

Multivariate Control Charts: The Hotelling T² Control Chart

A control chart normally monitors one variable over time. Perhaps this variable is machine uptime, a product characteristic, or on-time delivery. There are times, however, when the simultaneous monitoring of two or more related variables is important. The group of control charts that do this are called multivariate control charts. The most familiar one of these is the Hotelling T² control chart or just the T² control chart. This control chart is introduced in this publication.

In this issue:

- [Introduction to the T² Control Chart](#)
- [Constructing a T² Control Chart](#)
- [Out of Control Points](#)
- [Summary](#)
- [Quick Links](#)

Introduction the T² Control Chart

In 1947, Harold Hotelling introduced a statistic which allowed multivariate observations to be plotted on a single chart. This statistic is now called Hotelling's T² statistic. The statistic combines information from the mean as well as the dispersion of more than one variable. The calculations, which include some matrix algebra, are more difficult than those of "normal" control charts. This was a barrier to using multivariate control charts until software that could perform the calculations came along.

The T² control chart is used to detect shifts in the mean of more than one interrelated variable. The data can be in subgroups (like the \bar{X} -R control chart) or the data can be individual observations (like in the X-mR control charts).

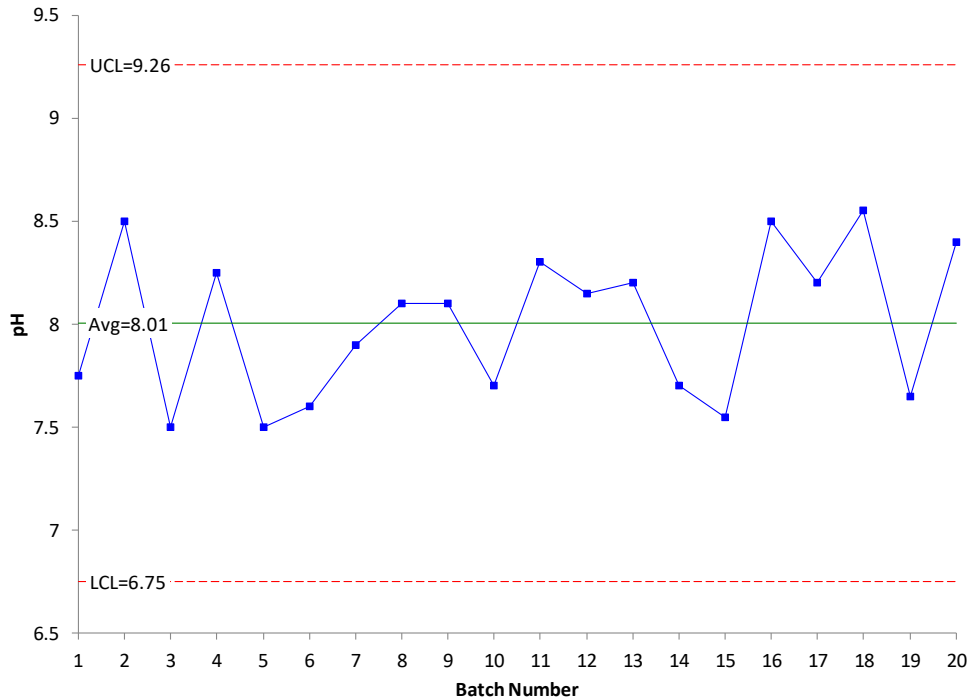
A few words of caution. The T² control chart, like other multivariate control charts, plots a value on the chart that you really can't explain too well. Suppose you have two variables that are important in an adhesive process. The adhesive is characterized by two variables, pH and viscosity, which need to be controlled. Data for 20 batches are shown below for pH and viscosity. The data in this example is adapted from "Advanced Topics in Statistical Process Control" by Donald Wheeler (www.spcpress.com).

