UNIVERSITY OF NAIROBI

COLLEGE OF BIOLOGICAL AND PHYSICAL SCIENCES

SCHOOL OF MATHEMATICS

PROJECT WORK

BIVARIATE PROCESS CONTROL FOR VARIABILITY

A PROJECT SUBMITTED TO THE SCHOOL OF MATHEMATICS IN PARTIAL FULFILMENT FOR THE AWARD OF MASTER OF SCIENCE DEGREE IN STATISTICS

AUGUST 2009

MUNG'AU M .K

Declaration

I the undersigned declare that this project is my original work and to the best of my knowledge has not been submitted for the award of a degree in any other University.

MUNG'AU M .K

Reg. No I/56/71095/2007

.... 13/08/2009 ....

Signature **Date**

Declaration By Supervisor

This project has been submitted for examination with my approval as supervisor

Dr. F. NJUI

School of Mathematics,

University of Nairobi,

P.O. Box 30197 Nairobi,

KENYA

.... 13th August 2009 ....

Signature **Date**

Dedicated to:

I dedicated this project to my dear wife Josephine, daughter Mumbi and my niece Mumbe for their sincere encouragement.

"Never settle for less than your dreams.
Somewhere , sometime, someday, somehow,
you'll find them"

Acknowledgement

The content and flow of this research work has been greatly influenced by a number of people and thus their contribution cannot be overlooked. I am very grateful to all those who have contributed in one way or the other towards the success of this work.

My deepest gratitude goes to DR.F.Njui for his tireless supervision through out this project, without his effort, it would have been very difficult for me to finish this work within the stipulated time frame. Allow me to mention his availability at all times and his guidance in various topics that has been so inspiring.

My sincere thanks and appreciation also go to members of my family for the support, cooperation and encouragement I received from them during the entire period of my studies.

Finally to my classmates, Alex, Macharia and Dorcas for the lovely atmosphere they provided during the entire course and for their constructive discussions across many issues that helped me patch up ideas into a well laid down research.

Abstract

It is important that a consumer get the quality of a product she/he wants. Consequently, the quality of a finished product should be controlled to a desired level. This is done through statistical quality control. We have Univariate process control and multivariate process control. In the first case, one is interested in controlling one quality characteristic of a product, whereas in the second case, one is interested in controlling more than one quality characteristics of a product. This project deals with both cases with special emphasis on the second.

Finally, we have looked at multivariate process control using the method of principal component analysis.

Table of Contents

Chapter 3

DATA COLLECTION AND ANALYSIS

Chapter 4

PRINCIPAL COMPONENTS AND FACTOR ANALYSIS

Chapter 1

NTRODUCTION

.1 Statistical quality control charts.

The best a manufacturer can do is to find out the causes of variability in the product and device means to control each contributing factor within appropriate limits, hereby controlling the quality of the product. This has to be done since the products should have the quality the consumer expects as the progress of an industry depends on the successful marketing (selling) of quality products. This is done by quality specifications from the arrival of raw materials, through their processing to the final product. However, complete elimination of the assignable causes of variation may not always be possible or even if possible, it may be uneconomical. Thus, statistical quality control is a system of planned collection and use of data for detecting causes of variation in the quality of a product. More specially it is a system of variation and maintenance of a desired level of quality by careful planning, using proper equipment, and controlling inspection and by taking corrective action where necessary. This is a most powerful productivity technique for effective diagnosis of lack of quality (lack of conformity to settled standards) in any of the materials, process, machines or end products. Statistical quality control therefore consists of techniques which help in the separation of the assignable causes from chance causes of variation thus signaling whenever assignable causes of variation are present in a process. To achieve this stated aim statistical quality control has three main objectives:

(i) It defines a goal or standards which the manufacturer may strive to achieve (specification),

(ii) It serves as an instrument for attaining the defined goal (production),

(iii) It serves as a means of judging whether the goal has been achieved

1.2 Setting limits on control charts

In order to exercise control over a process, we set limits on the variability due to chance occurrences. This is done by the use of control charts. If w is a sample statistic that measures some quality characteristic of interest with mean μ_w and standard deviation δ_w, then the control chart limits proposed by Shewhart (1931) are given by

$$UCL_w = \mu_w + k\delta_w$$

$$CL_w = \mu_w \qquad (1.1)$$

$$LCL_w = \mu_w - k\delta_w$$

where

UCL_w denotes the upper control limit

CL_w is the central line,

LCL_w is the lower control limit

k is the distance of the central line from the control limits.

When k=3, the control limits represented by (1.1) are called Shewhart 3-sigma limits. To apply these limits, a random sample of size n from a production process is taken at regular intervals. The relevant sample statistic is computed for each sample and plotted on a control chart with limits as in figure 1.1. The pattern of the plotted points on the control charts indicates whether the process is in control or otherwise in relation to the control chart limits.

Shewhart control chart

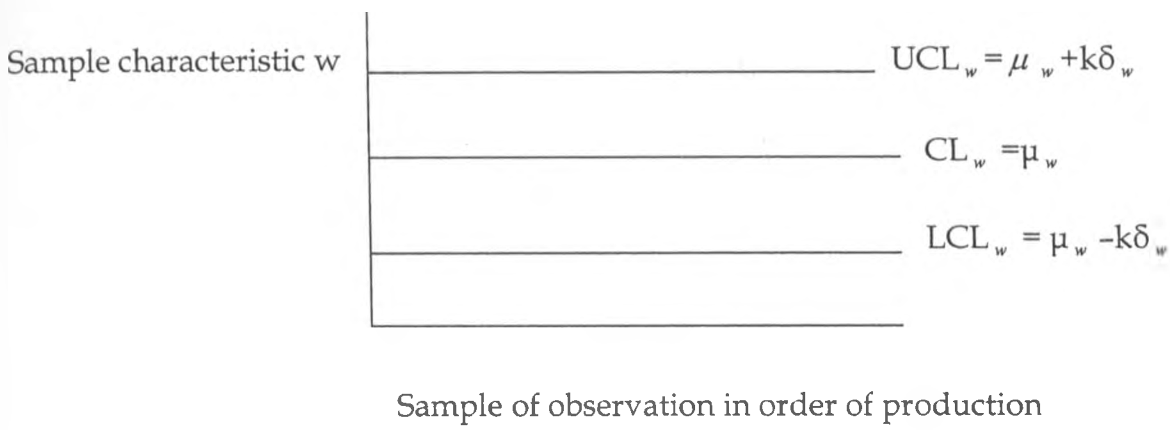


Figure 1.1

When a sample statistic w falls outside the control limits or when eight consecutive samples points plot above (or below) the central line, the process is said to be out of control,otherwise it is in control.

1.3 Types of control charts

Control charts can be classified into several types. Here, we briefly mention three types. The first type control charts are for variables which are used when the quality characteristics are measurable . These charts include the $\bar{X}$-chart, R- chart, S^2 - chart, S- chart and the moving average chart. The second type is the control charts for attributes. These charts are used for quality characteristics, which can be observed only as attributes classifying an item as defective or non-defective (conforming or not conforming to specifications). The P-chart and the C-chart are used in this case. The third type is the cumulative sum (CUSUM) control chart that is used primarily to maintain the current control of a process when detection of a small shift in the process is of interest.

In the next three sections, we shall briefly discuss some of the univariate control charts mentioned above.

1.3.1. $\bar{X}$ - control chart

The $\bar{X}$-chart monitors variability between sample sub-groups. Samples of size n are drawn from a production process and the sample means of the relevant quality characteristics are calculated. The means are then plotted on the $\bar{X}$-Chart. If we set the type 1 error of the test to be, say α =0.0027, then k = 3 and the 3-sigma limits of the $\bar{X}$ - chart are given by

$$UCL = \mu + 3\delta / \sqrt{n}$$

$$CL = \mu \qquad (1.2)$$

$$LCL = \mu - 3\delta / \sqrt{n}$$

The control status of the process is then determined as in figure 1.1

Generally, μ and δ are unknown and are estimated from preliminary samples taken when the process is believed to be in control.

1.3.2. R- control chart

If the measure of interest is the variance (δ^2) of the quality characteristic, several different control charts can be used. All of these control charts assume that random sample of size n is available and the characteristic is normally distributed. The range chart (R-Chart) is used to monitor process dispersion for small sample sizes (n≤10).

If $X_{i,j}$ be the j^{th} observation in the i^{th} for (j=1,2,.....,n;i=1,2,....,m) then the i^{th} sample mean is given by

$$\bar{X}_i = \sum X_{ij} / n$$

and the grand mean is given by

$$\bar{\bar{X}} = \sum \sum X_{ij} / nm$$

$$= \sum \bar{X}_i / m$$

The i^{th} sample range is given by

$$R_i = \max(x_{ij}) - \min(x_{ij}) \quad \text{for } i=1,2,.....,m$$

Similarly, the average sample range is given by

$$\bar{R} = \sum R_i / m$$

It can be shown that $E(R) = \delta_0 d_2$ and $var(R) = d_3^2 \delta_0^2$

Where δ_0 is the standard deviation of the sample characteristic assumed to be known and since most of the distribution of R is within the interval E (R) ± 3 [var (R)] the control limits for the R-Chart are given by

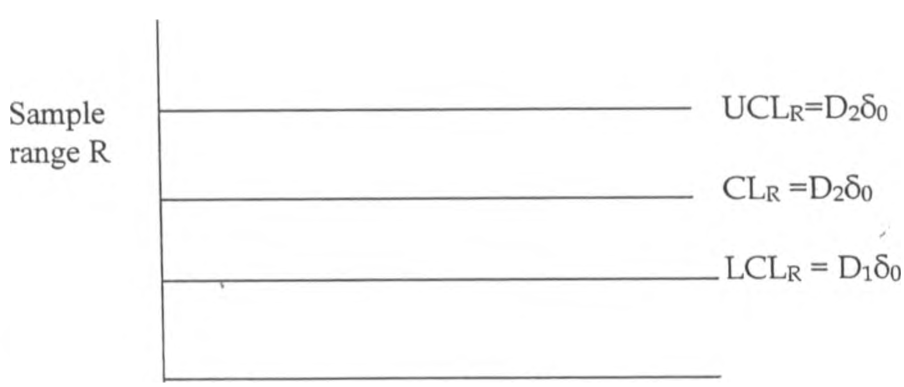
$$UCL = \delta_0 [d_2 + 3d_3] = D_2\delta_0$$

$$CL = \delta_0 D_2 \quad (1.3)$$

$$LCL = \delta_0 [d_2 - 3d_3] = D_1\delta_0$$

A 3-sigma R-chart is as in figure 1.2

R-Chart



Sample lots in order of production

Figure 1.2

Sample range falling outside the control limits implies that the process is out of control, otherwise it is in control.

