20, then the `mqcc()` function adds the new data on the right side of the control chart.

```
> library(IAcsSPCR)
> data(Xnew)
> # the next three statements format the new data for the mqcc
  function
> X1<-matrix(Xnew$x1,nrow=20,byrow=TRUE)
> X2<-matrix(Xnew$x2,nrow=20,byrow=TRUE)
> Xn = list(X1 = X1, X2 = X2) # a list of matrices, one for
     each variable
> library(qcc)
> qn = mqcc(X, type = "T2", newdata=Xn,add.stats=TRUE,
     limits=FALSE,pred.limits=TRUE,center=q$center,
     cov=q$cov,title="T2 chart for Phase II data")
```

Phase II T^2 Control Chart with subgrouped data

In the `mqcc()` function call shown on the last slide, the first argument `X` tells the function to calculate the control limits based on the data, `X`, used in making Figure 7.6. That data contained m=17 subgroups. The argument `newdata=Xn` tells the function to add the reformatted new data on the right. The arguments `limits=FALSE` and `pred.limits=TRUE` tell the `mqcc()` function to use the formula for control limits in Equation 7.5 rather than Equation 7.3. The Phase I control limits in Equation 7.3 are called `limits` by the `mqcc()` function, while the Phase II limits in Equation 7.5 are called `pred.limits`.

Finally, the arguments `center=q$center` and `!cov=q$cov` tell the function to use the estimates of the in-control mean and in-control covariance matrix when calculating the plotted T_i^2 values given by: $T_i^2 = n(\bar{\mathbf{x}}_\mathbf{i} - \bar{\bar{\mathbf{x}}})'\mathbf{S}^{-1}(\bar{\mathbf{x}}_\mathbf{i} - \bar{\bar{\mathbf{x}}})$. The in-control mean and covariance matrix were calculated when making Figure 7.6. The resulting control chart with additional data is shown in Figure 7.12.