Table 1: pH and Viscosity Data

Batch	pH	Viscosity	Batch	pH	Viscosity
1	7.75	5.48	11	8.30	5.58
2	8.50	5.98	12	8.15	5.44
3	7.50	4.12	13	8.20	3.11
4	8.25	5.34	14	7.70	4.34
5	7.50	4.36	15	7.55	4.08
6	7.60	4.26	16	8.50	5.96
7	7.90	4.50	17	8.20	5.80
8	8.10	5.16	18	8.55	5.94
9	8.10	5.56	19	7.65	4.00
10	7.70	4.08	20	8.40	5.86

One way of monitoring the pH and the viscosity is by using a control chart for each variable. An individuals chart (X-mR) could be used for both pH and the viscosity. Figure 1 is the X chart for pH and Figure 2 is the X chart for viscosity. The moving range charts are not shown here.

Figure 1: X Chart for pH



Figures 1 and 2 are both in statistical control. Look at the figures again. Notice anything? The two figures have very similar patterns. This implies that the two variables are correlated. A scatter diagram for the two variables is shown in Figure 3.

The scatter diagram shows that the two variables are related. As pH increases, the viscosity tends to increase. Note the point at the bottom of the scatter diagram for a pH of 8.2. This data pair looks like it might be an outlier, despite the X charts being in control. This is where the T^2 control chart can be used. It is designed to see the impact of multiple variables at the same time.

Figure 4 is the T^2 control chart for the data in Table 1.

The T^2 control chart shows an out of control point for batch 13. That out of control point does correspond to the low point on the scatter diagram above.

Look at the T^2 control chart in Figure 4. One of the first things you may notice about this control chart is that the value of T^2 has no resemblance to the original pH and viscosity data. So, looking at the value of T^2 really tells you nothing about the original data.

The only test for out of control points on this type of chart is points beyond the upper control limit (UCL). There is no lower control limit (LCL). There is one point beyond the control limit - but you don't know which variable (pH or viscosity) caused the out of control point by looking at the control chart. Information on how to determine which variable(s) is responsible is given below.

