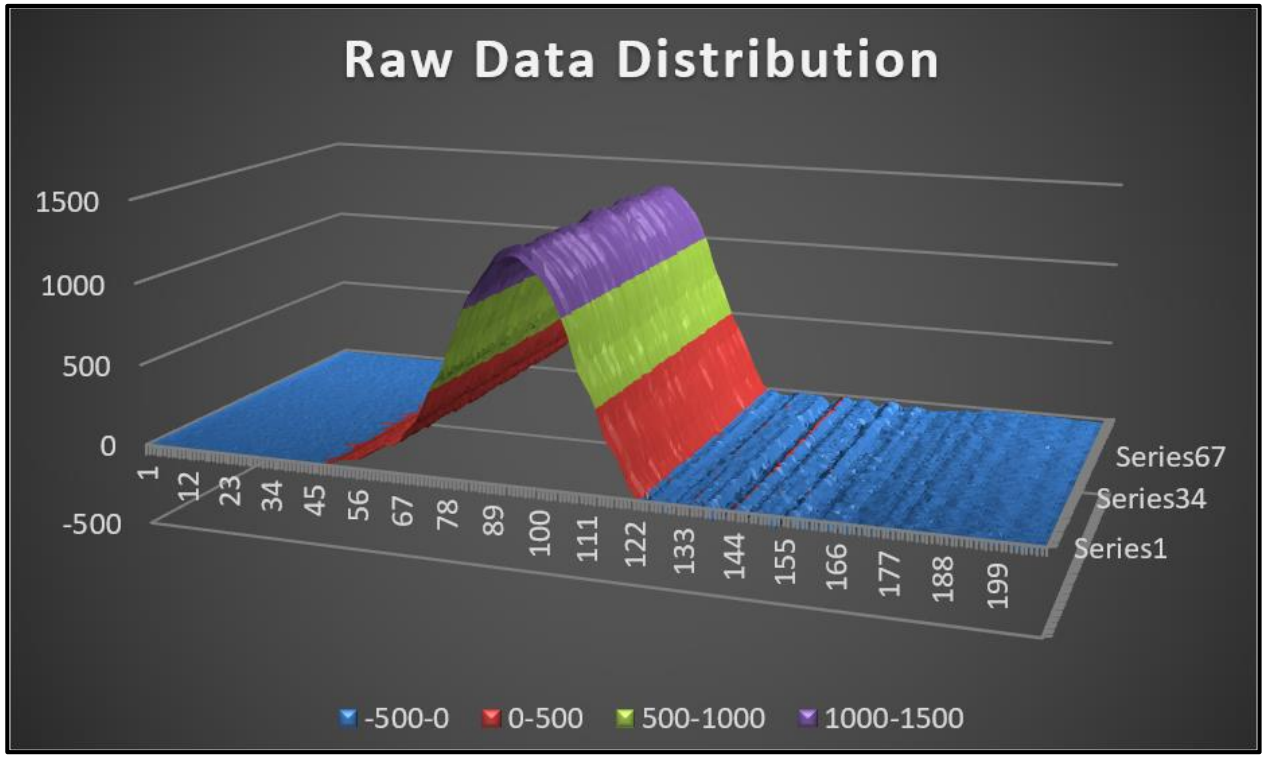
INDUSTRIAL & SYSTEMS ENGINEERING

TEXAS A&M UNIVERSITY

Phase I Analysis on Manufacturing process dataset

By:
Muhammad Hussain Anwar - 827008798
Paul George - 627003548
Jeremiah Prakash - 627005933
Yash Mehta - 228002156