Phase II T^2 Control Chart with subgrouped data

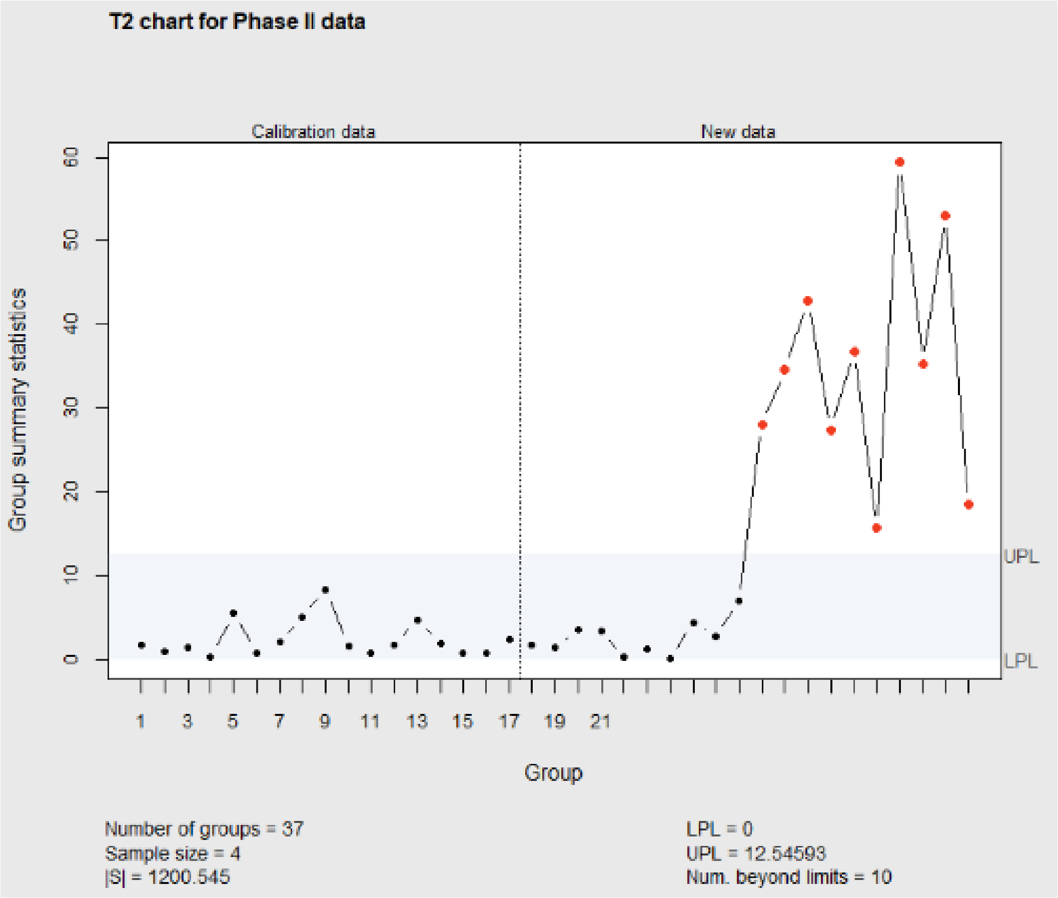


Figure 7.12 Phase II T^2 chart with 20 new subgroups of data

Phase II T^2 Control Chart with subgrouped data

Figure 7.12 shows shows that the process mean values for the 11th through 20th new subgroups have T^2 values that fall above the upper control limit. The UPL=12.54592 can be calculated using Equation 7.5 when $p = 2$, $m = 17$, and $n = 4$ by modifying the code above Figure 7.2.

There is no need to wait for 20 new subgroups of data to construct the Phase II control chart.

Phase I T^2 Chart with Individual data

Observations on 5 impurities from 30 individual lots Gonzales(2003-4)

```
> library(IAcsSPCR)
> qi<-mqcc(DrugI[,-1],type="T2.single",add.stats=TRUE,
        title="Historical data for Drug impurities")
> summary(qi)
-- Multivariate Quality Control Chart ------------


Chart type                 = T2.single
Data (phase I)             = DrugI[, -1]
Number of groups           = 30
Group sample size          = 1
Center =
        A         B         D         E         G
 23.33333  74.00000 625.66667 151.00000 684.00000
Covariance matrix =
           A          B          D           E          G
A  105.74713  -155.1724   294.2529   -51.72414  -551.7241
B -155.17241  4314.4828 -1399.3103  1813.10345  9424.8276
D  294.25287 -1399.3103 20459.8851  2645.86207 -2857.9310
E  -51.72414  1813.1034  2645.8621  9767.93103 10851.0345
G -551.72414  9424.8276 -2857.9310 10851.03448 67217.9310
|S| = 2.893802e+18

Control limits:
 LCL      UCL
   0 12.11333
```

Phase I T^2 Chart with Individual data

In the code on the last slide the argument `type="T2.single"` specifies that this is a T^2 chart for individuals. The output below the code shows the mean vector (over the 30 observations) for impurities A, B, D, E, G, and the covariance matrix calculated with the 30 observations.