1.3.3. S- control chart

If n>10, the range method for estimating δ_0 loses efficiency and we have to resort to the standard error chart (S- Chart) which also monitors process dispersion and variability within sample units. The 3-sigma control limits for the S- chart are given by

$$UCL_s = \delta_0 (C_4 + 3\sqrt{(1-C_4^2)}) = B_6 \delta_0$$

$$CL_s = \delta_0 C_4 \qquad (1.4)$$

$$LCL_s = \delta_0 (C_4 + 3\sqrt{(1-C_4^2)}) = B_5 \delta_0$$

A 3-sigma S-Chart is shown in figure 1.3

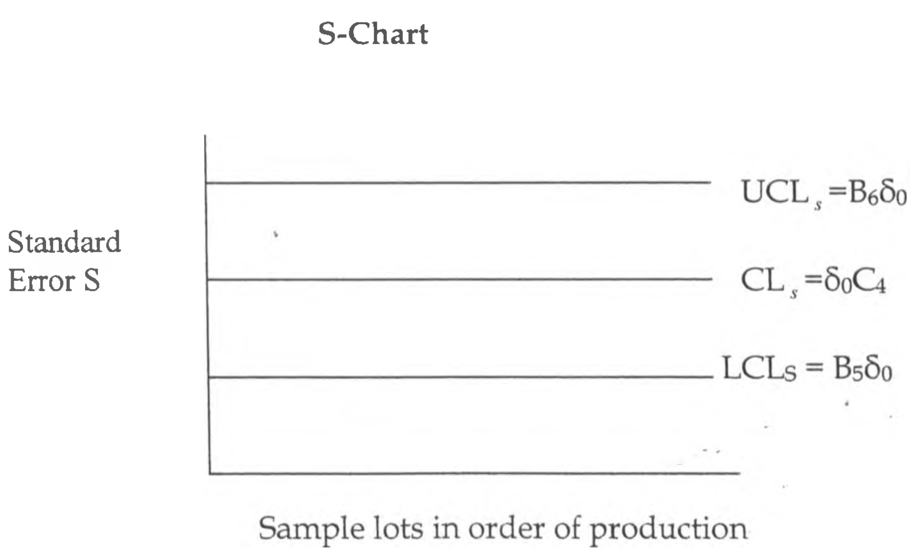


Figure 1.3

1.3.4 S^2 - control chart

This chart also monitors dispersion of a product. It is based on the sample variances. Here we assume that the sample characteristic is normally distributed. Since

$(n-1)\, S^2 / \delta_0^2$ has a chi-square distribution with (n-1) degrees of freedom, it follows that

$$P\,[\delta_0^2\, \chi^2_{n-1\,(1-\alpha/2)} / (n-1) \leq S^2 \leq \delta_0^2\, \chi^2_{(n-1),\,(\alpha/2)} / (n-1)] = 1-\alpha$$

The control limits for the S^2 -chart are then given by

$$UCL = \delta_0^2\, \chi^2_{n-1,\,(\alpha/2)\,/\,(n-1)} \qquad (1.5)$$

$$LCL = \delta_0^2\, \chi^2_{n-1,\,(1-\alpha/2)/\,(n-1)}$$

Guttman, Wilks and Hunter (1965) have pointed out that it is customary to use only an upper control limit, since the chi-square distribution is always positive. The upper control limit is given by

$$UCL = \delta_0^2\, \chi^2_{n-1}\, \alpha / (n-1)$$

Therefore, S^2 -chart is equivalent to the tests involving simple hypothesis against a simple alternative involving δ^2

$$H_0 : \delta_0^2 = \delta_0^2$$

against

$$H_1 : \delta_0^2 \neq \delta_0^2$$

The critical regions for this test are the regions above the UCL and below the LCL as in (1.5)

We note that, it is customary to constructing the charts for monitoring process dispersing i.e. the R- Chart, S- chart and S^2 -Chart. If any of these charts indicate that the process is in control, we construct the $\bar{X}$ -chart, which monitors variability between samples. If on the other hand the charts for monitoring process dispersion indicate the process is out of control, we do not need to construct the $\bar{X}$ -Chart.

1.4 Problem statement

No two objects are identical. Items produced in large quantity under the same operating conditions will differ in quality upon inspection. This variability may be due to chance or assignable causes. Chance causes are variations brought about by interacting factors which are random in nature and can neither be predicted nor controlled. Assignable causes on the other hand are variations resulting from multiplicity of factors e.g. defective or sub-standard raw materials, new operation, improper handling and setting of machines, mechanical defects e.t.c. Therefore in view of this the assignable causes can be identified, controlled and possibly eliminated where possible. When the variability of the product is only due to chance causes, the process is said to be under control (in control) otherwise it is out of control.

Chapter 2

MULTIVARIATE PROCESS CONTROL

2.1 Multi -characteristic control charts

There are many situations in which it is necessary to simultaneously monitor two or more correlated quality characteristics. Such problems are referred to as multivariate quality control problems. To be able to monitor two or more correlated quality characteristic we normally use multi-characteristics control charts.

2.2 Literature Review

The development of multi-characteristic control charts is necessarily based on Hotelling T^2-control charts. Hotelling (1974) proposed the use of the χ^2-random variable in a control chart setting for the testing of bombsights although he did not actually use χ^2-control charts because the covariance matrix (Σ_o) was unknown. His papers are primarily devoted to the case for Σ_o unknown. A fundamental assumption in the development of the χ^2-control chart is that the underlying distribution of the quality characteristics is multivariate normal.

Ghare and Torgerson (1968) discussed the use of a multi-characteristic control chart referred to as the Q-control chart to monitor the central tendency of a number of measurable quality characteristics on one control chart. The quadratic form of the multivariate normal distribution has chi-square distribution. Thus, an appropriate confidence region can be defined and a control chart constructed to monitor the stability of the pattern of variation of variables. As a special case of the Q-control chart, a bi-characteristic control chart was developed to identify the presence of a single assignable cause of variation. The Q-control chart is particularly effective when two or more quality characteristics are correlated. This is because the use of separate control charts to individually monitor each quality characteristic separately suffers from the weakness of ignoring the correlation between the variables which affects the type I error. Therefore in this case, we might erroneously conclude that a process is out-of-control when it is actually in control.

Montgomery and Klatt (1972) developed an appropriate model for the economic design of Hotelling T^2-control chart to maintain quality control for two or more related variables. They assumed that the process is subject to occurrence of a single assignable cause of variation and the time between occurrences has an exponential distribution. In order to formulate the cost function, they assumed that the state of the process (in control or out-of-control) is detected exactly at the same time a particular sample is drawn. This assumption underestimates the cost-function when the sample size is exceptionally very large and the inspection procedure is complicated. A two-stage grid search was used to find the optimal parameters of the Hotelling control chart (T^2-control chart). Montgomery and Klatt (1972) presented a cost model for a multivariate quality control procedure to determine the optimal sample size, sampling frequency and control chart for the sample means constant.

Montgomery and Klattt found that by minimizing the average run length (ARL) of an out-of-control process for a large fixed valued of the ARL of an in-control process, we can determine the sample size (n) and the control chart constant ($\chi^2{}_{p,\alpha}$) when there are two correlated quality characteristics, where p is the number of variables and α is the size of type I error) thus

(i) For a large positive correlation, ($\rho_0 > 0$), a large sample size is needed to detect large positive shifts in the sample means than small positive shifts. (ρ_0 is the correlation between the two variables)

(ii) A larger sample size is required to detect shifts for $\rho_0 > 0$ than for $\rho_0 < 0$.

Jackson (1980) presented an overview of principal components and its relation to quality control. Alt and Smith (1988) have given an excellent review of the multivate process control methods.

2.3 Multivariate control charts

To illustrate the need for multivariate control charts, consider a manufacturing plant where the product is a plastic film. Let the usefulness of the film depend on its transparency (x_1) and its tear resistance (x_2). Further assume that the two quality characteristics are jointly distributed as bivariate normal and the standard values are

μ_{01}, μ_{02}, δ_{01} and δ_{02} with a correlation ρ_0 between these two characteristics. We can therefore display these values as follows:

$$\mu_0 = \begin{bmatrix} \mu_{01} \\ \mu_{02} \end{bmatrix} \qquad (2.1)$$

$$\Sigma_0 = \begin{bmatrix} \delta_{01}^2 & \rho_0\delta_{01}\delta_{02} \\ \rho_0\delta_{01}\delta_{02} & \delta_{01}^2 \end{bmatrix} \qquad (2.2)$$

where Σ_0 is the covariance matrix.

A sample of size n is drawn from the process at regular intervals and measurements of both variables x_1 and x_2 are obtained.

If we focus our attention on monitoring the process means, one way of doing this is to ignore the correlation between the characteristics which results to type 1 error and monitor each process mean separately. For each sample of size n we take an unbiased estimate of μ_{01} which is denoted by $\bar{X}_1$ and plot it against sample lots on an $\bar{X}$-chart with the limits

$$UCL_1 = \mu_{o1} + 3\delta_{o1}/\sqrt{n}$$
$$CL_1 = \mu_{o1} \quad (2.3)$$
$$LCL_1 = \mu_{o1} - 3\delta_{o1}/\sqrt{n}$$

Here, 3-sigma limits developed by Shewhart [1931] are used to determine the control limits for the first quality characteristic.

Another $\bar{X}$-chart is also set up to monitor the process mean of the tear resistance variable (x_2).