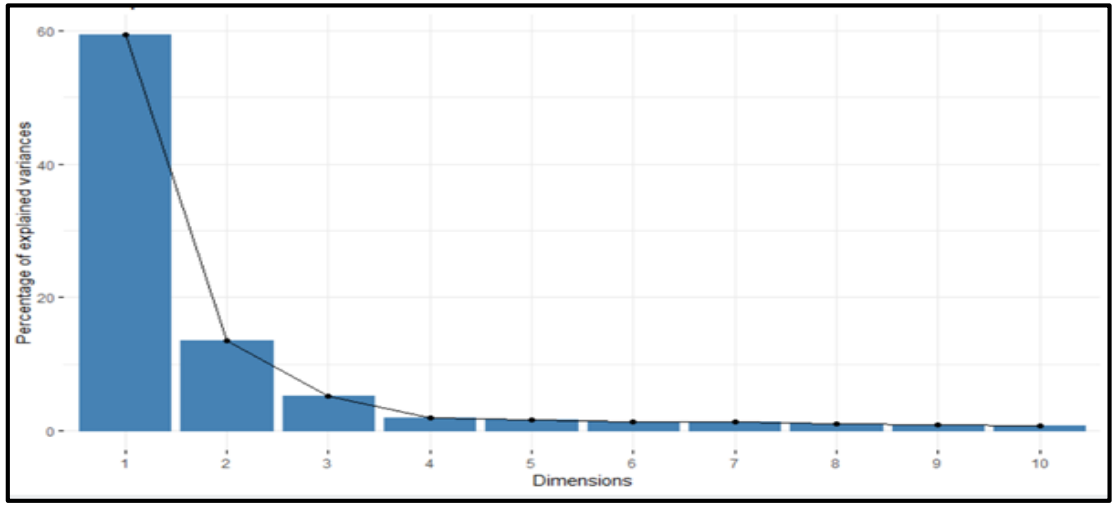
Preliminary Analysis:

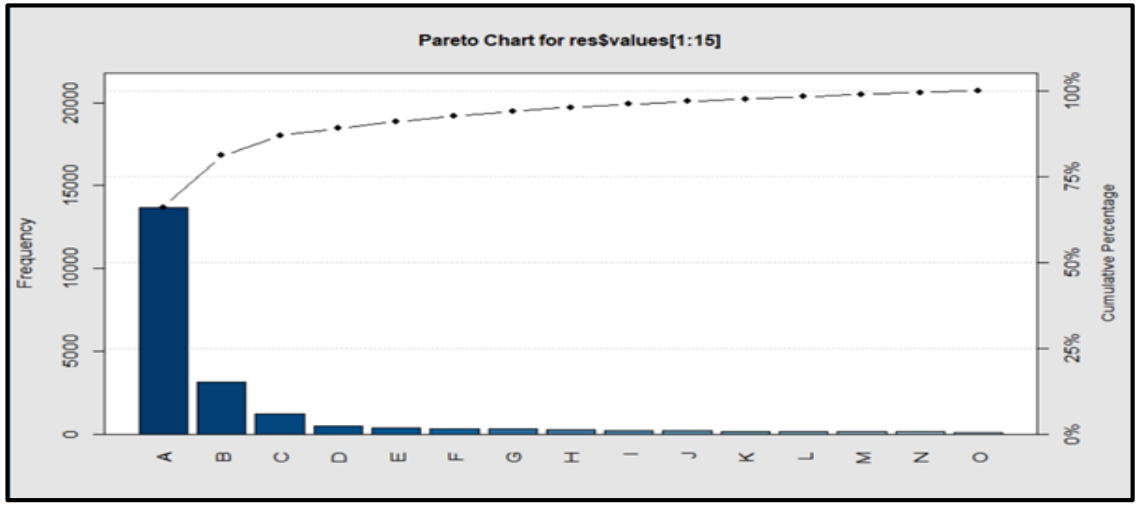


- Dataset contains of 552 observations for 209 attributes.
- The distribution for each attribute is normal distribution.
- By checking the co-variance and correlation matrix it was found that the attributes are highly correlated

Principal Component Analysis

- Since it is a high dimensional data and the attributes are highly correlated, the inherent noise in the data will drown the signals and make it harder to reject the null hypothesis.
- Use the concept of effect sparsity which states that it is always the "vital few" instead of the "trivial many" that matters.
- Used the PCA technique to reduce the dimension of the data.
- After finding the Principal Component Scores, scree plot and pareto plot were used to select the number of PCs to be used.



