UNIVARIATE AND MULTIVARIATE CONTROL CHARTS FOR MONITORING SUGAR PRODUCTION PROCESS

Zetty Azrah bt Sutirman, PM. Dr. Ismail Mohamad

1.1 INTRODUCTION

Quality control is a system to maintain quality of product or service to achieve specification standard of product. One of the most powerful tools is through graphical method which is control chart. This is because control chart easy to analyse the data and able to provide comprehensive information on existing product or process characteristics. There are two types of control chart. First, statistical process control (SPC) and second is multivariate statistical process control (MSPC).

Statistical process control commonly referred as SPC, was developed by Dr. Walter A. Shewhart in the mid-1920s. In general, statistical process control is to control and monitor the process of production line and detect abnormal process. However , the Shewhart control chart can only monitor single process variable at a time.

Multivariate statistical process control , MSPC was established by Hotelling in his 1947 pioneering paper. MSPC can simultaneously control and monitor more than one process variables at a time. The three most popular multivariate control statistics of multivariate control charts, such as Shewhart charts ($\bar{x}$ and Range charts) ,cumulative sum plots (CUSUM), and exponentially weighted moving average charts (EMWA) .

1.2 PROBLEM STATEMENT

Nowadays, statistical process control techniques, SPC are widely used in industry. However the characteristics of univariate quality control chart or Shewhart control chart itself , that can only monitor single process variable at a time are inadequate to control the process stability . This problem may affect the process variables and quality of the product. Usually, in industry, there are many situations in which the simultaneous monitoring or control in two or more related quality process characteristics is necessary. Monitoring these process variables independently can mislead the true process situation. MSPC charts overcomes this situation by considering the correlation between the variables and are able to analyze the stability of the process. In this research, we investigate method of multivariate controls charts and univariate control charts to identify a significant for monitoring and controlling the process.

1.3 LITERATURE REVIEW

1.3.1 Statistical Process Control (SPC)

SPC was pioneered by Walter A. Shewhart at Bell Laboratories in the early 1920s. Shewhart developed the control chart in 1924 and the concept of a state of statistical control. Statistical uses statistical methods. SPC is applied in order to monitor and control a process. Monitoring and controlling the process ensures that it operates at its full potential.The goal of SPC is to achieve higher quality of final product by elimintation of variability in the process while the main objective of SPC is to quickly detect the occurrence of assignable causes of abnormal process so that further investigation to the process and corrective measurements can be carried out. There are many ways to implement process control. The basic quality control tools include histogram , check sheet , pareto chart , cause and effect diagram , defect concentration diagram , scatter diagram and quality control charts.

1.3.2 Multivariate Statistical Process Control (MSPC)

Nowadays, in industry, there are many situations in which the simultaneous monitoring or control of two or more related quality–process characteristics is necessary. Monitoring these quality characteristics independently can be very misleading. Process monitoring of problems in which several related variables are of interest are collectively known as multivariate statistical process control. The most useful tool of multivariate statistical process control is the quality control chart.

1.4 OBJECTIVE

The objectives of this study are :

(i) To compare the ability of univariate control chart and multivariate chart using Hotelling's T^2 statistics in detecting out of control points

(ii) To construct the multivariate control chart using the HotellingT^2 statistics.

(iii) To identify significant method for monitoring multivariate process variables by using quality control charts.

1.5 METHODOLOGY

This chapter discuss about the methodology of the research . We will discuss the basic knowledge of control charts and the distributions of all the observation data. It concerns on univariate control chart and multivariate control chart by using HotellingT^2 for monitoring the quality of an industrial production process and detecting out of control points. The purpose is to determine whether this charts identified the same points are the out of control points.

1.5.1 Anderson- Darling Normality Test

All the observations should be tested wheter the data is associated with the normal distribution. The normality use in this study is Anderson Normality Test (AD). The test rejects the hypothesis of normality when the *p-value* is less than to 0.05 or confidence interval the null hypothesis is likely to be false. Failing the normality test allows you to state with 95% confidence the data does not fit the normal distribution. Passing the normality test only allows you to state no significant departure from normality was found. The Anderson-Darling test is defined as:

H_0 : The data follow a normal distribution

H_1 : The data do not follow the normal distribution

The AD test statistic is defined as

$$A^2 = - N - S$$

Where,

$$S = \sum_{i=2}^{N} \frac{(2i-1)}{N} [\ln F(Y_i) + \text{In}(1 - F(Y_{N+1-i}))]$$

F is the cumulative distribution function and Y_i is the ordered data.