The control limits are thus given by

$$UCL_2 = \mu_{o2} + 3\delta_{o2}/\sqrt{n}$$
$$CL_2 = \mu_{o2} \quad (2.4)$$
$$LCL_2 = \mu_{o2} - 3\delta_{o2}/\sqrt{n}$$

$\bar{X}$-chart for variable x_i (i=1, 2)

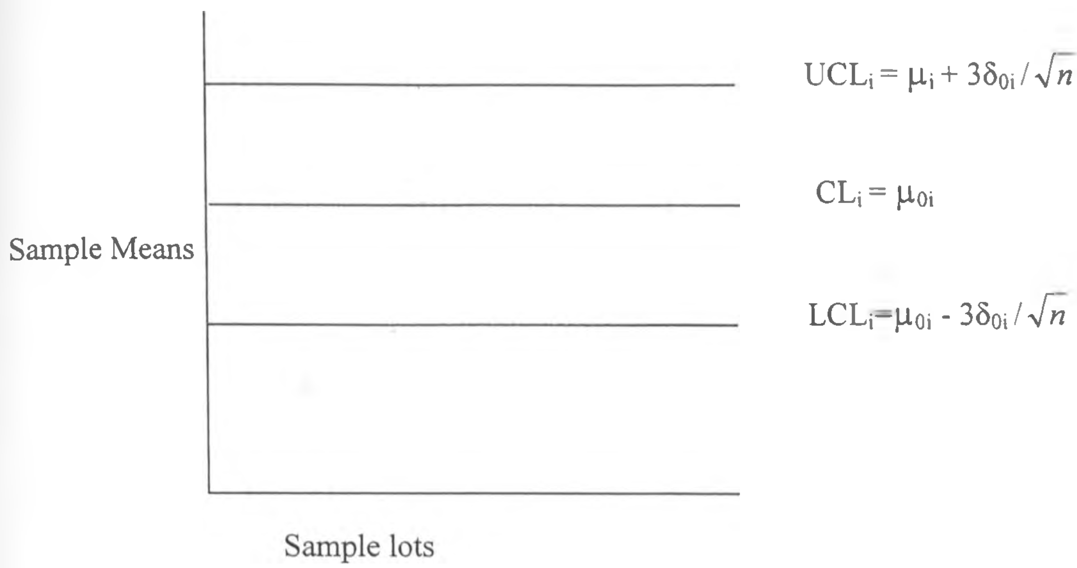


Figure 2.1

If the mean of any sample lots falls outside the respective control limits or if eight consecutive points falls above the central line, the process is said to be out of control hence, there is evidence of assignable causes of variation.

The Elliptical and Rectangular control Regions

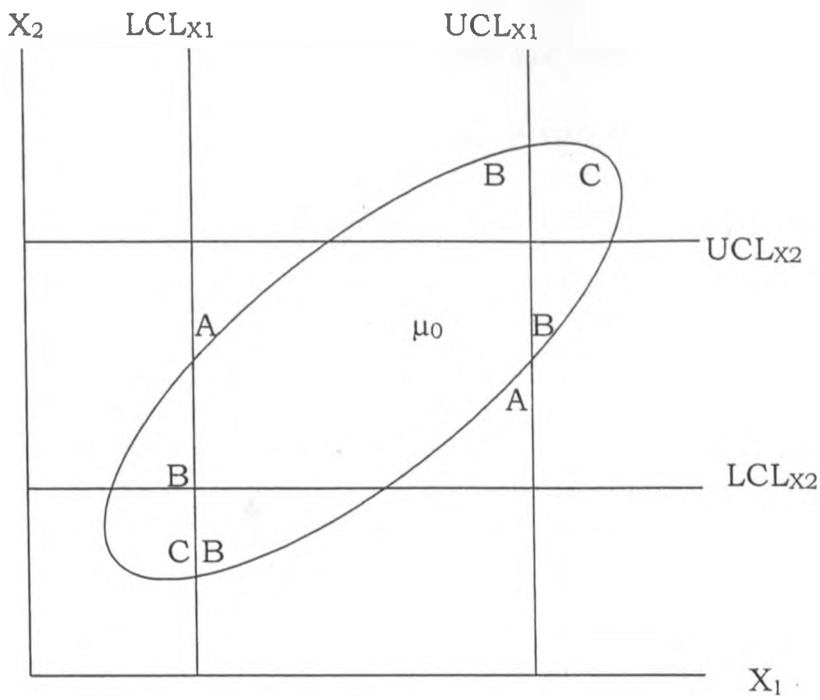


Figure 2.3

If the pair of sample means falls within the rectangular control region, the process is considered to be in control otherwise it is out of control.

The use of separate control charts or the equivalent rectangular region can be very misleading because when we ignore the correlation coefficient between the variables will results to type I error. The appropriate control region is elliptical in nature as in Figure 2.3 above. A process is considered out of control if the pair of means $(\bar{X}_1, \bar{X}_2)$ falls outside the elliptical region. On the other hand, if we use the rectangular region, we may erroneously conclude that both process means are in control which is demonstrated by region A, one is out of control and the other in control denoted by region B and both process means are out of control that is represented by region C.

In practice, individual, $\bar{X}$-charts are used in conjunction with the χ^2-chart to determine which process mean is out of control. When this is done, it is recommended that the type I error of each one of the charts be set equal to α/p where p is the number of variables and α is the overall type I error. For example when p=2 as in our case and α=0.0054, the type I error of each chart will be set at 0.0027 which means 3–sigma limits are computed as in equations (2.3) and (2.4).

In some instances, estimates of μ_0 and Σ_0 may be derived from such a large amount of past data that these values may be treated as parameters and not their corresponding estimates. Duncan (1974) states that the values for the parameters could also be selected by management to attain certain objectives. These are referred to as standard or target values. In a case where there are no set values, (standard values) μ_0 and Σ_0 are usually estimated from rational subgroups taken when the process is believed to be in control. In the sequel, control charts will be presented for both standards given and values estimated. In both cases, we will employ two characteristics for easy interpretation. These control charts will be referred to as bivariate control charts which can be extended to more than two characteristics.

2.4 Control charts for the process mean

In this section, the $\bar{X}$-chart when a process has more than one quality characteristic is considered. However, we will restrict ourselves to two quality characteristics that can be extended to three or more quality characteristics as we have mentioned above. The cases when μ_0 is known and when it is unknown are considered.

2.4.1 Control charts for the process mean when μ_0 is known

In the univariate case, the process has only one quality characteristic. If this quality characteristic is normally distributed with mean μ_0 and standard deviation δ_0, the probability that a sample mean will fall between

$$\mu_0 \pm Z_{\alpha/2}\ \delta_o / \sqrt{n} \tag{2.5}$$

is (1- α) where $Z_{\alpha/2}$ is such that

$$P(Z>Z_{\alpha/2}) = \alpha/2 \tag{2.6}$$

This is the basis for the control charts presented in equations (2.3) and (2.4) which were earlier discussed (univariate case). It is customary to use 3.0 in place of $Z_{\alpha/2}$ as proposed by Shewhart (1931), which gives

α=0.0027. In case of sample mean $\bar{X}$-plotting outside the control limits, assignable causes of variation are sought.

Suppose random samples of a given size are taken from a process at regular intervals and an $\bar{X}$-chart is maintained to determine whether or not the process mean is at the standard value μ_0. This is equivalent to repeated significance tests of the form.

$$H_o: \mu = \mu_o \quad \text{against} \quad H_1: \mu \neq \mu_o \tag{2.7}$$

Normally, the $\bar{X}$-chart is used with upper and lower control limits. Equivalently, we could use an $\bar{X}$-chart with only upper control limits (UCL). This is done by noting that

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{\operatorname{var}(\bar{X})}} \tag{2.8}$$

is distributed as a standard normal variable.

Here,

$$E[\bar{X}] = \mu_o \text{ and } \sqrt{\operatorname{var}(\bar{X})} = \delta_o/\sqrt{n}$$

which gives us

$Z = \sqrt{n}\,[\bar{X} - \mu_0]/\delta_0]$ as the standard normal variable.

Thus

$Z^2 = \sqrt{n}\,[\bar{X} - \mu_o] / \delta_o]^2$ has a chi-square distribution with 1 degree of freedom. The sample values of Z^2 are then plotted on a chart whose upper control limit is

$$\text{UCL} = \chi^2_{1,(\alpha)} \tag{2.9}$$

where $\chi^2_{1,(\alpha)}$ is the percentage point of a chi-square distribution on 1 degree of freedom.

The χ^2-chart has the disadvantage of not being able to distinguish the runs on either side of the mean. However, hypothesis testing based on χ^2-chart concept is important in that it provides the foundation for the extension of the univariate control charts.

In the present set up the univariate null hypothesis in (2.7) will be rejected if

$$\chi^2_0 > \chi^2_{1,(\alpha)} \tag{2.10}$$

where

$$\chi^2_0 = [\sqrt{n}(\bar{X} - \mu_0)\delta_o]^2 = n[\bar{X} - \mu_0]'(\delta^2_0)^{-1}(\bar{X} - \mu_0) \tag{2.11}$$

In quality control, rejection of H_0 would imply that the process is out of control. A natural generalization to a multivariate case is to reject the hypothesis in equation (2.7)

If

$$\chi^2_0 > \chi^2_{p,\alpha} \tag{2.12}$$

where

$$\chi_0^2 = n[\bar{X} - \mu_0]'\Sigma_0^{-1}(\bar{X} - \mu_0) \tag{2.13}$$

$\bar{X}$ denotes the (px1) vector of samples mean and Σ_0 is a pxp variance-covariance matrix

For p=2

$$\chi_0^2 = n(1-\rho_0^2)^{-1}[(\bar{X}_1 - \mu_{01})^2\delta_{01}^{-2} + (\bar{X}_2 - \mu_{02})\delta_{02}^{-2}$$

$$-2\rho_0\delta_{01}^{-1}\delta_{02}^{-1}(\bar{X}_1 - \mu_{01})(\bar{X}_2 - \mu_{02})] \tag{2.14}$$

Where ρ_0 is the correlation coefficient between the two variables. Equation (2.14) is that of an ellipse centered at (μ_{01}, μ_{02}) shown in figure 2.3. For bivariate quality characteristics, a control region is the interior of such an ellipse. In particular, a vector of sample means resulting in a point outside the elliptical region indicates that the process is out of control. By Bonferroni's inequality, the probability that each process mean is at the standard value is at least 1-α.