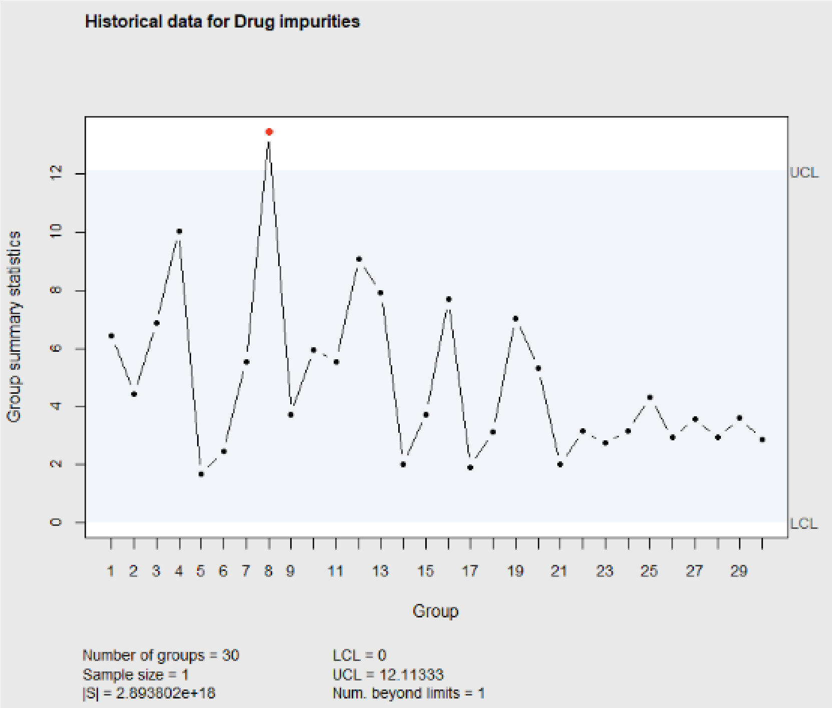


Figure 7.13 Phase I T^2 Chart of Drug Impurities

Phase I T^2 Chart with Individual data

The control chart should be re-created eliminating observation 8.

```
> # Redo eliminating observation 8
> DrugIe<-DrugI[-8,]
> qi<-mqcc(DrugIe[,-1],type="T2.single",add.stats=TRUE,
        title="Historical data for Drug impurities")
> summary(qi)
-- Multivariate Quality Control Chart ------------

Chart type                 = T2.single
Data (phase I)             = DrugIe[, -1]
Number of groups           = 29
Group sample size          = 1
Center =
        A         B         D         E         G
 23.44828  75.17241 630.00000 155.51724 665.86207
Covariance matrix =
           A          B          D           E          G
A  109.11330  -164.9015   289.2857   -69.70443  -506.6502
B -164.90148  4425.8621 -1607.1429  1713.30049 10422.1675
D  289.28571 -1607.1429 20607.1429  2132.14286  -517.8571
E  -69.70443  1713.3005  2132.1429  9482.75862 13784.3596
G -506.65025 10422.1675  -517.8571 13784.35961 59396.5517
|S| = 1.79408e+18
```