Figure 2: X Chart for Viscosity

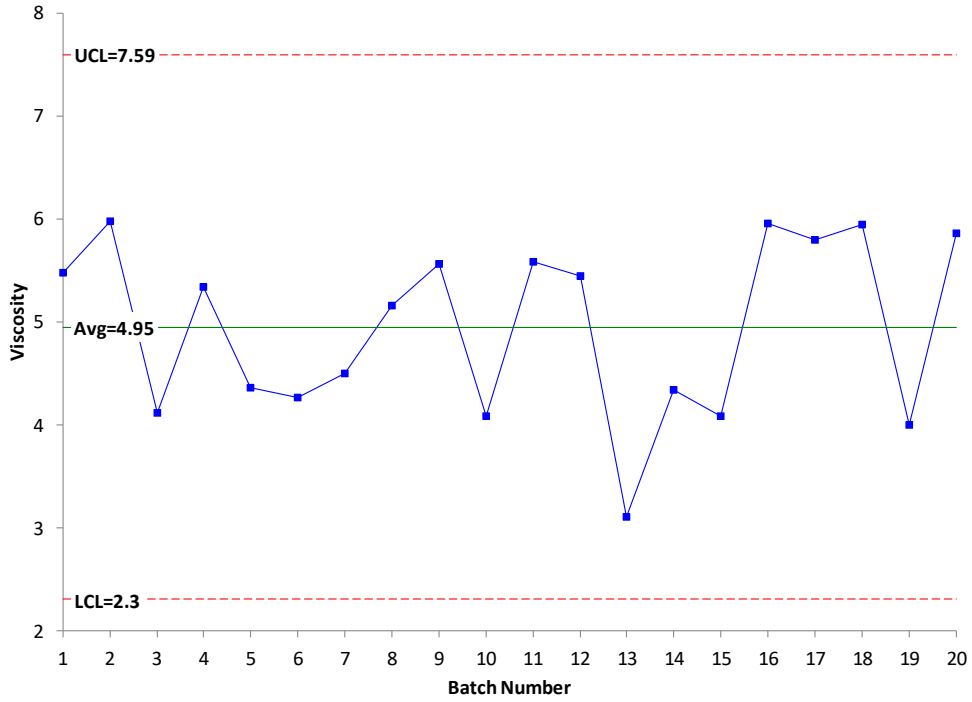


Figure 3: Scatter Diagram of pH vs Viscosity

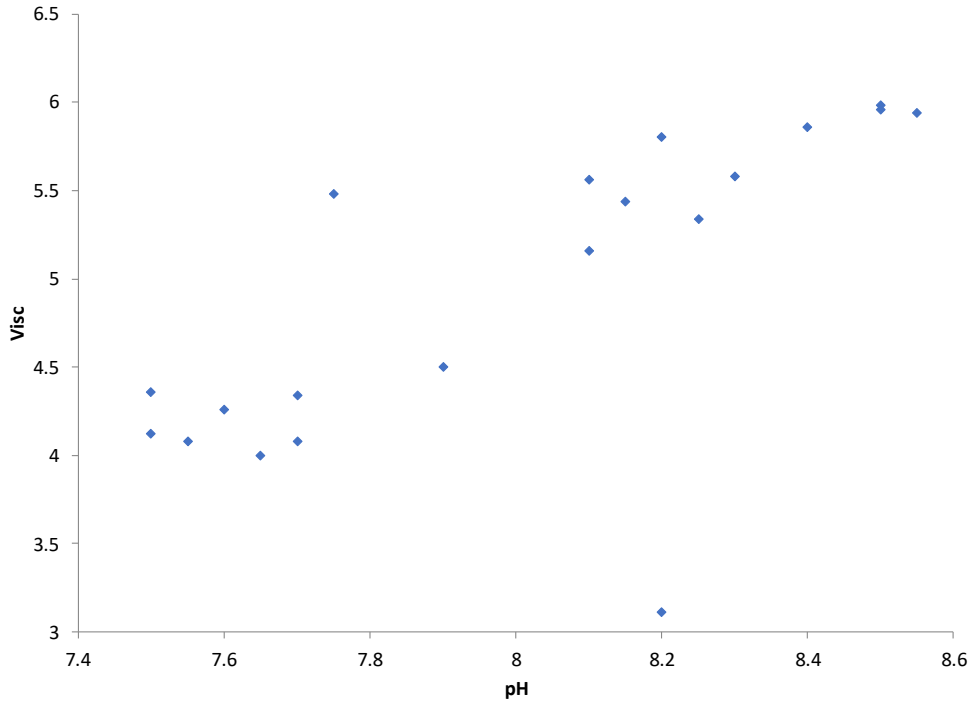
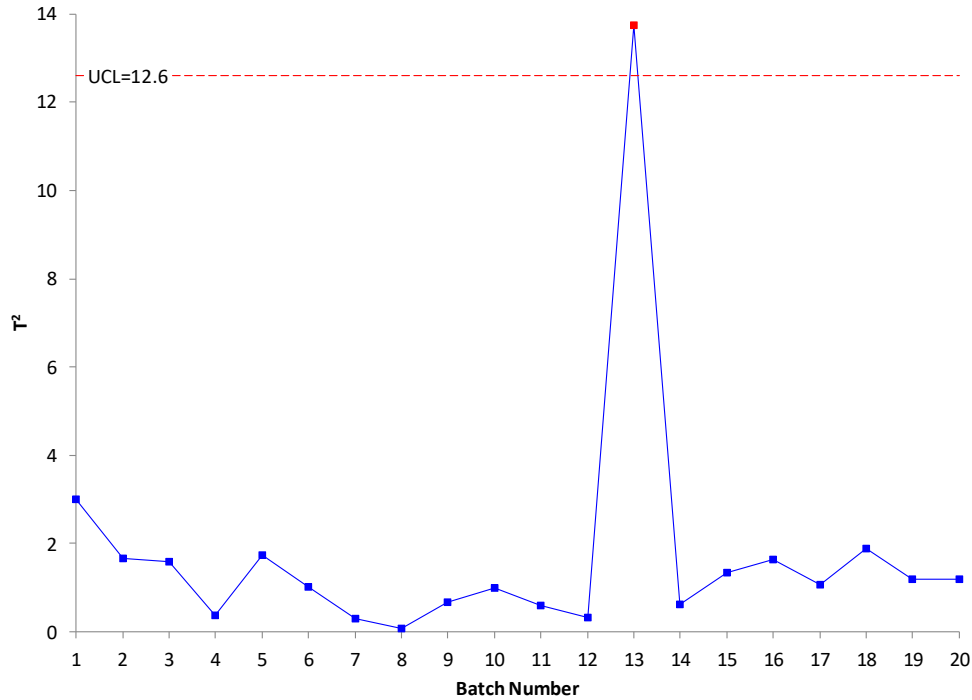


