ISEN 614 Advanced Quality Control
FINAL PROJECT
Phase 1 Analysis

Control Charts for multivariate data for manufacturing process

Industrial and Systems Engineering Department, Texas A&M University, College Station

TEAM MEMBERS

1 George Paul 627003548

2 Anwaar Hussain Muhammad 827008798

3 Prakash Jeremiah 627005933

4 Mehta Yash 228002156

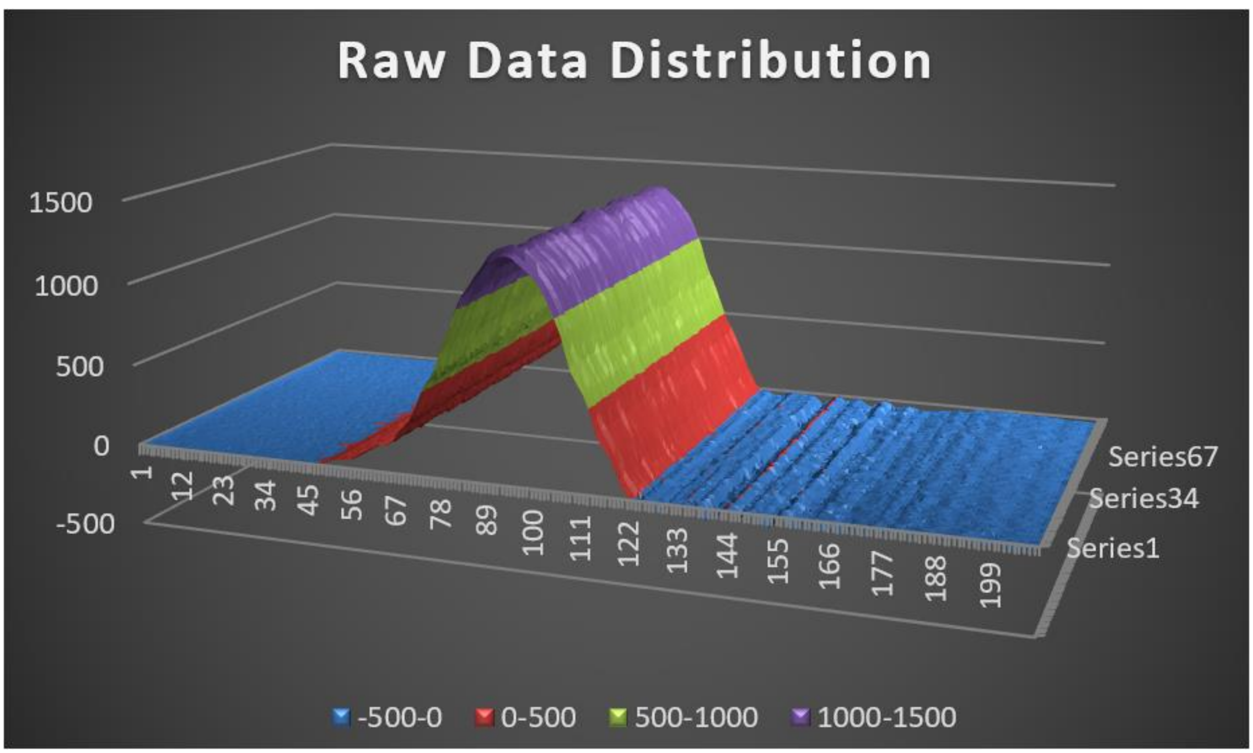
Contents

Introduction

In a manufacturing setting it is important to deliver the products with high quality. In this project we work on real life data generated from a system/ machine. There will be variations in the product dimensions due to inconsistencies from man and machine, our aim is to set in a system in order to detect these variations. We used PCA, univariate and Hoteling T^2 charts in order to set an in-control condition for our data as part of the Phase 1 analysis.

Approach

The raw data is initially interpreted in excel. We found that data is normally distributed for all 209 predictors. The correlation matrix that was generated suggests that the variables are highly correlated. The given high dimensional data needs to be reduced to a form where we can perform and set control limits for our future data. Various methods will be tried and tested such as the univariate charts, Hoteling T^2 limits etc. for different alpha values in order to adjust the ARL0 values which will in turn give us the best results in our phase 1 analysis.