Phase I T^2 Chart with Individual data

Now it can be seen the 4th observation falls above the UCL

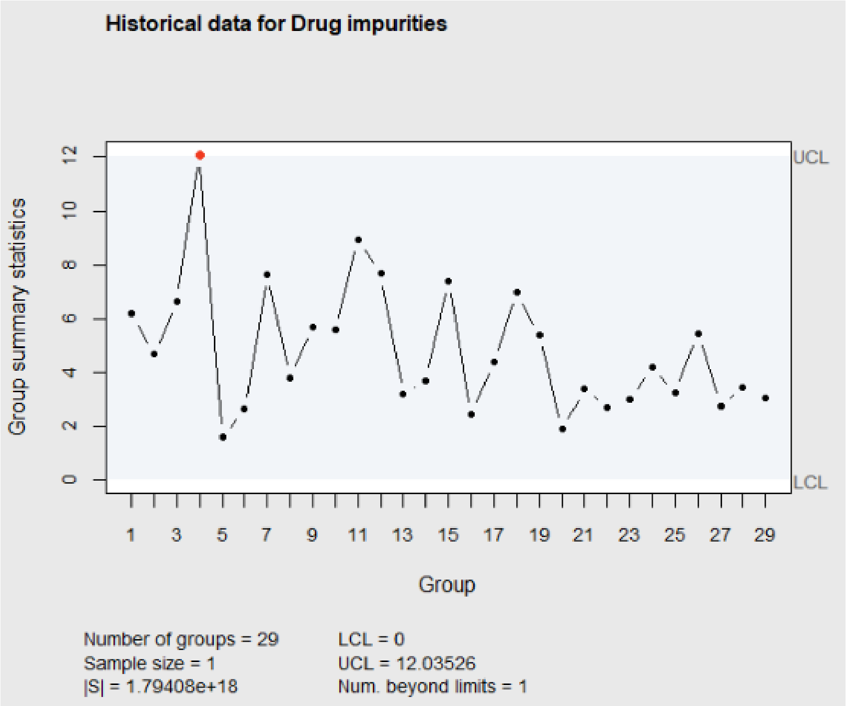


Figure 7.14 Phase I T^2 Chart of Drug Impurities after eliminating observation 8

Phase I T^2 Chart with Individual data

After investigating and eliminating the 4th observations, the process appears to be in control and the mean vector and covariance matrix can be used in Phase II monitoring.

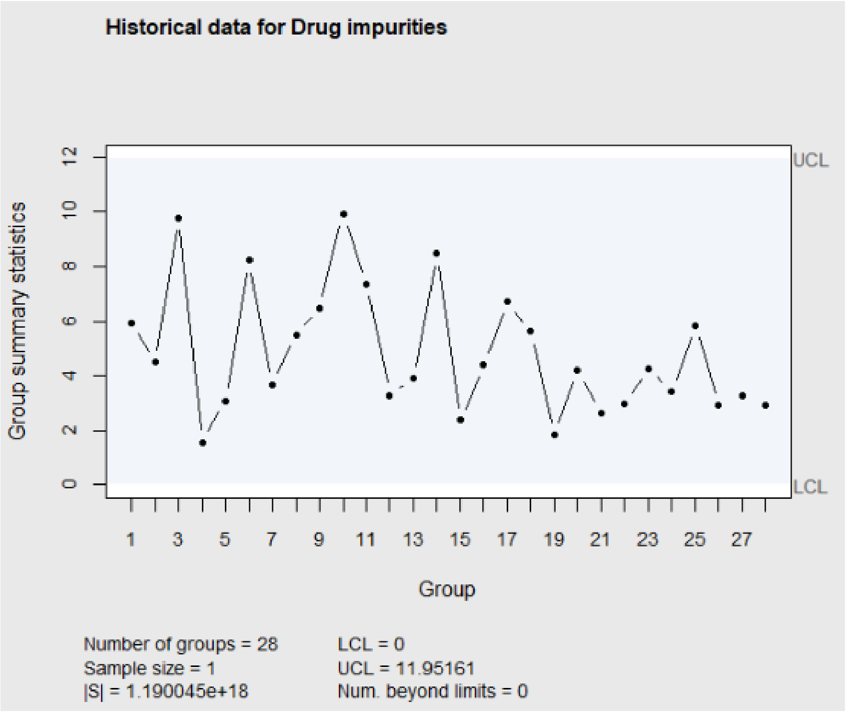


Figure 7.15 Phase I T^2 Chart of Drug Impurities after eliminating observation 8, and 4

Phase II T^2 Chart with Individual data

Continuing with Gonzales(2003-4)'s example, there were 160 additional observations from Phase II monitoring of drug lots. The dataframe `DrugIn` in the `IAcsSPCR` package contains the first 10 of these observations. The code below creates a Phase II control chart of the additional 10 observations on the right.

```
> library(IAcsSPCR)
> library(qcc)
> data(DrugIn)
> qn = mqcc(DrugIe2[,-1], type = "T2.single", newdata=DrugIn[,-1],add.stats=TRUE,
        limits=FALSE,pred.limits=TRUE,center=qi$center,cov=qi$cov,
        title="T2 Individual chart for Phase II data")
```

Phase II T^2 Chart with Individual data

In Figure 7.16 it can be seen that an assignable cause is signaled at the 5th observation in the additional data.

