**Multivariate Process Monitoring Procedures**

### for Jacob Hartzer

Assigned on: 4.11.2017 Submit on: 4.28.2017

There are many industrial settings where process performance is based on the behavior of set of variables that are interrelated. This is common in production units. An example of a production unit is one where natural gas as a form of energy is converted to steam or electricity. In this project we will develop control procedure that will detect unusual occurrence in variables. We need to perform a multivariate analysis because the characteristics of the variables are interrelated and hence they need to be examined together.

In multivariate analysis, we need to examine the variables relative to the relationships that exist among them. If the mean point is typical of sample data, one form of analysis is finding the distance of each point from the mean. The distance can be obtained using Euclidean distance.

The problem of using Euclidean distance is that it does not take care of the variance-covariance among the data. This is overcome by another statistical distance known as Hotelling’s T2 given by:

where is the mean and is the variance-covariance matrix of the sample.

One of the problems of the implementation of Hotelling’s T2 is the evaluation of the inverse of . The has to be non singular so that it can be inverted. A singular occurs when two or more variables are perfectly correlated (collinear). Collinearities can occur in a covariance matrix because of theoretical relation exist in the process or because of outliers in data. A convenient way to identify the collinearity is through examining the eigenvalues and eigenvectors of

We use principal component analysis (PCA) for examining collinearities of . PCA can help to identify subgroups of variables that are highly correlated and can be used to estimate the dimensionality of the system. The inverse of the can be expressed as:

where are the eigenvalues of and are the corresponding eigenvectors. If is close to zero, the inverse becomes large and will have a disproportionate effect on the calculation of and this will distort T2 calculation.

A recommended guideline for finding near singular is to estimate the square root of the ratio of maximum eigenvalue to each of eigenvalues. This is called condition indices. Condition index > 30 imply severe collinearity among variables. In that case if you have to implement T2 statistics, you have to remove the variable involved in collinearity. To determine which variables are involved in severe collinearity,

1. Select the eigenvector corresponding to the eigenvalue that has condition index > 30
2. Ignoring variables with small loadings, find the variables that have almost equal loadings. (You can also verify the large correlation between the two variables from their value in covariance matrix.)

One of the variables that have equal loadings is redundant and can be removed without loss of any information. Thus T2 statistics can be calculated by excluding the redundant variable. You can also directly estimate by removing the eigenvector corresponding to the near zero eigenvalue.

**Task**

For the given time variant data, estimate the Euclidian distance from the mean at different time steps. Plot Euclidian distance (in the ordinate) at different time (in the abscissa). Check if you can calculate T2 with the existing data, if not identify the redundant variable(s) responsible for near singularity. Estimate the T2 at different time steps after removing the redundant variable(s) for monitoring the system for outliers by plotting T2 with time. Give your analysis.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Time | X1 | X2 | X3 | X4 | X5 | X6 | X7 |
| 1 | 2020 | 165 | 661 | 5.9 | 1.6 | 82.5 | 86 |
| 2 | 2020 | 255 | 675 | 7 | 5 | 127.5 | 84 |
| 3 | 2014 | 266 | 675 | 7 | 3.9 | 133 | 85 |
| 4 | 1960 | 270 | 900 | 3 | 8 | 135 | 89 |
| 5 | 1870 | 185 | 850 | 4 | 4.8 | 92.5 | 90 |
| 6 | 1800 | 195 | 590 | 3 | 4 | 97.5 | 87 |
| 7 | 1711 | 201 | 663 | 4 | 3.6 | 100.5 | 88 |
| 8 | 1800 | 250 | 875 | 0 | 6 | 125 | 88 |
| 9 | 2011 | 182 | 710 | 2.4 | 7.7 | 91 | 86 |
| 10 | 1875 | 175 | 600 | 4.3 | 6 | 87.5 | 97 |