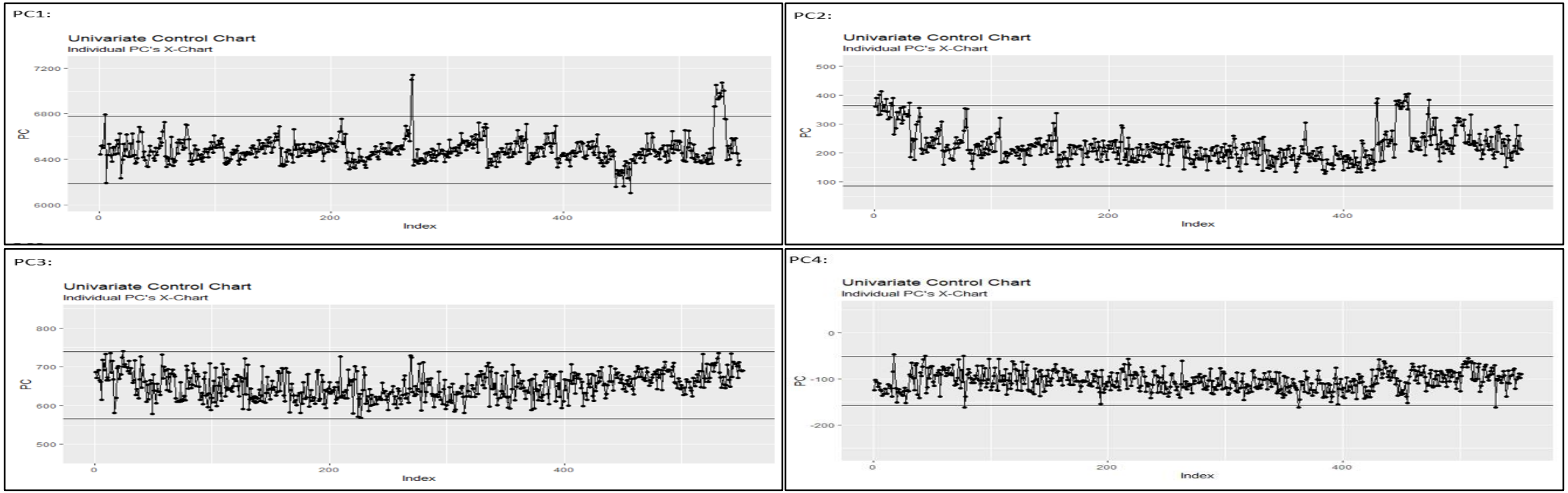


- Based on these plots, we decided to use first 4 PCs that explained a total of 80% variance in the data.

Control charts

INDUSTRIAL & SYSTEMS ENGINEERING
TEXAS A&M UNIVERSITY

- Univariate charts obtained using composite decision rules for obtaining a combined alpha level of 0.05

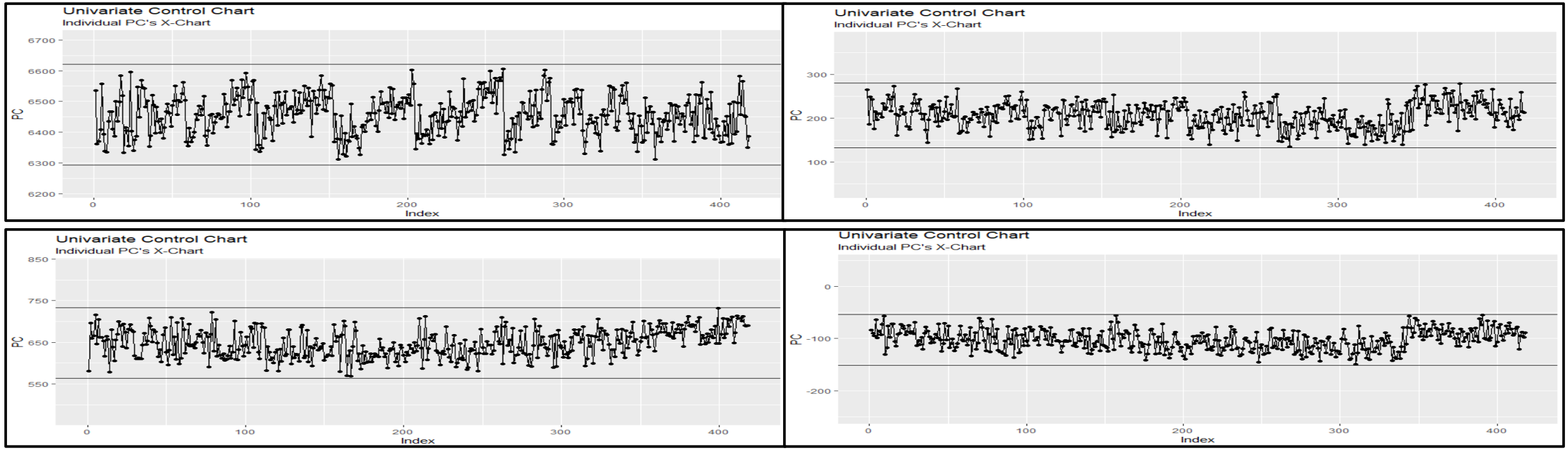


- Since, composite decision rules have to be used in case of multiple univariate charts, it leads to inflated α or β error unless parameters for individual charts are not tweaked.