1.5.2 HotellingT^2 statistics

Harold Hotelling was introduced Hoteling's T^2distribution which is a multivariate analogue of the univariate Student's *t*distribution . Harold Hotelling became the first person to discover the problem in analysing correlated variables in bombsight data from the perspective of the statistical control . He has controlled the process by using charting statistics and then the statistic charting was known as HotellingT^2.

Suppose a random sample of size ,*n* from normal distribution distribution with mean , μ variance σ^2 selected. Then,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(n-1)$$

(1.5.2.1)

This test statistic has a student t distribution with n -1 degrees of freedom. Where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the simple mean, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is the corresponding sample variance. The square of *t* statistic is given by

$$\boldsymbol{t^2 = \frac{(x - \mu)^2}{(s/\sqrt{n}\,)^2} = \frac{(x - \mu)^2}{s^2/n}}$$

(1.5.2.2)

$$\boldsymbol{t^2 = n(\bar{x} - \mu\,)^2\,(s^2)^{-1}(\bar{x} - \mu)}$$

(1.5.2.3)

Next, the Hotelling extended from the univariate statistic to multivariate statistic. When Equation 3.3 is generalized to *p* variables, T^2 follows*F* distribution as:

$$\boldsymbol{T^2 = n(\bar{x} - \mu)'(S)^{-1}(\bar{x} - \mu) \sim \frac{p(n-1)}{n-p} F_{(p,n-p)}}$$

(1.5.2.4)

Where *n* is the simple size, μ is the vector size *p*, $\bar{X}$ and *S* are the sample estimators of mean and covariance matrix are defined by:

$$\boldsymbol{\mu}' = [\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \dots \dots, \boldsymbol{\mu_n}]$$

(1.5.2.5)

$$\bar{X}_{(p\times 1)} = \frac{1}{n}\sum_{i=1}^{\mathrm{n}} x_{\mathrm{i}}$$

(1.5.2.6)

$$\boldsymbol{S}_{(\boldsymbol{p}\times\boldsymbol{p})} = \frac{\mathbf{1}}{\boldsymbol{n}-\mathbf{1}}\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} (\boldsymbol{x}_{\mathbf{i}} - \overline{\boldsymbol{x}})(\boldsymbol{x}_{\mathbf{i}} - \overline{\boldsymbol{x}})'$$

(1.5.3.7)

A control chart will be plotted based on the HotellingT^2statictics from all the observation against number of observation of time and compared to the upper control limit.[6]

$$\mathrm{UCL} = \frac{p(n+1)(n-1)}{n(n-p)} \; .F_{(\alpha\,;p\,,n-p)}$$

(1.5.2.8)

1.5.3 Mason Young Tracy Decomposition

MYT decompositions includes orthogonal components . These components consist of a series of conditional and unconditional condition T^2 terms . The general T^2 statistics for a *p*-dimensional observation vector *X'* = $(x_1, x_{2,} \dots x_3)$ can be presented as

$$T^2 = (X - \bar{X})'\mathrm{S}^{-1}(X - \bar{X})$$

(1.5.3.1)

Suppose the vector $(\boldsymbol{X}-\overline{\boldsymbol{X}})$is partitioned as

$$(\boldsymbol{X}-\overline{\boldsymbol{X}}) = [(\boldsymbol{X}^{(\mathbf{p-1})}-\overline{\boldsymbol{X}}^{(\boldsymbol{p-1})},(\boldsymbol{x}_{\mathbf{p}}-\overline{\boldsymbol{x}}_{\mathbf{p}})]'$$

(1.5.3.2)

Where $\boldsymbol{X}^{(\boldsymbol{p-1})'} = (\boldsymbol{x_1},\boldsymbol{x_2}, \ldots \boldsymbol{x_{p-1}})$represents the ($p$-1)-dimensional variable vector excluding the pth variable $\boldsymbol{x_p}$and$\overline{\boldsymbol{X}}^{(p-1)}$ represents the corresponding elements of the mean vector.