The χ^2-chart has associated with it an operating characteristic curve or equivalently a power curve. The power indicates the probability of detecting a shift in the process mean on the first sample taken after the shift has occurred.

If $\pi^{(\lambda)}$ denotes the power of the chat, then

$$\pi^{(\lambda)} = pr[\chi_p^2 > \chi_{p,\alpha}^2] \tag{2.15}$$

where χ_p^2 denotes the non central chi-square random variable with p degrees of freedom and non centrality parameter

$$\lambda = n\,(\mu-\mu_o)'\,\Sigma_o^{-1}(\mu-\mu_o)$$

For two quality characteristics

$$\lambda=n(1-\rho_0^2)^{-1}((\mu_1-\mu_{01})^2\delta_{01}^{-2}+(\mu_2-\mu_{02})^2\delta_{02}^{-2}$$

$$-2\rho_0\,\delta_{01}^{-1}\,\delta_{02}^{-1}(\mu_1-\mu_{01})(\mu_2-\mu_{02})) \qquad (2.16)$$

A fundamental assumption in the development of the χ^2-chart is that the underlying distribution of the quality characteristics is multivariate normal

2.4.2 Control charts for the process mean when μ_0 is unknown

If the mean μ_0 is unknown, it must be estimated from preliminary samples taken when the process is believed to be in control. These preliminary samples are referred to as rational subgroups. We shall denote by *m* the number of subgroups. The standard values used in section 2.4.1 are replaced by their unbiased estimates obtained from the *m* rational subgroups. For example μ_0 in equation (2.3) would be replaced by the average of the sample means obtained from each rational subgroups and any one of the several measures of variability would be used in place of δ_{01} .Hillier (1969) developed a two-stage procedure using probability limits for determining whether the data for the first m subgroups come from a process that is in control and whether future subgroup data from this process exhibit statistical control.

For each of the m subgroups, a random sample of size n is obtained and the (px1) vector of sample means ($\bar{X}_i$) is calculated so is the (PxP) sample variance–covariance matrix (S_i). If there is statistical control within each subgroup then estimates of the process mean vector and the process variance – covariance matrix are given by :-

$$\bar{X}_i = \Sigma X_{ij}/n \text{ and } S_i = \Sigma(X_{ij}-X_i)^2/(n-1)$$

therefore

$$\bar{\bar{X}} = \Sigma \bar{X}_i/m \text{ and } \bar{S} = \Sigma S_i/m$$

where $\bar{X}_i$ and S_i denote the i[th] sample mean and sample variance of the m subgroups. The overall mean and variance of the m subgroups are given by $\bar{\bar{X}}$ and $\bar{S}$ respectively.

For known standard values of μ and Σ, the test statistic is given by equation (2.13). If the values of μ and Σ are unknown, then they are replaced by their unbiased estimates and the resulting statistics is

$$T^2_{0,1} = n(\bar{X}_i-\bar{\bar{X}})'S^{-1}(\bar{X}_i-\bar{\bar{X}}) \qquad (2.17)$$

for i=1,2,...,m

where

$T^2_{0,1}$ follows the Hotellings T^2-distribution.

For two quality characteristics, equation (2.17) becomes

$$T^2_{0,1} = (n/\det(S))\,[(\bar{X}_{1,i} - \bar{\bar{X}}_1)^2 S_2^2 + (\bar{X}_{2,i} - \bar{\bar{X}}_2)^2 S_1^2$$

$$-2(\bar{X}_{1i} - \bar{\bar{X}}_1)(\bar{X}_{2,i} - \bar{\bar{X}}_2) S_{12}] \qquad (2.18)$$

$$\text{Det}(S) = S_1^2 S_2^2 - S_{12}^2$$

where

$$S_1^2 = \Sigma\, S^2_{1,i}/m$$

$$S_2^2 = \Sigma\, S^2_{2,i}/m$$

$$S_{12}^2 = \Sigma S_{12,i}/m$$

It can be shown that $T^2_{0,1}$ is distributed as

$$C_1[m,n,p,]F_{p,mn-m-p}$$

where

$F_{p,mn-m-p}$ is the F-distribution

and

$$C_1[m,n,p) = \frac{p(m-1)(n-1)}{(mn-m-p+1)} \qquad (2.19)$$

To determine whether the process is in control when the first *m* subgroups are obtained, the *m* values of $T^2_{0,1}$ are plotted on a chart with

$$UCL_x = C_1[m,n,p)\, F_{p,mn-m-p+1,\alpha}$$

$$= \frac{p(m-1)(n-1)}{(mn-m-p+1)} F_{p,mn-m-p+1,\alpha} \qquad (2.20)$$

$$LCL_x = 0$$

If $T^2_{0,1}$, for one or more of the m initial subgroups falls out of control, such subgroups are discarded. Control limits are then recalculated using the remaining subgroups. $\bar{\bar{X}}$ and $\bar{S}$ are also recomputed and new control limits are determined with (m minus the number of discarded rational subgroups). In the case (p=1) equation (2.19) becomes

$$C_1(m,n,1) = \frac{(m-1)(n-1)}{(mn-m)} = \frac{(m-1)(n-1)}{m(n-1)}$$

therefore

$$UCL_{x1} = \frac{(m-1)}{m} F_{1,m(n-1),\alpha} \qquad (2.21)$$

and

$$T^2_{0,1} = \frac{n(\bar{X}_i - \bar{\bar{X}})^2}{S^2} \qquad (2.22)$$

where S^2 is the average of the sample variances obtained from each subgroup.

$$F_{1,m(n-1),\alpha} = t^2_{m(n-1),\alpha/2}$$

therefore

$$1-\alpha = pr[(\bar{X}_1 - \bar{\bar{X}})^2/S^2 \le [(m-1)/m) F_{1,m(n-1),\alpha}]$$

$$= pr[(\bar{X}_1 - \bar{\bar{X}}) \le \sqrt{(S^2((m-1)/m))}\, t_{m(n-1),\alpha/2}]$$

$$1-\alpha = pr[\bar{\bar{X}} - A_4\sqrt{S^2} \le \bar{X}_i \le \bar{\bar{X}} + A_4\sqrt{S^2}] \qquad (2.23)$$

where

$$A_4 = \sqrt{((m-1)/m)}\, t_{m(n-1),\alpha/2}$$

thus, the multivariate result reduces to the univariate one, hence the intervals for the individual characteristics are obtained by using

$$A_4 = \sqrt{((m-1)/m)}\, t_{m(n-1),\alpha/2p}$$

and for P = 2, the upper and lower control limits for each variable are given by

$$\overline{\overline{X}} \pm A_4 \sqrt{S^2}$$

2.5 Control charts for the process dispersion

In Multivariate situation, it is usually desired that the covariance matrix of the process remains at a standard value Σ_o. This is checked by taking a random sample of size n and the value of some sample statistic is determined from the (pxn) data matrix. Thus if S denotes the (pxp) sample variance-covariance matrix.

$$S = \begin{bmatrix} S_{11} & S_{12} \cdots & S_{1p} \\ S_{21} & S_{22} \cdots & S_{2p} \\ S_{p1} & S_{p2} \cdots & S_{pp} \end{bmatrix}$$

where the diagonal elements $S_{ii} = S_i^2$ (i=1,2,....,p) are the sample variances and the off diagonal elements $S_{ij}=S_{ji}$ ($i \neq j$=1,2,...,p) are the sample covariance's.

For bivariate characteristics we get

$$S = \begin{bmatrix} s_1^2 & s_{12} \\ s_{21} & s_2^2 \end{bmatrix}$$

where

$$S_1^2 = \sum_{l=1}^{2} (x_{1k} - x_1)^2$$

$$S_{12} = \sum_{l=1}^{2} (x_{1k} - x_1)(x_{2k} - x_2)$$

$$S_{21} = \sum_{l=1}^{2} (x_{1k} - x_1)(x_{2k} - x_2)$$

$$S_2^2 = \sum_{l=1}^{2} (x_{2k} - x_2)^2$$

The sample correlation coefficient for the i^{th} and j^{th} variables is given by

$$Y_{ij} = S_{ij} / S_i S_j$$

The generalized sample variance denoted by $|S|$ is a commonly used scalar measure of multivariate dispersion. For bivariate characteristics

$$|S| = S_1^2 S_2^2 - S_{12}^2 = S_1^2 S_2^2 (1 - r_{12}^2).$$

Johnson and Wichern (1982) pointed out that one of the properties of $|S|$ is that distinctly different covariance matrices can have the same generalized variance. In view of the last property it is recommended that any procedure based on $|S|$ is to be accompanied by the appropriate univariate procedures to monitor dispersion.

2.5.1 Control charts for the process dispersion when Σ_0 is known

2.5.1.1 The $|s|^{1/2}$-control chart when Σ_0 is known

The $|S|^{1/2}$-control chart is the multivariate analogue of the S-chart. This control chart makes use of two approaches. In the first approach, we make use of the distributional properties of $|S|^{1/2}$. For bivariate quality characteristics, it can be shown that

$2(n-1)\ |S|^{1/2} / |\Sigma_o|^{1/2}$ is distributed as χ^2_{2n-4}. In view of this expression, the control limits for the $|S|^{1/2}$-chart are as follows:

$$UCL = [|\Sigma_o|^{1/2}\ \chi^2_{2n-4,\alpha/2}]\ /2(n-1)$$

$$\tag{2.24}$$

$$LCL = [|\Sigma_o|^{1/2}\ \chi^2_{2n-4,1-\alpha/2}]\ /2(n-1)$$

where

$$|\Sigma_o|^{1/2} = \delta_{01}\delta_{02}\sqrt{(1-\rho_0^2)}.$$

therefore, for each random sample of size n,

$|S|^{1/2} = (S_1^2\ S_2^2 - S_{12}^2)^{1/2} = S_1S_2\sqrt{(1-r_{12}^2)}$ is computed.

If

$$|S|^{1/2} > UCL$$

or

$$|S|^{1/2} < LCL,$$

then dispersion of the process is said to be out of control thus assignable causes are to be sought. The exact distribution of $|S|^{1/2}$ for more than two quality characteristics is unknown.