Figure 4: T² Control Chart for pH and Viscosity



Constructing a T² Control Chart

As stated before, the T² control chart can be used with data in subgroups or data that are individual observations. We will use individual observations to show how to construct a T² control chart. Note that the calculations are different for data in subgroups.

The first step in creating a T² control chart is to calculate the values of T². This is where matrix algebra comes in. The value of T² is given by:

$$T^2 = (\mathbf{x} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}})$$

where $\bar{\mathbf{x}}$ is the sample mean vector and \mathbf{S} is the sample covariance vector. This is clearly now a little more complicated than the calculations for the basic control charts, but we will give you the general idea of how it is done for the case of two variables.

We will start with the $(\mathbf{x} - \bar{\mathbf{x}})$ term in the T² equation. To create this term, you subtract the average value for the variable from each individual value for the variable. The average for pH and viscosity are given below.

$$\bar{x} \text{ for pH} = 8.005$$

$$\bar{x} \text{ for viscosity} = 4.9475$$

From Table 1, the first point for pH and viscosity are 7.75 and 5.48 respectively. Then, for the first point,

$$(\mathbf{x} - \bar{\mathbf{x}}) = [7.75 - 8.005 \quad 5.48 - 4.9475]$$

This will be eventually calculated for each point. Note that it is a matrix with one row for each point and a column for each variable.

Now, $(x - \bar{x})'$ is the transpose of $(x - \bar{x})$. The transpose of $(x - \bar{x})$ is:

$$(x - \bar{x})' = \begin{bmatrix} 7.75 - 8.005 \\ 5.48 - 4.9475 \end{bmatrix}$$

The transposed matrix has a column for each point and a row for each variable.

The **S** matrix is a little more difficult to find. It is found from the vector of moving differences for each variable. For each variable, v_i is found where

$$v_i = x_{i+1} - x_i$$

So, for pH, v_1 is the difference between the second pH value and the first pH value:

$$v_1 = 8.50 - 7.75 = 0.75$$

This is done for both variables and the vector **V** contains the results for both variables:

$$\mathbf{V} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_{m-1} \end{bmatrix}$$

where m = number of samples. The components of **V** are given below for the two variables.

pH	Viscosity
0.75	0.5
-1	-1.86
0.75	1.22
-0.75	-0.98
0.1	-0.1
0.3	0.24
0.2	0.66
0	0.4
-0.4	-1.48
0.6	1.5
-0.15	-0.14
0.05	-2.33
-0.5	1.23
-0.15	-0.26
0.95	1.88
-0.3	-0.16
0.35	0.14
-0.9	-1.94
0.75	1.86

So **V** has a column for each variable and $m - 1$ rows. **S** can now be found using the following:

$$S = \frac{1}{2} \frac{V'V}{(m-1)}$$

where V' is the transpose of V . You can easily multiply two matrices in Excel using the function MMULT. If you multiply V' by V , you get the following:

$$V'V = \begin{bmatrix} 6.1325 & 9.9235 \\ 9.9235 & 29.0938 \end{bmatrix}$$

S is then found by dividing each term in $V'V$ by $2(m-1) = 2(19) = 38$.

$$S = \begin{bmatrix} 0.161382 & 0.261145 \\ 0.261145 & 0.765626 \end{bmatrix}$$

Remember that T^2 is given by:

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

S^{-1} is the inverse of S . You can use the Excel function MINVERSE to find the inverse. Using that function gives:

$$S^{-1} = \begin{bmatrix} 13.8296 & -4.71709 \\ -4.71709 & 2.915055 \end{bmatrix}$$

Now we have all three terms: $(x - \bar{x})'$, S^{-1} , and $(x - \bar{x})$.