Partitioning the matrix S so that

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_{\mathrm{XX}} & \boldsymbol{s}_{x\mathrm{X}} \\ \boldsymbol{s}'_{xX} & \boldsymbol{s^2}_{p} \end{bmatrix}$$

(1.5.3.3)

Where $\boldsymbol{S}_{XX}$ is the (p -1) × (p-1) covariance matrix for the first (p-1) variables ,s_p^2 is the variance of $\boldsymbol{x}_{\mathbf{p}}$ and $\boldsymbol{s}_{xX}$ is a (p-1)-dimensional vector containing the covariances between $\boldsymbol{x}_p$ and the remaining (p-1) variables .

The T^2 statistics can be partitioned into two independent parts given by

$$\boldsymbol{T^2} = \boldsymbol{T^2}_{(\boldsymbol{p-1})} + \boldsymbol{T^2}_{\boldsymbol{p},\mathbf{1},\mathbf{2}\ldots\ldots\boldsymbol{p-1}}$$

(1.5.3.4)

where

$$\boldsymbol{T^2}_{(\boldsymbol{p-1})} = (\boldsymbol{X}^{(\boldsymbol{p-1})}-\overline{\boldsymbol{X}}^{(\boldsymbol{p-1})})'\boldsymbol{S}_{\boldsymbol{XX}}^{-1}(\boldsymbol{X}^{(\boldsymbol{p-1})}-\overline{\boldsymbol{X}}^{(\boldsymbol{p-1})})$$

(1.5.3.5)

and the second term

$$\boldsymbol{T}^{2}_{\boldsymbol{p}.\mathbf{1},\mathbf{2},\ldots\,\boldsymbol{p}-\mathbf{1}} = {(\boldsymbol{x_p} - \overline{\boldsymbol{x}}_{\boldsymbol{p}.\mathbf{1},\mathbf{2}\ldots\boldsymbol{p}-\mathbf{1}})^{2}}\Big/{\boldsymbol{s}^{2}_{\boldsymbol{p}.\mathbf{1},\mathbf{2}\ldots\ldots\boldsymbol{p}-\mathbf{1}}}$$

(1.5.3.6)

where

$$\overline{\boldsymbol{x}}_{\boldsymbol{p}.\mathbf{1},\mathbf{2}\ldots\boldsymbol{p}-\mathbf{1}} = \overline{\boldsymbol{x}}_{\boldsymbol{p}} + \boldsymbol{B}'_{\boldsymbol{p}}(\boldsymbol{X}^{(\boldsymbol{p}-\mathbf{1})} - \overline{\boldsymbol{X}}^{(\boldsymbol{p}-\mathbf{1})}$$

(1.5.3.7)

and

$$\boldsymbol{B}'_{\boldsymbol{p}} = \boldsymbol{S}^{-\mathbf{1}}_{\boldsymbol{XX}}\boldsymbol{s}_{\boldsymbol{xX}}$$

(1.5.3.8)

While the conditional variance is given as

$$\boldsymbol{s}^{2}_{\boldsymbol{p}.\mathbf{1},\mathbf{2}\ldots\boldsymbol{p}-\mathbf{1}} = \boldsymbol{s}^{2}_{\boldsymbol{p}} - \boldsymbol{s}'_{\boldsymbol{xX}}\boldsymbol{S}^{-\mathbf{1}}_{\boldsymbol{XX}}\boldsymbol{s}_{\boldsymbol{xX}}$$

(1.5.3.9)

Continuing to iterate and partition the MYT decomposition of a T^2 statistic is given by

$$\boldsymbol{T}^{2} = \boldsymbol{T}^{2}_{\mathbf{1}} + \boldsymbol{T}^{2}_{\mathbf{2}.\mathbf{1}} + \boldsymbol{T}^{2}_{\mathbf{3}.\mathbf{1},\mathbf{2}} + \cdots \boldsymbol{T}^{2}_{\boldsymbol{p}.\mathbf{1},\mathbf{2}\ \ldots\ldots\boldsymbol{p}-\mathbf{1}}$$