Principal Component Analysis

It is difficult to analyze or perform any statistical procedure on a data set with 209 variables and 552 observation for each series. Hence, we adopt principal component analysis as the dimension reduction technique, in order to reduce the number of variables that is to be considered for our analysis. By PCA our aim is to obtain the variables that gives us the vital information required for

our analysis. PCA can be performed on both the covariance matrix and the correlation matrix. But since there is enough evidence from the initial analysis of raw data that there is significant correlation between successive variables, we can move forward in our principal component analysis considering the covariance matrix as the difference between the correlation matrix and covariance analysis is normalization of its elements.
Once the Principal component analysis was performed, we used a combination of scree plots and pareto charts to identify the PC's that gave us most information by considering the elements that considered the most variations. These components or variables were considered for further analysis.

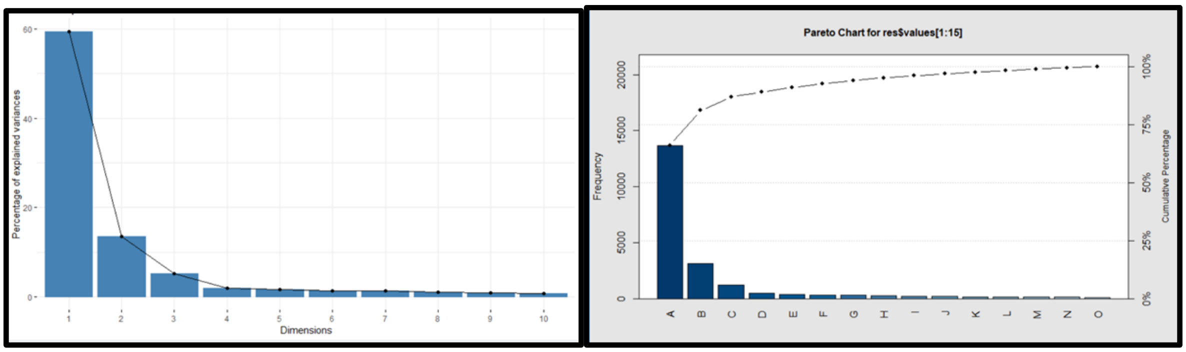


Figure 1: Scree Plot and pareto charts of the principal components. We consider 4 PC's to explain 80% of variation.

Analysis

We use a scree plot in order to determine the number of principal components needed to explain the variance of the data. We look for the elbow in the graph and using that as a reference we select the number of principal components.
From the Figure 1 we can see that the elbow bends at around the third or fourth principal component. Thus, we select four principal components to explain the variance in the data.

Univariate Control Chart Approach:

Since the data is uncorrelated, we can use multiple univariate control charts to detect the out of control points. Multiple composite decision rules could be used to determine whether a point can be considered OOC such as if only all charts signals or either of the charts signal thus creating ambiguity in the process. Also, depending on the decision rule used, either the alpha error or the beta error could be seriously inflated. We therefore decided to use the Hoteling T^2 charts instead so there was no confusion on the decision rules as there is only one way we can decide if a point is OOC.