The second approach to the construction of the $|S|^{1/2}$ control chart utilizes only the first two moments of $|S|^{1/2}$ and the property that most of the distribution of $|S|^{1/2}$ is within three standard deviations of its expected value. Since

$$|S| = (n-1)^{-p} |\Sigma_o| \pi \chi^2_{n-k},$$

where the chi-square random variables are independent, it follows that

$$E[|S|^r] = (n-1)^{-pr} |\Sigma_o|^r \pi \Gamma(r+(n-k)/2)/ \Gamma[(n-k)/2]$$

It follows therefore that

$$E[|S|^{1/2}] = |\Sigma_o|^{1/2} (2/(n-1)^{p/2} \Gamma(n/2)/ \Gamma[(n-p)/2]$$

$$= |\Sigma_o|^{1/2} b_3 \qquad (2.25)$$

and

$$E[|S|] = |\Sigma_o| (n-1)^{-p} \pi(n-k)$$

$$= |\Sigma_o| / b_1 \qquad (2.26)$$

Now

$$Var[|S|^{1/2}] = E[|S|] - (E[|S|^{1/2})]^2$$

$$= |\Sigma_o| (b_1 - b_3^2) \qquad (2.27)$$

with this, the upper and lower limit for a $|S|^{1/2}$-control chart are given respectively by

$$UCL = E[|S|^{1/2}] + 3\sqrt{var[|S|^{1/2}]}$$

$$LCL = E[|S|^{1/2}] - 3\sqrt{var[|S|^{1/2}]}.$$

It follows that the control limits for Σ_0 are given by

$$UCL = |\Sigma_o|^{1/2} (b_3 + 3\sqrt{(b_1 - b_3^2)}$$

$$CL = |\Sigma_o|^{1/2} b_3 \qquad (2.28)$$

$$LCL = |\Sigma_o|^{1/2} (b_3 - 3\sqrt{(b_1 - b_3^2)}$$

In the univariate case, $b_3 = c_4$, $b_1 = 1$,

$|\Sigma_o|^{1/2} = \delta_0$ and the control limits in (2.28) reduce to those stated in the univariate case(1.1). In the bivariate case

$b_1 = (n-2) / (n-1)$

$b_3 = (2/(n-1) (\Gamma(n/2)/ \Gamma[(n-2)/2)]$.

If n=10, $b_1 = b_3 = 0.889$ and $b_1 - b_3^2 = 0.099$. Then, the UCL=1.831$|\Sigma_o|^{1/2}$ and since LCL is negative, hence UCL=0

2.5.1.2 The s²-control chart when Σ_o is known

The S^2 –control chart is also used in the multivariate case to monitor process dispersion. This is equivalent to repeated tests of significance. Anderson (1984) has shown that the likelihood ration test for

$H_o: \Sigma = \Sigma_o$

against

$H_1: \Sigma \neq \Sigma_o$

modified to be unbiased (the power of the test is greater than or equal to the significance level) is based on the following statistic:

$$W^* = -p(n-1) - (n-1)\ln[|S|] + (n-1)\ln[|\Sigma_0|] + (n-1)\,tr(\Sigma_0^{-1}S) \qquad (2.29)$$

where $tr(\Sigma_0^{-1}S)$ is the sum of the diagonal elements of $\Sigma_0^{-1}S$. For bivariate quality characteristics

$$tr(\Sigma_0^{-1}S) = (1-\rho_0^2)^{-1}[S^2_1/\delta^2_{01}) + (S^2_2/\delta^2_{02}) - 2\rho_0(S_{12}/\delta_{01}\delta_{02})]$$

He further showed that W* is asymptotically distributed as $\chi^2_{p(p+1)/2}$ and has presented the upper 5% and 1% points for the exact distribution of W*. For two quality characteristics and n-1=9 degrees of freedom, the upper 5% and 1% percentage points are 8.52 and 12.38 respectively. In our case the process dispersion is considered out of control at the 5% level if W*>8.52 and at the 1% level if W*>12.38. A natural generalization is that a process dispersion will be out of control if W*>UCL at the specified level of significance.

We notice that $|S|^{1/2}$ is plotted on the charts described by both equation (2.24) and (2.28). The difference between the two charts is that the control limits in equation (2.24) and (2.28) are probability limits and 3-sigma limits respectively. We also note that the range chart is used to monitor the variability of each quality characteristics but the multivariate analogue is not presented, as it is relatively intractable.

2.5.2 Control charts for the process dispersion when Σ_O is unknown

When there are multiple quality characteristics, two variations of the $|S|^{1/2}$ -chart were presented for monitoring process dispersion when Σ_O is known. The first was a probability limit chart with control limits as stated in equation (2.24). These limits are applicable only when there are two quality characteristics.

Let $|S^*|^{1/2}$ denote the average of the square roots of the generalized sample variance. This implies that

$$|S^*|^{1/2} = 1/m\Sigma\ |S_i|^{1/2}.$$

then, $|S^*|^{1/2}/\ b_3$ is an unbiased estimate of $|\Sigma_O|^{1/2}$ thus the control limits for process dispersion when Σ_O is unknown are as follows:

$$UCL = [\,|S^*|^{1/2}\ \chi^2_{2n-4,(\alpha/2)}]2b_3(n-1) \qquad (2.30)$$

$$LCL = [\ |S^*|^{1/2}\ \chi^2_{2n-4,(1-\alpha/2)}]2b_3(n-1)$$

where the constant b_3 is as defined earlier. The chart for $|S^*|^{1/2}$ when Σ_O is known uses the 3-sigma limits already discussed in the previous section and is applicable for any number of quality characteristics.

Therefore for Σ_O unknown, the control limits are obtained by substituting $|S^*|^{1/2}$ b_3 for $|\Sigma_O|^{1/2}$ in equation (2.28).

This results to the following limits for process dispersion when Σ_0 is unknown:

$$UCL = |S^*|^{1/2} \left(1+(3/b_3)\sqrt{(b_1-b_3^2)}\right)$$

$$CL = |S^*|^{1/2} \quad (2.31)$$

$$LCL = |S^*|^{1/2} \left(1-(3/b_3)\sqrt{(b_1-b_3^2)}\right)$$

For the univariate case the limits given in (2.31) for a $|S^*|^{1/2}$- control chart are identical to those for the S-chart(2.19) with the b_3 and b_4 factors stated as

$$C_2\,(m,n,p) = p(n-1)\,(m+1)\,/\,(mn-m-p+1)$$

Another procedure which could be used to investigate process dispersion when Σ_o is unknown is obtained from equation (2.29). This is the likelihood ratio statistic for testing

$$H_o: \Sigma = \Sigma_o$$

against

$$H_1: \Sigma \neq \Sigma_o$$

To obtain the corresponding procedure in this section, we need the unbiased estimates of $|\Sigma_o|^{1/2}$ and Σ_o^{-1}.

If we let $|S_0|$ denote the average of the generalized sample variances from the *m* rational subgroups, then

$$|S_o| = 1/m\, \Sigma |S_i|.$$

Using the result in equation (2.26), it can be verified that $|S_o|/b_1$ is an unbiased estimate of $|\Sigma_o|$.

Further, let S_i^{-1} denote the inverse of the sample variance-covariance matrix for subgroup i, i=1, 2, ...,m.

Kshirsagar (1972) showed that $(n-p-2)\, S_i^{-1}/(n-1)$ is an unbiased estimate of Σ_o^{-1}.

Therefore if

$$S^* = 1/m \, \Sigma \, S_i^{-1},$$

then

$$(n-p-2)\, S^{*-1}/(n-1)$$

is also an unbiased estimate of Σ_o^{-1} which is obtained for the *m* rational subgroups. Therefore to obtain the procedure for determining the control limits for this section, we substitute $|S_o|/b_1$ for $|\Sigma_o|$ and $(n-p-2)S_*/(n-1)$ for Σ_o^{-1} in equation (2.29). The revised values of W*, i=1,2,...m, would still be plotted on a control chart with

$$UCL = \chi^2_{p(p+1)/2}$$

The control limit factors used in this section for both p=1 and p>1 are independent of the number of subgroups. However, at times these factors depend on the number of subgroups (i.e. they should be a function of *m*)

Chapter 3

DATA COLLECTION AND ANALYSIS

3.1 Simulation study

In this section, we generated 25 samples of size 10 each from a motor production company where the products are piston rings of various diameters (X_1) and thickness (X_2).The two quality characteristics are jointly distributed as bivariate normal with means μ_{01} = 30, μ_{01}= 15, variances δ_{01}^2 = 8, δ_{02}^2 = 4 and correlation coefficient ρ = 0.5.

Table 3.1 provides data for $\overline{X}$ and χ^2 control charts for the means of X_1 and X_2 at α = 0.05.

Table 3.2 provides data for $\overline{X}$ and χ^2 control charts for the means when they have been disturbed as indicated 3.1.1.

Table 3.3 provides data for dispersion control charts namely R-chart, S-chart, S^2-chart, $|S|^{1/2}$-chart (probability chart), $|S|^{1/2}$-chart (1.96-sigma limits chart) and W*-Chart at α = 0.05.

Table 3.4 provides data for control charts in Table 3.3 when the variances have been disturbed as indicated 3.1.1.

Control chart data for the means of $X_1$and X_2

Sample Number	$\overline{X}_1$	$\overline{X}_2$	χ_0^2
1	30.546	14.204	3.63331
2	30.488	15.528	0.71887
3	29.711	14.171	1.86531
4	29.494	15.274	1.00377
5	28.883	13.903	3.20267
6	28.855	14.440	1.71905
7	29.771	15.635	1.77423
8	31.301	16.123	3.58110
9	29.898	15.039	0.03179
10	29.836	14.483	0.73594
11	29.299	15.020	0.85338
12	29.870	15.103	0.10014
13	30.240	14.394	1.66292
14	29.000	15.691	4.88697
15	39.632	14.528	0.55891
16	29.237	15.674	3.69666
17	29.880	14.709	0.22396
18	30.590	14.396	2.63617
19	29.259	14.718	0.68769
20	29.260	14.697	0.6920
21	29.221	14.938	0.91037
22	30.090	14.420	1.25787
23	29.844	15.196	0.24068
24	30.740	14.293	3.81197
25	30.372	15.432	0.47394

Table 3.1

$UCL_1 = 32.68$ $UCL_2 = 16.90$ $UCL = \chi^2_{2,0.05} = 5.99$

$CL_1 = 30.00$ $CL_2 = 15.00$

$LCL_1 = 27.32$ $LCL_2 = 13.10$ $\rho_0 = 0.5$ (correlation coefficient)

Table 3.1 shows an in-control situation as all values of $\overline{X}_1$ and $\overline{X}_2$ lie within their respective control limits and the values of $\chi_0^2 < UCL = 5.99$

3.1.1 Means and variances when disturbed.

a For these samples μ_{01} is increased by 30 + 0.5i for i = 1,........,10

b For these samples μ_{02} is increased by 15 + 0.5i for i = 1,........,10

c For these samples, μ_{01} and μ_{02} are increased by 30 + 0.5i and 15 + 0.5i respectively for i = 1,......., 5

Samples 1 to 10 the variance is shifted as (δ_{01}^2 = 8 × 0.2i,for i= 1,........,10).