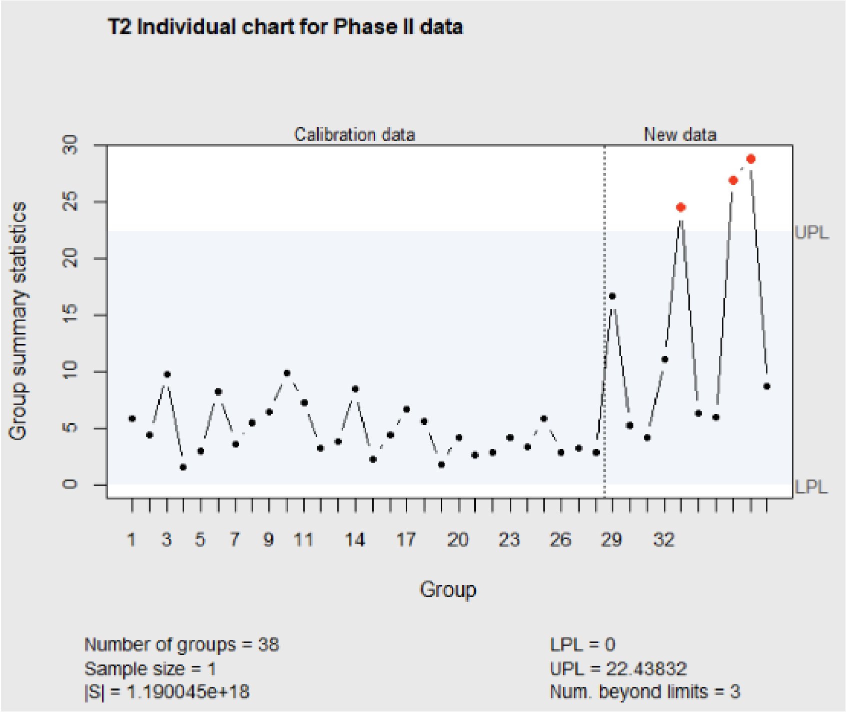


Figure 7.16 Phase II T^2 Chart of Drug Impurities

Interpreting Out-of-control Signals

Which of the p quality characteristics (or subset of them) is mainly responsible for an out of control signal on a T^2 chart.

- When there are only two quality characteristics, the ellipse plot made by the `ellipseChart()` function can help.
- When there are more than two quality characteristics, making a barchart of the the percentage change of each of the quality control characteristic from its mean value is useful.
- To better determine the contribution of each quality characteristic, Runger, Alt and Montgomery(1996) discussed the decomposition of the T^2 statistic for an out-of-control point.

Barchart of Impurity Percentage Changes from the Mean for Observation 8

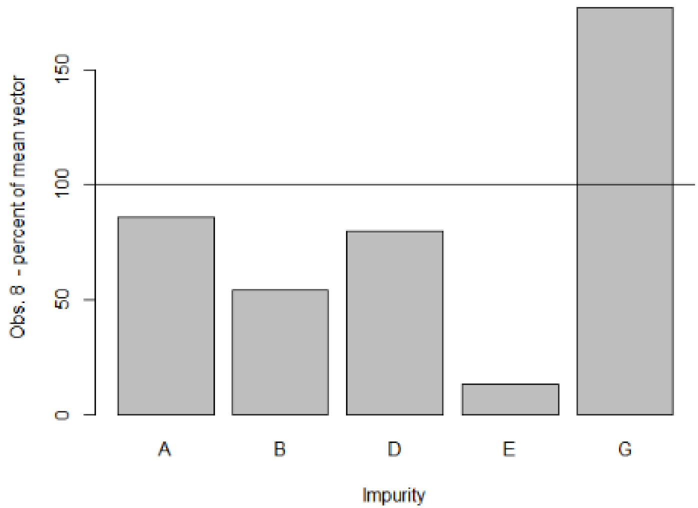


Figure 7.17 Barchart of Imurity percentage changes from the mean for observation 8

Decomposition of the T^2 statistic

- T^2 is the T^2 value of an out-of-control point
- $T^2_{(i)}$ is the T^2 for the same observation omitting quality characteristic i and only using the other $p - 1$ quality characteristics
- Runger, Alt and Montgomery recommend computing the values d_i (i= 1, 2, …, p) and then focusing on the variables for which the d_i values are relatively large.
- The T^2 values for each of the observations can then be retrieved as the statistics component in the object created by `mqcc()` i.e., `object$statistics`.