(1.5.3.10)

From the above equation , the T_1^2 is the unconditional components of the T^2 statistic. Meanwhile ,$T^2_{2.1,}T^2_{3.1,2,}\ldots,T^2_{p.1,2\ldots p-1}$ is the conditional components . The other ways and easier approach of computing the terms of the MYT decomposition is given by

$$\boldsymbol{T}^{2}_{\boldsymbol{x_1},\boldsymbol{x_2}\ldots\ \boldsymbol{x_j}} = (\boldsymbol{X}^{(\boldsymbol{j})} - \overline{\boldsymbol{X}}^{\boldsymbol{j}})'\boldsymbol{S}^{-\mathbf{1}}_{\boldsymbol{jj}}(\boldsymbol{X}^{(\boldsymbol{j})} - \overline{\boldsymbol{X}}^{(\boldsymbol{j})}$$

(1.5.3.11)

Where $X^{(j)}$ represents the appropriate subvector ,$\bar{X}^{(j)}$ is the corresponding subvector mean and S_{jj} denotes the corresponding covariance submatrix obtained from the overall S matrix by deleting the unused rows and columns.

However , the T_1^2 term can be computes as

$$\boldsymbol{T_1^2 = (x_1 - \bar{x}_1)^2/s_1^2}$$

(1.5.3.12)

Hence , the MYT decomposition can be computed as follows

$$\boldsymbol{T^2_{p.1,2,\ldots,p-1} = T^2_{(x_1,x_{\ldots\ldots},x_p)} - T^2_{(x_1,x_2\ldots\ldots x_{p-1})'}}$$

$$\boldsymbol{T^2_{p.1,2,\ldots,p-2} = T^2_{(x_1,x_2\ldots\ldots,x_{p-1})} - T^2_{(x_1,x_2,\ldots\ldots,x_{p-2})'}}$$

....

....

....

$$\boldsymbol{T^2_{2.1} = T^2_{(x_1,x_2)} - T^2_1}$$

$$\boldsymbol{T_1^2 = (x_1 - \bar{x}_1)^2/s_1^2}$$

Therefore, this technique will make the better confirmation whcih variables effect to the enlargement of limit in T^2 statistic control chart that lead to the out of control points.

1.6 RESULT AND DISCUSSION

The case study of sugar production process will be conducted. The process measurements are made on five variables which is steam temperature (X_1) , Sugar density (X_2) , Cool temperature (X_3) , Sugar length (X_4) and Sugar weight (X_5).

Table 1.6.1: Sugar production process

Number of observation	X_1	X_2	X_3	X_4	X_5
1	850	53	170	21	29.2
2	847	53	170	21.2	29.2
3	848	53	170	21.5	29.2
4	850	54	170	22.1	29.1
5	849	54	170	21.7	29.2
6	850	54	170	21.8	29.1
7	848	53	170	21.3	29.2
8	848	53	170	21.5	29.1
9	848	54	170	22	29.1
10	850	54	170	21.9	29.1
11	848	53	171	21	29.1
12	848	53	171	20.4	29.1
13	847	53	170	19.8	29.1
14	849	53	170	21.4	29.1
15	846	54	170	21.8	29.1
16	844	53	171	21.3	29.1
17	843	53	171	20.7	29.1
18	842	53	170	21.5	29.1
19	841	53	170	21.5	29.1
20	842	53	171	21.3	29.1

21	843	53	171	21.8	29.1
22	841	54	168	22.2	29.1
23	850	54	170	22	29.2
24	846	54	170	21.8	29.2
25	843	54	170	16	29.2
26	846	52	170	21.5	29.2
27	845	53	170	20.8	29.2
28	847	53	170	21.1	29.2
29	843	53	170	21.4	29
30	844	53	171	20.8	28.9
31	845	53	170	21.4	28.9
32	843	54	170	21	28.9
33	842	53	170	20	28.9
34	845	53	171	18.9	29
35	842	52	171	20.1	29
36	844	53	171	16.9	29
37	844	52	171	19.8	29
38	844	52	171	21.5	29
39	844	53	170	21.3	29
40	844	53	170	21.3	29
41	844	53	170	21.4	29
42	850	53	170	21	29
43	846	53	170	21	29