Samples 11 to 20 the variance is shifted as (δ_{02}^2 = 4 × 0.2i, for i= 1,......, 10)

Samples 21 to 25 the variances are shifted as (δ_{01}^2 = 8 × 0.2i and δ_{02}^2 = 4 × 0.2i, for i=1,......,5).

Control chart data when means are disturbed.

Sample Number	$\overline{X}_1$	$\overline{X}_2$	χ_0^2
1 [a]	29.562	14.152	1.526
2 [a]	29.708	15.528	1.683
3 [a]	30.321	14.171	2.320
4 [a]	31.337	15.274	7.562 [f]
5 [a]	31.439	13.903	19.213 [f]
6 [a]	32.893 [d]	14.440	13.167 [f]
7 [a]	32.907 [d]	15.635	17.231 [f]
8 [a]	33.117 [d]	16.123	36.782 [f]
9 [a]	34.234 [d]	15.039	18.607 [f]
10 [a]	34.675 [d]	14.483	25.453 [f]
11 [b]	29.299	14.528	26.741 [f]
12 [b]	29.870	15.609	34.191 [f]
13 [b]	30.240	15.920	21.311 [f]
14 [b]	29.000	16.962 [e]	13.819 [f]
15 [b]	39.632	16.978 [e]	41.132 [f]
16 [b]	29.237	17.313 [e]	15.234 [f]
17 [b]	29.880	17.401 [e]	29.327 [f]
18 [b]	30.590	18.211 [e]	32.103 [f]
19 [b]	29.259	18.347 [e]	16.263 [f]
20 [b]	29.260	18.912 [e]	16.816 [f]
21 [c]	29.142	15.674	3.193
22 [c]	29.614	19.123	4.121
23 [c]	31.312	16.432	2.342
24 [c]	32.009	16.919 [e]	12.183 [f]
25 [c]	32.238	17.021 [e]	14.389 [f]

Table 3.2

UCL = $\chi^2_{2,\,0.05} = 5.99$

The values in $\overline{X}_1$ chart marked by *d* falls outside the control limits indicating an out of control status in the process.

The valued in $\overline{X}_2$ chart marked by *e* falls outside the control limits indicating an out control situation in the process.

The values in χ_0^2 chart marked by *f* shows an out of control situation since their value are greater than UCL=5.99

Control chart data for process dispersion.

Sample Number	R_1	R_2	S_1	S_2	S_{12}	$\|S\|$	$\|S\|^{½}$	W*
1	8.09	6.22	2.74	1.97	1.03	28.08	5.30	1.311
2	6.71	6.83	2.05	2.04	2.65	10.47	3.24	2.629
3	6.36	7.18	1.90	2.09	2.51	9.17	3.03	3.850
4	9.67	8.23	2.83	2.69	5.34	29.44	5.43	2.552
5	8.79	6.18	2.83	2.30	3.23	31.93	5.65	0.459
6	6.42	7.52	1.96	2.18	0.28	18.18	4.26	3.922
7	8.32	4.96	2.38	1.47	2.71	4.90	2.21	5.526
8	7.95	3.81	2.49	1.17	1.194	4.72	2.17	5.924
9	7.29	5.85	2.60	1.66	0.39	18.48	4.30	1.928
10	6.75	5.10	2.36	1.50	2.76	4.91	2.22	5.527
11	11.22	8.93	2.12	2.05	0.26	18.82	4.34	2.983
12	8.34	6.56	2.66	2.32	3.04	28.84	5.37	0.655
13	5.93	5.38	2.35	1.54	2.45	7.09	2.66	3.172
14	6.79	5.08	3.37	1.38	1.61	19.04	4.36	3.413
15	6.84	7.47	2.34	2.16	3.61	12.51	3.54	2.369
16	8.02	7.05	2.42	2.37	3.56	20.22	4.50	1.62
17	7.72	5.91	2.58	1.91	0.66	23.85	4.88	1.582
18	10.98	3.06	2.91	0.99	1.63	5.64	2.38	7.215
19	7.71	4.04	2.92	1.42	3.05	7.89	2.81	4.420
20	9.28	4.95	3.66	1.77	4.42	22.43	4.74	2.721
21	7.43	5.69	2.40	1.96	1.52	19.82	4.45	0.659
22	13.2	6.08	3.39	1.74	2.72	27.40	5.23	1.373
23	11.30	9.23	2.37	2.35	1.28	29.38	5.42	2.451
24	6.92	5.61	2.43	2.63	1.21	39.38	6.28	4.581
25	8.59	5.21	2.54	2.15	0.133	29.71	5.45	2.921

Table 3.3

Control limits for the process dispersion in table 3.3.

Chart	Quality Characteristic I	Quality Characteristic 2
R-Chart	$UCL_1 = 15.471$	$UCL_2 = 10.94$
	$CL_1 = 8.712$	$CL_2 = 6.16$
	$LCL_1 = 1.949$	$LCL_2 = 1.378$
S-Chart	$UCL_1 = 4.720$	$UCL_2 = 3.337$
	$CL_1 = 2.751$	$CL_2 = 1.945$
	$LCL_1 = 0.779$	$LCL_2 = 0.52$
S^2 -Chart	$UCL_1 = 15.04$	$UCL_2 = 7.52$

In table 3.3 the process is in control because all the values for the process dispersion charts falls within their respective control limits as shown above.

Control chart data for process dispersion when variances are disturbed.

Sample Number	S_1	S_2	S_{12}	$\|S\|$	$\|S\|^{1/2}$	W*
1	1.264	2.000	1.03	5.33	2.31	7.744
2	1.789		2.65	5.78	2.40	5.989
3	2.191		2.51	12.90	3.59	1.460
4	2.530		5.34	2.91	1.71	11.259
5	2.828		3.23	21.56	4.64	0.318
6	3.098		0.28	38.31	6.19	3.590
7	3.347		2.71	37.47	6.12	1.042
8	3.578		1.94	47.44	6.89	2.952
9	3.793		0.39	57.40	7.58	6.902
10	4.000		2.76	56.38	7.51	4.455
11	2.288	0.894	0.26	6.32	2.51	7.872
12		1.263	3.04	3.52	1.88	9.604
13		1.549	2.45	13.19	3.63	1.381
14		1.789	1.61	23.00	4.86	0.562
15		2.000	3.61	18.96	4.35	0.456
16		2.191	3.56	25.72	5.07	0.220
17		2.366	0.66	44.33	6.66	3.864
18		2.530	1.63	48.53	6.97	3.401
19		2.683	3.05	48.27	6.95	2.830
20		2.828	4.42	44.42	6.66	3.069
21	1.264	0.894	1.52	1.03	1.01	11.903
22	1.789	1.263	2.72	2.29	1.51	6.959
23	2.191	1.549	1.28	9.88	3.14	1.675
24	2.530	1.789	1.21	19.02	4.36	0.726
25	2.828	2.000	0.33	31.88	5.65	2.738

Table 3.4

Control limits for process dispersion in table 3.4

Probability Limits for $|S|^{1/2}$ - Chart ($\alpha = 0.05$)

UCL= 7.851

LCL= 1.880

1.96-Σ Limits for $|S|^{1/2}$ - Chart

UCL= 7.3750
CL= 4.3544
LCL= 0

W* - Chart ($\alpha = 0.05$)

UCL = 8.52

Table 3.4 indicates an out–of control situation for samples 9 and 10 using 1.96-Σ limits on $|S|^{1/2}$ - chart.

The probability limits for $|S|^{1/2}$-chart shows an out–of control status for samples 4, 21 and 22. Both $|S|^{1/2}$-chart and W*- chart register out – of control on samples 4 and 21.

It seems that $|S|^{1/2}$-chart with the probability limits has performed better than the other charts for W*-Chart, this is not surprising since it is based on a large sample test statistic. A better picture would probably emerge if larger studies involving more than 100 samples were conducted since this reduces the sub-group variability.

Chapter 4

PRINCIPAL COMPONENTS AND FACTOR ANALYSIS

4.1 Residuals associated with principal components

Most statistical techniques involve operations on single response variables such as weight, pressure specific gravity, temperature, concentration. However, there are a number of occasions where more than one response variable is of interest in a problem .These variables should be studied collectively so as to take advantage of the information about the relationships among them. This is the field of multivariate analysis as pointed out earlier. Multivariate techniques are extension of univariate techniques such as t-tests or the analysis of variance.

Principal components and factor analysis are two techniques, which are finding increasing application by quality engineers who are concerned with processes with more response variable. This method is a procedure in its own right. It is used to simplify simultaneous interpretation of a number related variable.

Principal components are used as a data reduction technique, a diagnostic tool as well as a control device. In this case, we are concerned with control situation using a method of principal components together with its associated residual analysis.

In the early days of principal component analysis, most of the attention was devoted to ways of obtaining characteristic root and vectors from a covariance matrix and the interpretation of the root and vectors. Effort has since then shifted to the problems of inference, such as estimation and test of hypothesis concerning these parameters to deeper examination of new tools for the application of these techniques.

The improved understanding of principal components as a data reduction tool, their role in applications such as regression analysis and multivariate quality control and the availability of high speed computer and convenient software packages have made incorporation of principal components techniques in routine data analysis not only feasible but also common.