- Therefore, we decided to use the Hotelling T^2 chart instead as it only has 1 decision rule and is unambiguous.

Control Charts

- The univariate charts obtained after iteratively removing OOC points



- It took 16 iterations to achieve the in-control data points.

- Since using univariate charts is computationally intensive and leads to ambiguity in interpretation, we decided to go ahead with Hotelling T^2 chart instead.

Hotelling T^2 chart

- Using the mean and covariance matrix from the PC scores, initial T^2 statistics were calculated.
- As the T^2 statistic in this case follows chi-square distribution, UCL was determined using chi-squared table at α level of 0.05
- The Out of Control(OOC) points were removed iteratively by recalculating the T^2 after each iteration using new mean and covariance matrix.

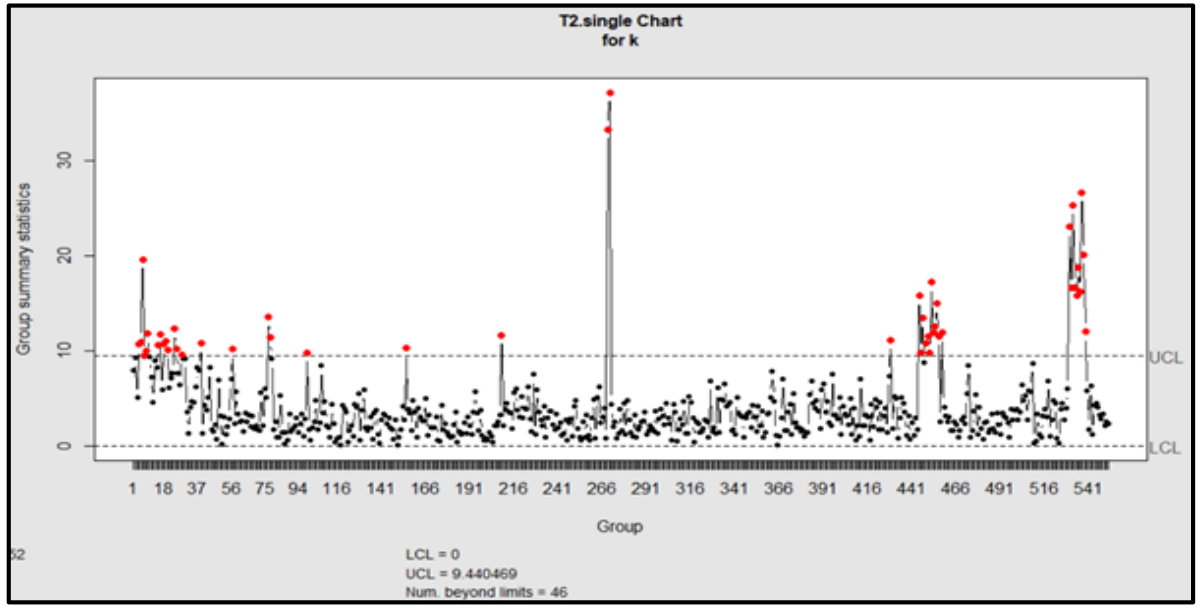


Fig: T^2 chart for original observations of PCs

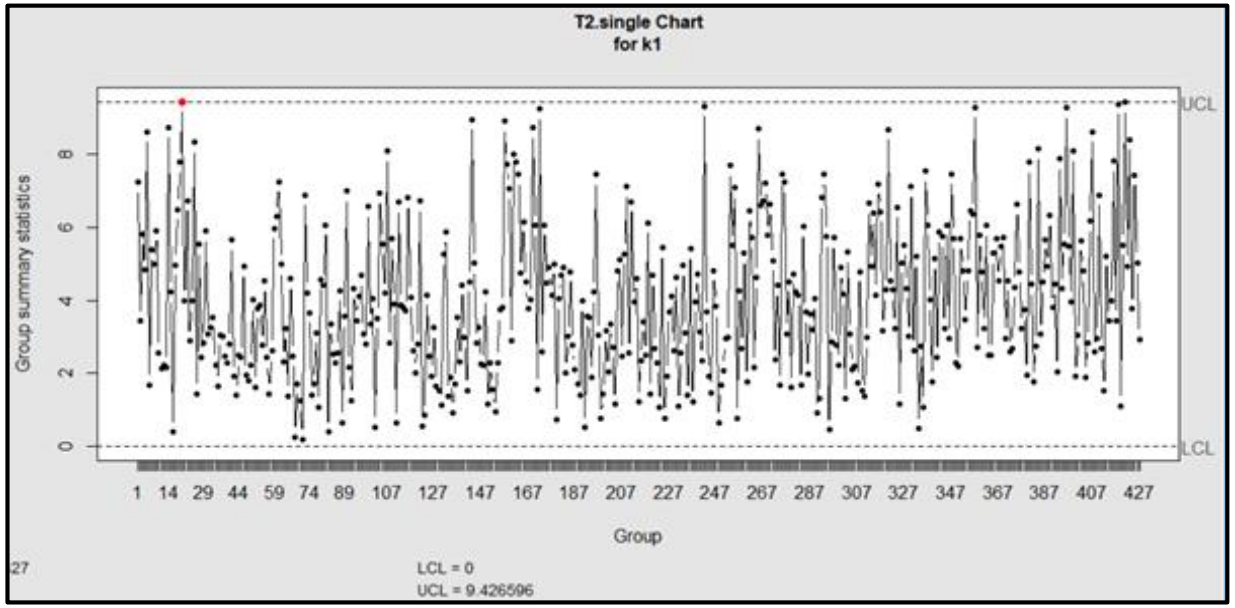


Fig: T^2 chart after removing all OOC points

- Using the IC data points, the IC mean and covariance matrix were calculated.

M-CUSUM chart

- To establish the parameters for future monitoring, M-CUSUM chart was used to set the control limits for detection of mean shift.

- The control limits were designed to obtain an ARL_0 of 200 and ARL_1 of 4 for a mean shift of 5%using Monte Carlo simulation technique.