So, for the first point:

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

$$T^2 = \begin{bmatrix} 7.75 - 8.005 \\ 5.48 - 4.9475 \end{bmatrix} \begin{bmatrix} 13.8296 & -4.71709 \\ -4.71709 & 2.915055 \end{bmatrix} \begin{bmatrix} 7.75 - 8.005 & 5.48 - 4.9475 \end{bmatrix}$$

You can use the Excel MMULT function to multiply these three matrices together. The result is given below.

$$T^2 = 3.006896$$

This is the value of T^2 for the first data point. The above process can be used to generate the T^2 values for the rest of the data. The values of T^2 are shown in the table below.

These are the values that are plotted in the T^2 control chart in Figure 4. The other calculation that is needed is for the UCL. And, of course, this is a little more complicated here as well. The upper control limit based on individual observations is given by the following:

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

where m = number of samples, p = number of variables, β = Beta distribution, α = the confidence level and

$$q = \frac{2(m-1)^2}{3m-4}$$

Table 2: Values of T2 for pH and Viscosity

Batch	pH	Viscosity	T2	Batch	pH	Viscosity	T2
1	7.75	5.48	3.006896	11	8.30	5.58	0.609407
2	8.50	5.98	1.674520	12	8.15	5.44	0.324114
3	7.50	4.12	1.580571	13	8.20	3.11	13.748666
4	8.25	5.34	0.371990	14	7.70	4.34	0.614283
5	7.50	4.36	1.734041	15	7.55	4.08	1.333028
6	7.60	4.26	1.019390	16	8.50	5.96	1.648693
7	7.90	4.50	0.292941	17	8.20	5.80	1.076091
8	8.10	5.16	0.065993	18	8.55	5.94	1.876166
9	8.10	5.56	0.669462	19	7.65	4.00	1.186581
10	7.70	4.08	0.984076	20	8.40	5.86	1.184571

You can use the BETAINV function in Excel to determine the value of the beta distribution. Note that the control limits do not depend on the values of T^2 . Using $\alpha = 0.0027$ gives an UCL of 12.6 as shown in Figure 4. (Note: if you use the BETAINV function, α is $1 - 0.0027$.) There is no lower control limit.

Out of Control Points

When there is an out of control point on a T^2 control chart, more work is needed to determine which variable could be the cause of it. If there are points out of control, the results are decomposed to find out which variables could be responsible for the out of control situation. Each variable is removed from the calculations and the T^2 calculations repeated. The difference between the value of T^2 with the variable present and without the variable present is calculated. The larger the difference, the more likely the variable is to be the cause of the out of control point.

For the out of control point in Figure 4, the difference due to pH is 9.34 and the difference due to viscosity is 13.51. Based on this analysis, it appears that viscosity is the variable responsible for the out of control point.

Summary

This publication has introduced the T^2 control chart. This control chart is used to monitor multiple variables on one chart. This publication demonstrated how to do the calculations for the case of two variables. This included the calculation of T^2 as well as the upper control limit. It was shown how to determine which variable might be responsible for an out of control point. Computer software, like SPC for Excel, can easily handle the calculations presented in this publication.

Quick Links

[Visit our home page](#)

[SPC for Excel Software](#)

[SPC Training](#)

[SPC Consulting](#)

[SPC Knowledge Base](#)

[Ordering Information](#)

Thanks, so much for reading our publication. We hope you find it informative and useful. Happy charting and may the data always support your position.

Sincerely,

Dr. Bill McNeese
BPI Consulting, LLC