This increased use implies imperatives such as model fit questions in general and examination of residuals in the model in particular. The problem is similar to those in regression analysis and need a test for outliers

4.2 Matrix algebra to principal components analysis

Given two related quality characteristics, we can carry out the method of principal components in order to check the control status of the process. We first obtain the means of each quality characteristics, the variances and the covariance between the two qualities. In this case, the means will be donated by

$$\mathbf{X} = \begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \end{bmatrix}$$

and the covariance matrix is

$$S=\begin{bmatrix} S_{11}^2 & S_{12} \\ S_{12} & S_{22}^2 \end{bmatrix}$$

where $\bar{X} = \Sigma x_{ij}$ is the mean of the i^{th} quality characteristics and

$S_{ij} = (n\Sigma x_{ik}\, x_{jk} - \Sigma x_{ik}\Sigma_{jk})/n(n-1)$ is the covariance between the i^{th} and the j^{th} quality characteristics

The correlation coefficient r between x_i and x_j is given by

$$r_{ij} = \frac{Cov(X_i, X_j)}{\sqrt{(\text{var}(X_i)}\sqrt{\text{var}(X_j)}}$$

The important point from matrix algebra related to the method of principal component is that a pxp symmetric non-singular matrix such as the covariance matrix S may be reduced to a diagonal matrix L by premultiplying and post multiplying by a particular orthonormal matrix U as shown below:

$$L = U'SU \tag{4.1}$$

The diagonal elements of L are $I_1, I_2, \ldots, I_p$ which are the characteristics roots, or eigenvalues of S. The columns of the orthonormal matrix U denoted by $U_1, U_2, \ldots, U_p$ are the characteristics vectors or eigenvectors of S.

The characteristic root can be obtained without using formula (4.1).

This can be done by using the characteristic equation

$$|S-1I|=O \qquad (4.2)$$

where I is a 2x2 identity matrix i.e.

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

On evaluating equation (4.2) we get a p^{th} degree polynomial in I from which the values of characteristics roots are obtained.

for p=2, we have

$$|S-I1| = \begin{bmatrix} S_1^2 - 1 & S_{12} \\ S_{12} & S_2^2 - 1 \end{bmatrix} = 0$$

that is

$$S_{12}^2 - (S_1^2 S_2^2 - S_1^2 1 - S_2^2 1 + 1^2) = O$$

We can therefore solve for 1 whose values will be 1_1 and 1_2.The characteristic vectors may be obtained by solving the equations

$$[S-I_iI)\ t_i = 0 \qquad (4.3)$$

and

$$U_i = \frac{t_i}{\sqrt{t_i' t_i}} \qquad (4.4)$$

for i =1, 2,..., p

Equation (4.3) is evaluated as

$$\begin{bmatrix} S_1^2 - I_1 & S_{12} \\ S_{12} & S_2^2 - I_1 \end{bmatrix} \begin{bmatrix} t_{11} \\ t_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

These are homogeneous equations with two unknowns therefore, in order to solve them, assume t_{11} =1 which helps us to work with one equation. Therefore

$$U_1 = \frac{t_1}{\sqrt{t_1{}'t_1}}$$

Similarly, we can use l_2 and let t_{22}=1 hence

$$|S - I_2 I|\ t_2 = \begin{bmatrix} S_1^2 - I_2 & S_{12} \\ S_{12} & S_2^2 - I_2 \end{bmatrix} \begin{bmatrix} t_{21} \\ t_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$U_2 = \frac{t_2}{\sqrt{t_2{}'t_2}}$$

The matrix U= $[u_1 | u_2]$ is orthonormal hence

$u_1{}'u_1 = 1, u_2{}'u_2 = 1$ and $u_1{}'u_2 = 0$.

Therefore

$$U'SU = \begin{vmatrix} I_1 & 0 \\ 0 & I_2 \end{vmatrix} = L$$ is a diagonal matrix.

Equation (4.2) and (4.3) are used for small values of p. For large values of p, iterative procedures for obtaining characteristic vectors are used.

4.3 Principal components to statistical analysis

The sample covariance matrix S is the basis for the statistical applications of the method of the principal components. For p-quality characteristics the matrix S is given as shown below

$$S = \begin{bmatrix} S_1^2 & S_{12} \dots\dots & S_{1p} \\ S_{21} & S_2^2 \dots\dots & S_{2p} \\ & & \\ S_{p1} & S_{p2} \dots\dots & S_p^2 \end{bmatrix}$$

Where the leading diagonal elements S_i^2' s are the variances of the i^{th} variables and the off diagonal elements S_{ij} are the covariance's between the i^{th} and j^{th} variables. If the covariances are non-zero, then the variable are related otherwise they are not.

A principal axis transformation transforms p correlated variables $x_1, x_2, \dots, X_p$ into p new uncorrelated variables $z_1, z_2, \dots, z_p$ where the coordinate axes are described by the vectors u_i. Thus,

$$Z = U'(X - \bar{x}) \quad (4.5)$$

where x and $\bar{x}$ are px1 vector of the original variables and their means respectively .These transformed variables are the principal components of x. The i^{th} principal component is denoted by

$$Z_i = u'(x - \bar{x}) \quad (4.6)$$

which has mean zero and variance l_i .For example, if p=2,

Then

$$Z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = U'(X - \bar{X})$$

Therefore

$$\text{Var}(Z_1) = l_1, \ \text{var}(Z_2) = l_2$$

If you wish to transform a set of variables x by a linear transformation

$Z=U'(x-\bar{x})$, whether U is orthonormal or not, the covariance matrix of the original variables by the formula

$$S_z=U'SU \tag{4.7}$$

Although we have said that U is orthonormal, this is not a sufficient condition for the Z_i'S to be independent. Only a transformation such as principal axis transformation will produce an S_z, which is a diagonal matrix L as earlier stated. The fact that S_z is diagonal means that principal components are uncorrelated. If the coefficients for the first vector are nearly equal and both positive, this means that the first principal component is related to the variability which both measurements have in common .If there are no correlated errors of measurements, it would be assumed to represent process variability. For diagnostic purpose, we determine the correlation of each of the original variables.

The correlation of the i^{th} principal component z_i and j^{th} original variable x_j is given by

$$Yz_i x_j = U_{ji}\frac{\sqrt{I_i}}{S_j} \tag{4.8}$$

Principal component have also another property in that equation (4.5) can be re-written as

$$X=\bar{X}+U_z \tag{4.9}$$

Since U is orthonormal that is $U^{-1}=U^{1}$.This helps us to determine the original data if we know the values of the principal components. In other words each observation is made up of a linear combination of the principal components. If we know the population covariance matrix Σ, we would operate on it just as with S. The characteristic roots of Σ would be obtained by

1', 2' , ………., p'

The characteristic vectors associated with the roots would be the population values.

4.4 Scaling of principal components

We have two ways to scale characteristic vectors as indicated in the following equations

(i) $V_i = \sqrt{l_i}\,u_i$ (4.10)

(ii) $W_i = u_i / \sqrt{l_i}$ (4.11)

The first transformation gives

$$V'V=L \qquad (4.12)$$

which means that the vectors are orthogonal but not of unit length.

Further,

$$L^2 = V'\,SV \qquad (4.13)$$

which implies that the transformation given by the following expression

$$\sqrt{l_i}\,z_i = V_i'(x-\bar{X}) \qquad (4.14)$$

will produce a new variable with zero mean and variance l_i^2 which is uncorrelated. These are now principal components, which are scaled to their respective characteristic root rather than a unit length as in equation (4.5)

Another useful relationship

$$S = VV' \tag{4.15}$$

shows that the covariance matrix can be obtained directly from its characteristics vectors. Scaling the principal components as in equation (4.15) is useful since the components will be in the original units of the problem if the x variables were all in the same units.

Using the transformation in (4.11), we have

$$Y_i = W_i'(x - \bar{x}) \tag{4.16}$$

from which we have

$$L^{-1} = W'W \tag{4.17}$$

and

$$L = W'SW \tag{4.18}$$

The variables obtained from this transformation are uncorrelated and have unit variance.

Relations (4.5), (4.14) or (4.16) are used to express principal components as regards scaling. These relations differ only by a scale factor. U-vectors are desirable from a diagnostic point of view since the vectors have the advantage that the coefficients are restricted from -1 to +1.

V-vectors have the advantage that the principal components in that mode are expressed in terms of the units of the original variables.W-vectors produce components that have variance.

4.5 Generalized measures of principal components to variability

We present two ways to describe the variability of a set of related variables namely

(i) the determinant of the covariance matrix $|S|$

(ii) the sum of the variances of the variables

$$S_1^2+S_2^2 \ldots . +S_p^2$$

here $S_1^2+S_2^2+\ldots .+S_p^2=$ trace of S written as TrS. Trace of S means the sum all leading diagonal elements of S which are the variances.

There are other measures of generalized variability but the above two are commonly used. An important property of principal components is that the variability as specified by any measure is preserved as follows

$$|S|=|L| \tag{4.19}$$

where the determinants are related to the area or volume generated by asset of variables .Equivalently,

$$TrS=TrL \tag{4.20}$$

Relation (4.20) implies that the sum of the characteristic roots (the variances of the principal components) is equal to the sum of the variances of the original variables.

This shows that the characteristic roots may be treated as components of variance. If we obtain the ratio of each characteristic roots to the total, we obtain a proportion of the total variability associated with each principal component.

4.6 Principal components for quality control

When we use the method of principal components for quality control, one has to calculate the deviations from the means and then use any of the three principal component transformation given by equations (4.5),(4.14) or (4.16) to calculate the measure of overall variability T^2 where

$$T^2 = Y'Y \qquad (4.21)$$

which follows the Hotellings T^2-distribution as discussed earlier.

for p=2

$$T^2 = Y_1^2 + Y_2^2$$

The scaling for principal components adopted is a matter of choice .The scaling in (4.16) is preferred since it gives a set of principal components with unit variances. Control charts for these components can be constructed except that individual components are controlled instead of their means and the tabulated value of T^2 is given by

$$T^2_{p,\,n,\,\alpha} = \frac{P(n-1)}{n-P} F_{p,,n-p,\alpha} \qquad (4.22)$$

Therefore, any observation vectors with values of $T^2 > T^2_{p, n, \alpha}$ will imply an out of control situation on the T^2-chart. We point out that T^2 –chart has only an upper limit since T^2 is always positive. It is worth noting, that the control charts for the principal components should be used together with the T^2-chart .If T^2 is in control, one should proceed normally otherwise if it is out of control, one should revert to the individual control chart for the principal components so as to determine the nature of the trouble, that is to check which variable cause an out of control status.