44	844	53	169	17	29
45	843	52	170	20.9	29.1
46	844	53	170	21.5	29.1
47	845	54	170	18.3	29.2
48	843	53	170	20.6	29.2
49	842	53	170	21	29.2
50	845	53	170	21	29.2

1.6.1 Result of Anderson Darling test and Univariate Control Charts

Table 1.6.2 showed that the summary of sugar production process data. The figures are used to check the normality of each variables or quality characteristics based on histogram and normality is testing using Anderson Darling test. The distribution of the data may identify either fit the normal distribution if $p - value < \alpha$ the data is not normal but if $p - value \geq \alpha$, the data is said to follow the normal distribution.

Next, it showed that the summary of univariate control chart sugar production process. The figures are used to check the out of control points of five quality characteristics. Points that are outside the control region or control limits indicate that the process is out of statistical control.

Table 1.6.2: Result of Sugar Production Process

Quality Characteristics/Variables	*Normality test*	*Univariate Control Chart*
Steam temperature, X_1	*Not Normal*	*In Control*
Sugar density, X_2	*Not Normal*	*In Control*
Cool Temperature, X_3	*Not Normal*	*Out of Control*
Sugar length, X_4	*Not Normal*	*Out of Control*

Sugar weight, X_5	*Not Normal*	*In Control*

*From the Table 1.6.2, all the variables*Steam temperature (X_1), Sugar density (X_2), Cool Temperature, (X_3), Sugar length (X_4), and Sugar weight (X_5) showed they are come from not normal distribution. In univariate control charts, there are several observations from), Cool Temperature, (X_3), Sugar length (X_4) had detected out of control limits.

1.6.2 Result of Muitivariate Control Charts using HotellingT^2 statistics

***Table 1.6.3**: Summarize T^2 Statistical Control Charts*

No of variables	*Variables*	*The capability detecting out of control points*
Two	X_3 and X_4	*Capable*
Three	X_1 , X_3 and X_4	*Capable*
	X_2 , X_3 and X_4	*Capable*
	X_3 , X_4 and X_5	*Capable*
Four	X_1 , X_2 , X_3 and X_4	*Capable*
	X_1 , X_3 , X_4 and X_5	*Capable*
	X_2 , X_3 , X_4 and X_5	*Capable*
Five	X_1 , X_2 , X_3 , X_4 and X_5	*Capable*

From the Table 1.6.2, we can see the movement of capability in multivariate control charts to detect out of control points. It observed that multivariate control charts by using Hotelling T^2 statistics is capable to detect the out of control points for two, three, four and five variables. This shows that Hotelling T^2 able to detect the out of control points for each variables and does not require MYT Decomposition method to detect the signals.

1.7 CONCLUSION

In this study, we discussed univariate control chart and multivariate control chart. In the beginning of this research, we manufactured product involve several quality characteristics or variables which when univariate control is used, the control chart must be constructed for each variable which is cumbersome process because in real production process is very complex it may involve several process variables. Univariate chart can only monitor a single process variable. The multivariate control charts using HotellingT^2 statistics overcomes this problem by monitor all the process variables in a single control chart.

The univariate control charts are constructed with a ±3 standard deviation. The univariate control charts detect that X_3 and X_4 are out of control since there are points outside the control limits. The multivariate charts were constructed using HotellingT^2 statistics by combination of different number of variables.

Supposedly, from the provious research on the same topic found that the HotellingT^2have some weaknessess to detect the out of control points during the number of varibales increase. The reason for this weaknessess because the control region for the T^2 become larger as the control limit becomes greater to the enlargement of the value upper control limit.

But from the results we gained from this reasearch, we found that HotellingT^2 become capable to detect the $X_3$and$X_4$to beout of control points by meant multivariate control charts is good in detecting out of control points. However, it is known that there are some out of control points which shown by univariate control charts. Maybe the factor of sample size of data and the number of variables used in this research causes the results quite different from the previous reaseacrh.