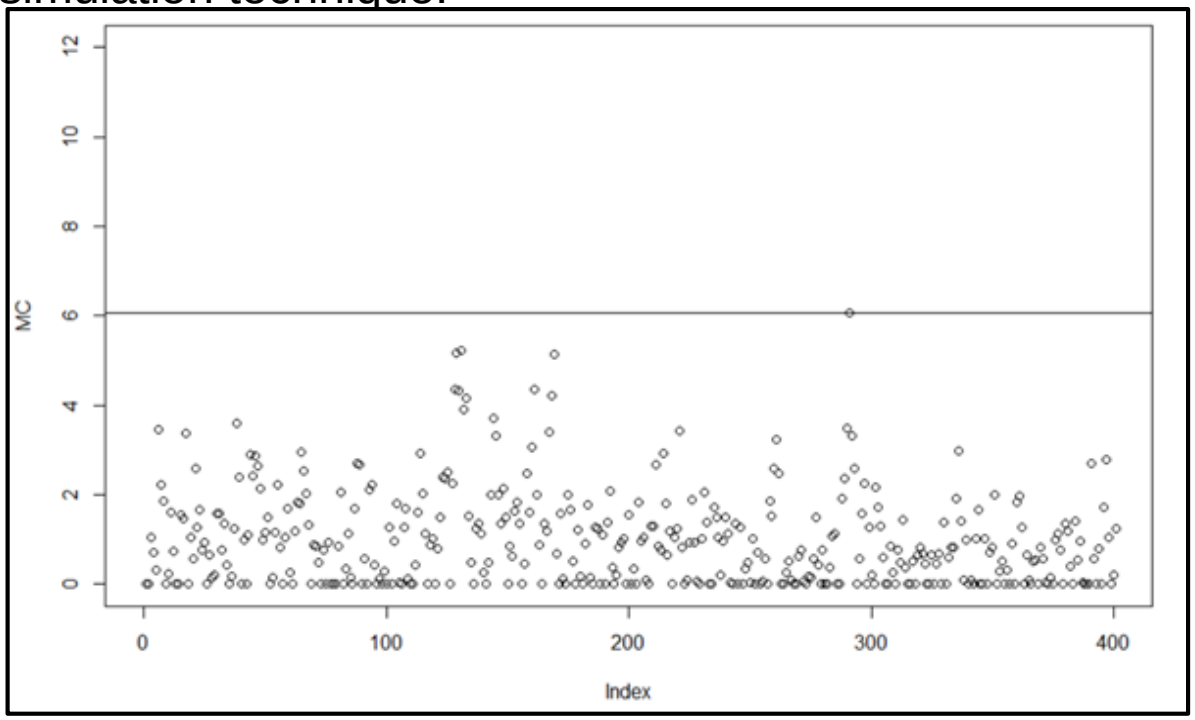


Fig: Data simulated using IC parameters

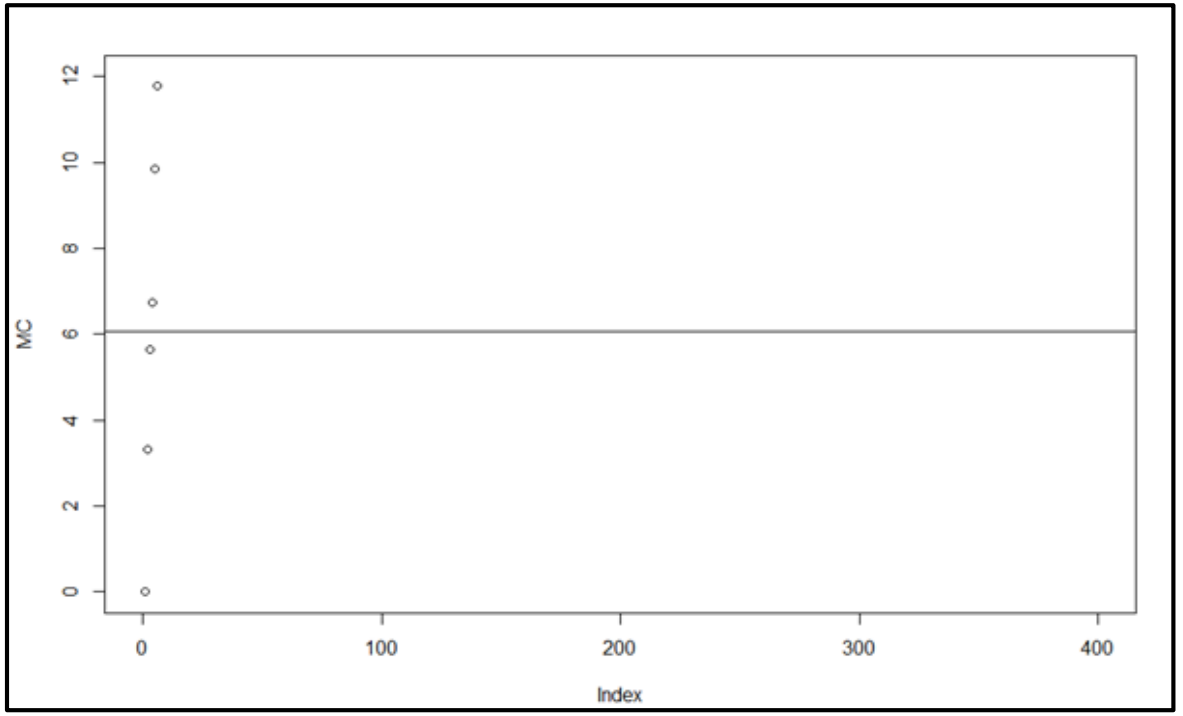


Fig: Data simulated using 5% mean shift parameters

Learnings and conclusions

- It is difficult to use high dimensional, highly correlated data for quality control purposes as it makes it harder to reject null hypothesis.
- Dimension reduction through PCA is a powerful tool to capture majority of the information in data in a fraction of its size.
- Even though multi–variate charts can be used on reduced data as it is un-correlated, it makes it difficult to determine appropriate control limits to achieve desired results.
- Hotelling T^2 chart is easy to use in case of multivariate data to detect OOC points in quality control process.
- M-CUSUM is useful in detecting a small mean shift in the process and should be used in conjunction with other charts for quality control process.

Thank You