A control ellipse can also be used to judge whether a process is out of control or otherwise .However, this is only possible for two quality characteristics. It is important to note that T^2 can be obtained directly from the original variables without the use of principal components. However, the principal component approach is easier to handle and has many desirable properties that cannot be obtained directly from the original variables especially when a larger number of variables is under study.

4.7 Extension of principal components to more than two variables.

In the previous sections, we have discussed the approach of principal components when there are two correlated quality characteristics. In this section, we briefly outline how the method of principal component is applied in the case of more than two variables. It is necessary to note that the principal component method works well for a large number of variables however, the two variable situations has an advantage in that, the operation and relationships can be demonstrated in a simple manner.

A greater part of the computational procedures to obtain the roots and vectors in a sequential manner is as follows, we begin with the largest root and its associated vector, the next largest and so on. When we scale these V-vectors the variability left unexplained by the first principal component is $S-V_1V'_1$, the variability unexplained by the second two is $S-V_1V'_1-V_2V'_2$ and so on.

If the covariance is matrix is of full rank p, there will be p positive characteristic roots whose rank will be in descending order .If however, the covariance is not of full rank say r<p, there will be r positive root and the remaining p-r roots will be zero. This occurs when one or more linear relationships exist among the original variables so that knowledge of a subset of them would allow us to determine the remainder without error. In such a case, we will require r principal components to reconstruct the original data.Suppose it is given that p=2, this is extended to p=3 by setting $X_3=X_1+X_2$ and immediately the first two variables are obtained, the third variable would be known by default. The covariance matrix would therefore be

$$V_1V_1'+V_2V_2'=\begin{bmatrix} S_1^2 & S_{12} & S_{13} \\ S_{21} & S_1^2 & S_{23} \\ S_{31} & S_{32} & S_1^2 \end{bmatrix} = VV'$$

where v has only two columns instead of three as the two principal components will completely reconstruct the original data.

It is customary to have the first few roots, say $k<p$ explain most of the total variability where we use percentage of TrS as a yardstick .If the remaining (p-k) roots are equal, it means that the remaining (p-k) vectors are not different and are all of the same length with arbitrary rotation. The closer together the roots are, the larger will be the standard errors. If the remaining (p-k) roots are equal and do not differ significantly such that we use only k vectors, then it is impossible to reproduce the exact covariance matrix. However, if the percentage of the total trace represented by these remaining roots is small, the first k vectors will be used adequately for reconstruction.

4.8 Simulation study for principal components

In this section, we generated one sample of size 10 each from a motor production company where the products are piston rings of various diameters(X_1) and thickness(X_2).The two quality characteristics are jointly distributed as bivariate normal with means μ_{01}=30,μ_{02}=15,variances δ_{01}^2=8, δ_{02}^2=4 and correlation coefficient ρ=0.5.

A and B represent process shift with μ_{01}=32, μ_{02}=17, and μ_{01}=33, μ_{02}=19, respectively.

C represents recording error whereas D is an outlier.

Observations

Sample number	x_1	x_2
1	27.54	14.32
2	35.35	14.38
3	31.67	15.69
4	27.87	15.23
5	29.46	14.72
6	28.28	16.38
7	33.58	13.92
8	29.93	15.18
9	28.93	15.83
10	27.26	15.28
A	35.93	19.10
B	37.02	18.67
C	26.23	12.18
D	36.12	18.36

Table 4.1

Principle components data chart

mple number	$X_1 - \bar{X}_1$	$X_2 - \bar{X}_2$	Z_1	Z_2	$Z_1\sqrt{l_1}$	$Z_2\sqrt{l_2}$	Y_1	Y_2	T^2
1	-2.44	-0.77	-0.06	0.50	-0.24	1.09	0.36	0.31	0.23
2	5.37	-0.71	3.74	3.07	15.17	6.70	0.84	-0.08	0.71
3	1.69	0.60	2.68	2.72	10.87	5.93	0.95	0.44	1.10
4	-2.11	0.14	-1.02	-0.71	-4.14	-1.55	-0.14	0.61	0.39
5	-0.52	-0.37	-1.65	-1.83	-6.69	-3.99	-0.69	-0.63	0.87
6	-1.75	1.29	-1.70	-1.59	-6.90	-3.47	-0.51	0.04	0.26
7	3.60	-1.17	3.83	3.60	15.54	7.85	1.16	-0.03	0.21
8	-0.05	0.90	-1.14	-1.16	-4.63	-2.53	-0.40	-0.20	0.20
9	-1.05	0.74	-0.05	0.18	-0.20	0.39	0.14	0.53	0.30
10	-2.72	0.19	-4.63	-4.79	-18.78	-10.45	-1.70	-0.98	3.85
A	5.95	4.01	2.27	1.06	-4.31	-3.16	-2.61	2.21	11.70
B	7.04	3.58	1.28	-3.12	1.30	-3.63	2.32	-3.36	13.03
C	3.75	-2.91	-1.70	2.13	-0.18	1.41	-4.03	-2.08	16.25
D	6.14	3.27	1.53	-1.77	-3.89	2.81	3.74	2.04	18.27

Table 4.2 $T^2_{2,\ 10,\ 0.05}$=10.04

In Table 4.2 samples 1 to 10 shows an-in control status since T^2 values are less than critical value 10.04.

Samples A, B, C and D indicate an out-of control situation with T^2-critterion as T^2>10.04, therefore control charts for Y_1 and Y_2 would be constructed to determine the nature of the trouble. These in conjunction with charts for X_1, X_2 and T^2 Could be used as diagnostic checks on assignable causes such as outliers, recording errors and process shift which fail to be signaled by the charts for the individual characteristic x_1 and x_2.

4.9 Conclusion

Given some data from a production process, we start by constructing the process dispersion charts namely the R-chart, S-chart, and S^2-chart, which monitor the variability within samples. If the process dispersion registers an in-control status, we proceed to construct the $\overline{X}$-chart, which monitors variability between samples

The numerical results in table 3.1, shows that for those points, which registered an in-control status using χ^2 –chart, all the individual charts registered the same .Further, those points which registered an out-of-control situation with the χ^2 –chart registered the same with at least one of the individual charts. It would therefore seem reasonable to construct the individual charts to determine which quality characteristic is the cause of the trouble. The $|S|^{1/2}$-chart with the probability limits seems to perform better than the $|S|^{1/2}$ –chart using 1.96 –sigma limit and W*-chart.

4.10 Areas of further Research

In the principal component analysis, in case the T^2 -chart registers an out-of control status, one should revert to the individual chart to discover the nature of the trouble.

Further, since our project is based on bivariate process control in which we restricted ourselves to two quality characteristics, we recommend that more research should be carried out on more than two quality characteristics.

REFERENCES

1. Alt, F.B. (1973). **Aspects of multivariate control charts.** *M.SC. thesis, Georgia Institute of Technology, Atlanta.*
2. Alt, F.B. (1982). **Bonferroni inequalities and intervals.** *In S. Kotz and N.L. Johnson eds., Encyclopedia of Statistical Sciences vol.1, John Wiley, New York, pg 294-300.*
3. Alt, F.B. (1985). **Multivariate Quality control.** *In S. Kotz and N.L. Johnson eds., Encyclopedia of Statistical Sciences vol.1, John Wiley, New York, pg 110-122.*
4. Alt, F.B. and Smith N.D. (1988). **Multivariate Process Control.** *In P.R. Krishnrah all C.R. Rao eds., Quality Control and probability: Handbook of Statistics, Amsterdam:-North Holland, vol.7, pg 333-351.*
5. Anderson, T.W. (1984). **An Introduction to Multivariate Statistical Analysis.** *2nd ed., Wiley, New York.*
6. Bowker, A.H. and Lieberman, G.J. (1959). **Engineering Statistics**. *Prentice-Hall, Englewood Cliffs, NJ.*
7. Duncan, A.H. (1974). **Quality Control and Industrial Statistics.** *4th ed. Richard D. Irwin, Homewood, IL.*
8. Ghare, P. M. and Torgerson, P.E. (1968). **The Multicharacteristic Control Chart.** *The Journal of Industrial Engineering, pg 269-272.*
9. Guttman, I., Wilks, S.S. and Hunter. (1965). **Introductory Engineering Statistics.** *Wiley, New York.*
10. Hillier, F.S. (1969). **X and R chart control limits based on a small number of subgroups**. *Journal of Quality Technology vol.1, pg 17-26,*

11. Hotelling, H. (1947). **Multivariate quality Control, Illustrated by the air testing of sample bombsights.** ***In; C. Eisenhart, H. Hastay, and W.A. Wallies, eds., Techniques of Statistical Analysis. McGraw-Hill, New York, pg 111-184.***

12. Jackson, J.E. (1980). **Principal Components and Factor Analysis: Part I – Principal Components.** ***Eastman Kodak Company, Rochester, New York 14650.***

13. Johnson, R.A. and Wichern, D.W. (1982). **Applied Multivariate Statistical Analysis.** ***Prentice-Hall, Englewood Cliffs, NJ.***

14. Kshirsagar, A.M. (1972). ***Multivariate Analysis. Marcel Dekker, New York.***

15. Montgomery, D.C. and Klatt, P.J. (1972). **Economic Design of T^2 – control charts to maintain current control of a process.** ***Management Science vol.19, pg 76-89.***

16. Montgomery, D.C. (1985). **Introduction to Statistical Quality Control.** ***Wiley, New York.***

17. Morrison, D.F. (1967). **Multivariate Statistical Methods.** ***McGraw- Hill Book Con., New York.***

18. Schilling, E.G. and Nelson, P.R. (1967). **The Effect of** ***Journal of Quality Technology vol. 8***. **Nonnormality on Control Limits of X-chart.**

19. Shewhart, W.A. (1931). **Economic Control of Quality of Manufactured Products.** ***D.Van Mostrand, New York.***