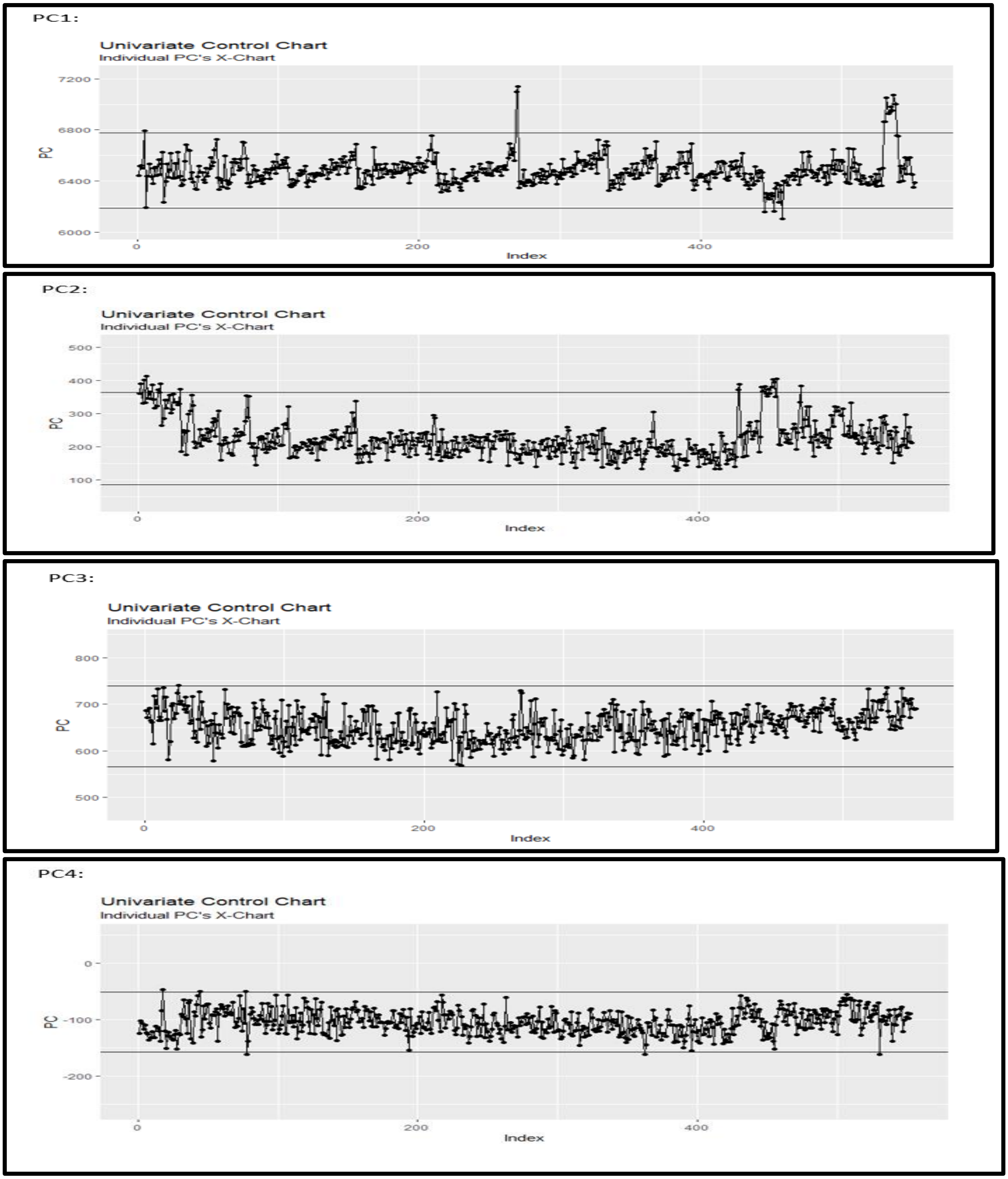


Figure 2: Univariate approach results for each principal component analysis. Here the results of each PC are set to a separate control limit.

Hoteling T Square

Since this is a Phase-I analysis with n=1, we used chi-squared distribution with alpha set to 0.05 to set the UCL at 9.44. We used this UCL to determine the OOC points and eliminate them from the dataset. Once the OOC points are removed, the T^2 statistic is re-calculated for all the remaining data points as the mean and covariance changes. This process is repeated until no points fall outside the control limit. Figure 3 is the plot obtained for the raw data, we can clearly see that a lot of points fall outside the UCL and hence we remove these points and perform multiple iterations until we obtain a chart where there are no points that lie above the UCL.