Multivariate EWMA Charts with Individual Data

- When $\boldsymbol{\mu_0}$, and $\boldsymbol{\Sigma}$ are known, the T^2 chart

$$T_i^2 = n(\mathbf{x_i} - \boldsymbol{\mu_0})'\boldsymbol{\Sigma}^{-1}(\mathbf{x_i} - \boldsymbol{\mu_0}) \sim \chi_p^2$$

  is based on the most recent observation, $\mathbf{x_i}$.

- Therefore, it is insensitive to detecting small shifts in the mean vector $\boldsymbol{\mu_0}$.

Multivariate EWMA Charts with Individual Data

Lowry et. al.(1996) developed a Multivariate EWMA T^2 control chart (called a MEWMA) that has greater sensitivity for detecting small changes in $\mu_{\mathbf{0}}$.

$$\mathbf{z_i} = \mathbf{R}(\mathbf{x_i} - \mu_{\mathbf{0}}) + (\mathbf{I} - \mathbf{R})\mathbf{z_{i-1}}. \tag{9}$$

- where $\mathbf{R}$ =diag$(r_1, r_2, \ldots, r_p)$ and $0 < r_i \leq 1$, $i = 1, \ldots, p$.
- If there is no apriori reason to weigh past observations differently for the p quality characteristics, then $r_1 = r_2 = \ldots = r_p = r$, and $\mathbf{z_i}$ simplifies to:
- $\mathbf{z_i} = r(\mathbf{x_i} - \mu_{\mathbf{0}}) + (1 - r)\mathbf{z_{i-1}}$.
- The covariance matrix of $\mathbf{z_i}$ is $\mathbf{\Sigma_{z_i}} = (r/(2-r))\mathbf{\Sigma}$
- An out-of-control signal for the MEWMA occurs when $T_i^2 = \mathbf{z_i}'\mathbf{\Sigma_{z_i}}^{-1}\mathbf{z_i} > h_4$
-

Multivariate EWMA Charts with Individual Data

Lowry et. al.(1996) used simulation to generate the upper control limit h_4 that would result in and $ARL_0 \approx 200$.

Table 1: Upper control limits h_4 for MEWMA charts with $r = .1$

p	h_4
2	8.66
3	10.79
4	12.73
5	14.56
10	22.67
20	37.01

Multivariate EWMA Charts with Individual Data

Consider the the following situation. A random sample of 10 observations from a multivariate normal distribution is generated with an in-control mean of $\mu_{\mathbf{0}} = \begin{pmatrix} 25 \\ 10 \\ 17 \\ 15 \end{pmatrix}$ and a known covariance matrix.

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} 5.40000000 & 0.09583333 & 2.0583333 & 3.1291667 \\ 0.09583333 & 0.48400000 & 0.2963333 & 0.2686667 \\ 2.05833330 & 0.29633330 & 2.2993333 & 1.0056667 \\ 3.12916667 & 0.26866667 & 1.0056667 & 2.2310000 \end{bmatrix}.$$

Multivariate EWMA Charts with Individual Data

Next, another random sample of 15 observations is generated from a multivariate normal distribution with the same covariance matrix, but the mean is shifted to $\mu = \mu_{\mathbf{0}} + \mu_{\mathbf{\Delta}}$, where $\mu_{\mathbf{\Delta}} = \begin{pmatrix} 1.10 \\ 0.60 \\ 1.00 \\ 0.70 \end{pmatrix}$. The non-centrality factor for the shift in the mean vector is

$$\lambda = (\mu_{\mathbf{\Delta}})' \mathbf{\Sigma}^{-1} (\mu_{\mathbf{\Delta}}) = 1.045907.$$

Multivariate EWMA Charts with Individual Data

The code below generates this data and then appends the first sample with the second sample below it to create the matrix D.

```
> # generate random sample of 10  multivariate obs. with p=4 and mean vector muv
> muv<-c(25,10,17,15)
> covm<-matrix(c(5.40000000, .09583333, 2.0583333, 3.12916667, .09583333, .48400000,  .2963333,
               .26866667, 2.05833333, .29633333, 2.2993333, 1.00566667, 3.12916667, .26866667,
               1.0056667, 2.23100000),nrow=4,ncol=4 )

> # generate a random sample of 10 obs. with p=4 and mean vector muv and covariance matrix covm
> library(MASS)
> set.seed(100)
> d1<-mvrnorm(10,mu=muv,Sigma=covm)
> # generate random sample of 15 multivariate obs. with p=4 and mean vector muv+mud and
> #  covariance matrix covm
> mud<-c(1.1,.6,1.0,.70)
> d2<-mvrnorm(15,mu=muv+mud,Sigma=covm)
> # calculate the noncentrality parameter for detecting the mean shift
> lambda<-mud%*%solve(covm)%*%mud
> lambda
         [,1]
[1,] 1.045907

> #combine the data in the combined data the mean shifts after the 10th obs.
> D<-rbind(d1,d2)
```

Multivariate EWMA Charts with Individual Data

The R package **IAcsSPCR** that contains the data sets and functions from this book contains the function `MEWMA()` that calculates Lowry et. al.'s MEWMA control chart. This function uses a small value of $r = 0.1$ in order to be sensitive to small shifts in the mean vector.

```
# MEWMA chart assuming mean and covariance matrix is known
> library(IAcsSPCR)
> MEWMA(D,covm,muv,Sigma.known=TRUE)
$name
[1] "UCL="

$value
[1] 12.73

$name
[1] "Covariance matrix="

$value
           [,1]       [,2]      [,3]      [,4]
[1,] 5.40000000 0.09583333 2.0583333 3.1291667
[2,] 0.09583333 0.48400000 0.2963333 0.2686667
[3,] 2.05833330 0.29633330 2.2993333 1.0056667
[4,] 3.12916667 0.26866667 1.0056667 2.2310000
```

Multivariate EWMA Charts with Individual Data

```
$name
[1] "mean vector"


$value
[1] 25 10 17 15
```