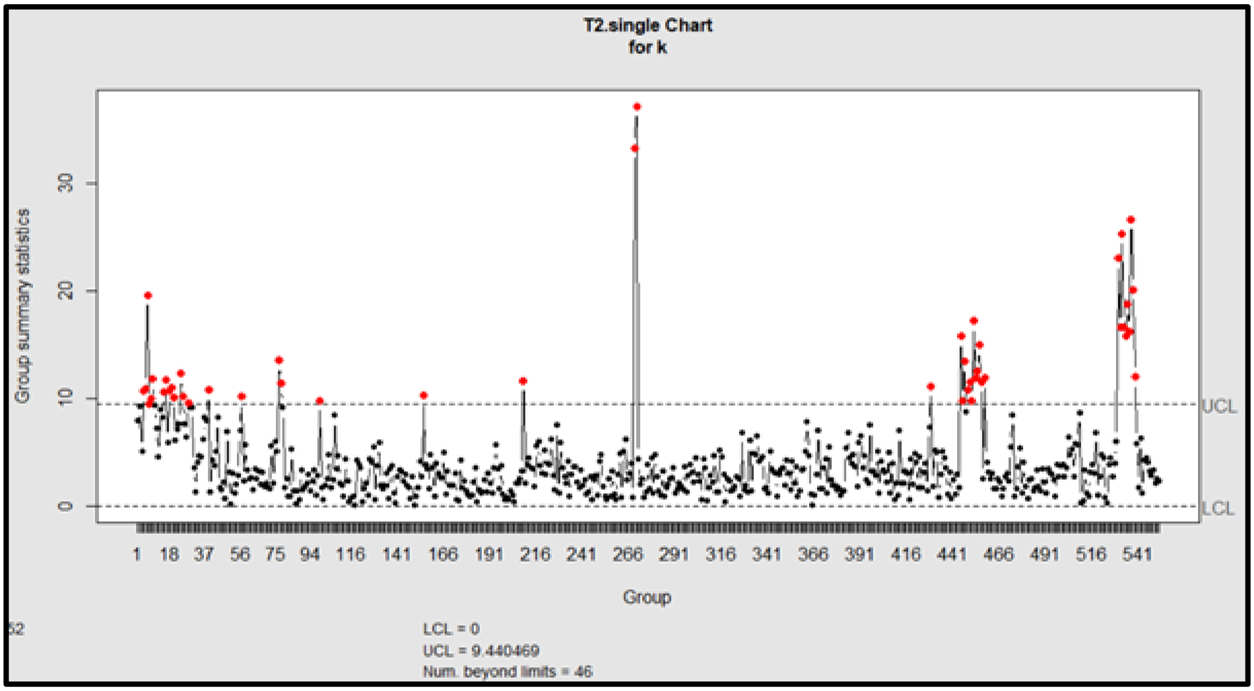


FIGURE 3: T2 plot obtained for the dataset using four PC's at UCL set at 9.44.

Figure 4 is the plot obtained after iterating the process 10 times. Here we can see that all the data points fall below the UCL and hence we can state that all OOC points have been removed. Thus, this data can now be used as the training data for alpha value 0.05 and we can conclude with the PHASE I analysis. It was observed that out of the 552 data points, 141 data points were OOC and hence were removed from the training dataset. Thus, using the results obtained we can now perform the PHASE II analysis if necessary.

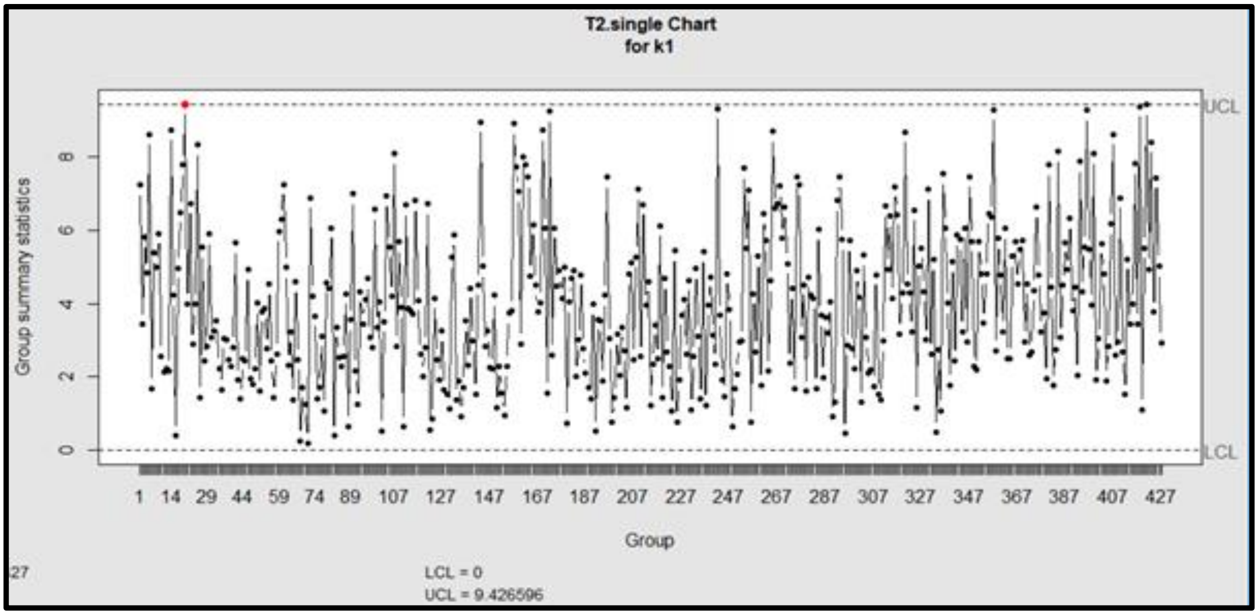


FIGURE 4: T2 plot obtained for the dataset using four PC's at UCL set at 9.42. Here no OOC points are seen.

Parameters for future monitoring:

After finding the in control mean and covariance matrix, we simulated data using these parameters to find out what the ARL0 and the ARL1 would be for our data. For finding the ARL0, we used the in-control parameters to simulate the data points and find the RL, this was repeated to get enough data, to get the ARL0. Similarly, to detect the ARL1 we increased the mean by 5% to assess the detection capability in the event of a small mean shift. The results obtained are as follows:

For alpha = 0.05 and UCL = 6.06

ARL0 = 215, ARL1 = 4

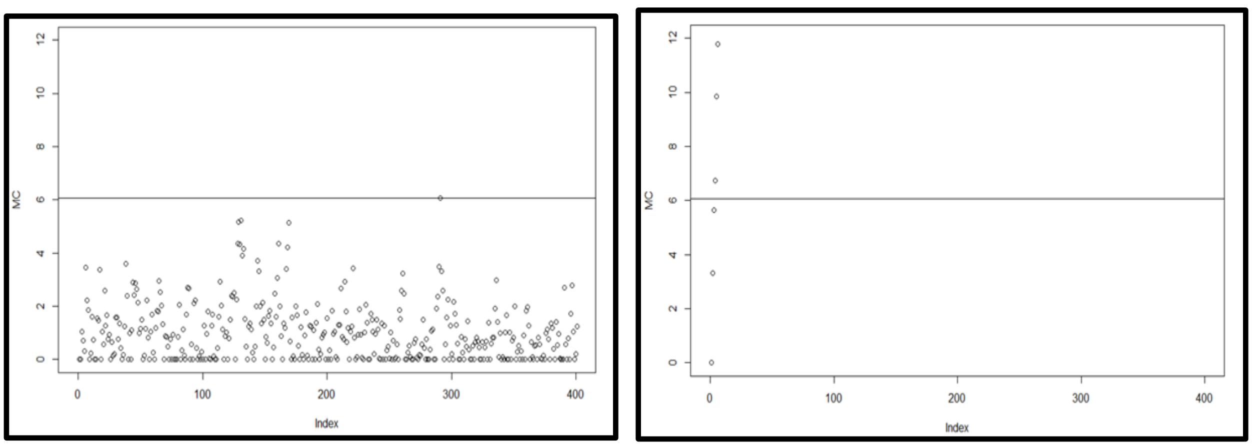


Figure 5: CUSUM charts. Simulated data with IC parameters (left). Simulated data with 5% mean shift (right)

Conclusion

In this paper we have learned how to apply and set control chart limits to real life manufacturing data. Dimension reduction is a handy method to initiate the multivariable problem, as it helps to retain all the vital information without actually changing the actual meaning of the process. Tried different iterations and methods for the data to achieve the best ARL0 value for phase one analysis. We realized that hoteling T squared chart will help us set the training data by removing the out of control points for phase 1 analysis. While univariate control charts will help to set controls it's difficult to compare and monitor multiple graphs at the same time when there are a greater number of variables. There is inflation in ARL values which makes it difficult to detect. We have also done a MCUSUM analysis to ensure the results obtained are comparable and acceptable.