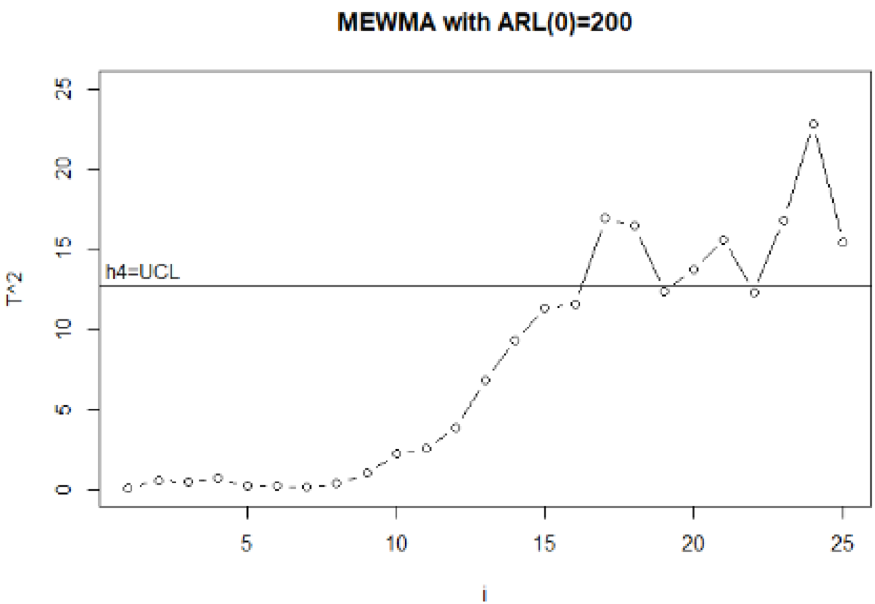


Figure 7.18 MEWMA chart of Lowry et.al data

Multivariate EWMA Charts with Individual Data

- The first argument can either be a matrix or dataframe with one row for each observation and one column for each of the p quality characteristics being monitored. This function works for $2 \leq p \leq 10$ quality characteristics.
- In this example, it is assumed that $\mathbf{\Sigma}_0$=`covm` is known and it is supplied to the function as the second argument.
- The in-control mean vector `muv` is supplied as the third argument
- The fourth argument `Sigma.known=TRUE` tells the function that these are the assumed known parameters. When `Sigma.known=FALSE` then the function calculates `mu` and `covm` from the data and they don't need to be supplied to the function.

Multivariate EWMA Charts with Individual Data

The first out of control point occurs at the 17th observation. Since the mean shifted after the 10th observation, it took the MEWMA chart 7 observations to detect this change.

The T^2 chart is used on the same data, and the results are shown in Figure 7.19

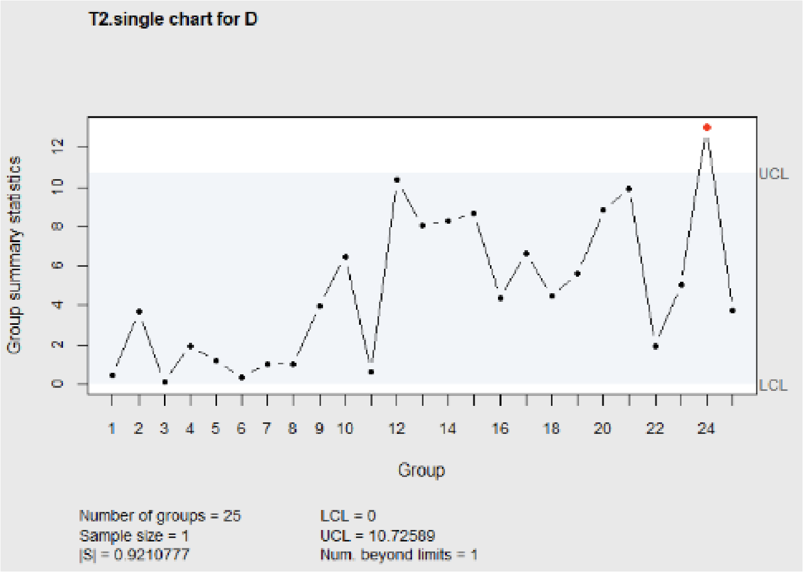


Figure 7.20 T^2 chart of Lowry et.al data

Multivariate EWMA Charts with Individual Data

TABLE 7.2: ARL Comparisons between T^2 and MEWMA Charts - note: Source Lowry et. al. *Technometrics* 34:46-53, 1992

λ	T^2	MEWMA	T^2	MEWMA	T^2	MEWMA
	p=2	p=2	p=3	p=3	p=4	p=4
0.0	200	200	201	200	200	201
0.5	116	28.1	130	31.8	138	34.7
1.0	42.0	10.2	52.6	11.3	60.9	12.1
1.5	15.8	6.12	20.5	6.69	24.6	7.23
2.0	6.9	4.41	8.8	4.86	10.6	5.18
2.5	3.5	3.51	4.4	3.83	5.19	4.10
3.0	2.2	2.92	2.6	3.2	2.83	3.41
λ	T^2	MEWMA	T^2	MEWMA	T^2	MEWMA
	p=5	p=5	p=10	p=10	p=20	p=20
0.0	200	200	200	200	200	200
0.5	145	31.8	162	48.1	174	64.1
1.0	68.1	11.3	92.8	15.9	117	20.0
1.5	28.5	6.69	44.7	9.16	66.2	11.3
2.0	12.4	4.86	20.6	6.55	34.3	8.00
2.5	5.99	3.83	9.90	5.15	17.4	6.26
3.0	3.31	3.20	5.20	4.28	9.1	5.19

In this table, it can be seen that the MEWMA has shorter ARL's for noncentrality factors $\lambda \leq 2.5$.