PHASE 1 ANALYSIS TO IDENTIFY IN-CONTROL DATA & ESTIMATE DISTRIBUTION PARAMETERS

ISEN 614 ADVANCED QUALITY CONTROL

TEAM MEMBERS: JAY SHAH(826008661) & KISHAN SHAH(426009889)

ANALYSIS APPROACH

Preliminary Analysis

- Simulated data set is given.
- Data involves 25 quality characteristics & 1000 data records.

Principal Component Analysis

- Use of covariance matrix to convert given data into a linearly independent principal components.
- Pareto plot, Scree plot tools are used to reduce the data dimension by identifying principal components which contributes approx. 80% of total variance.

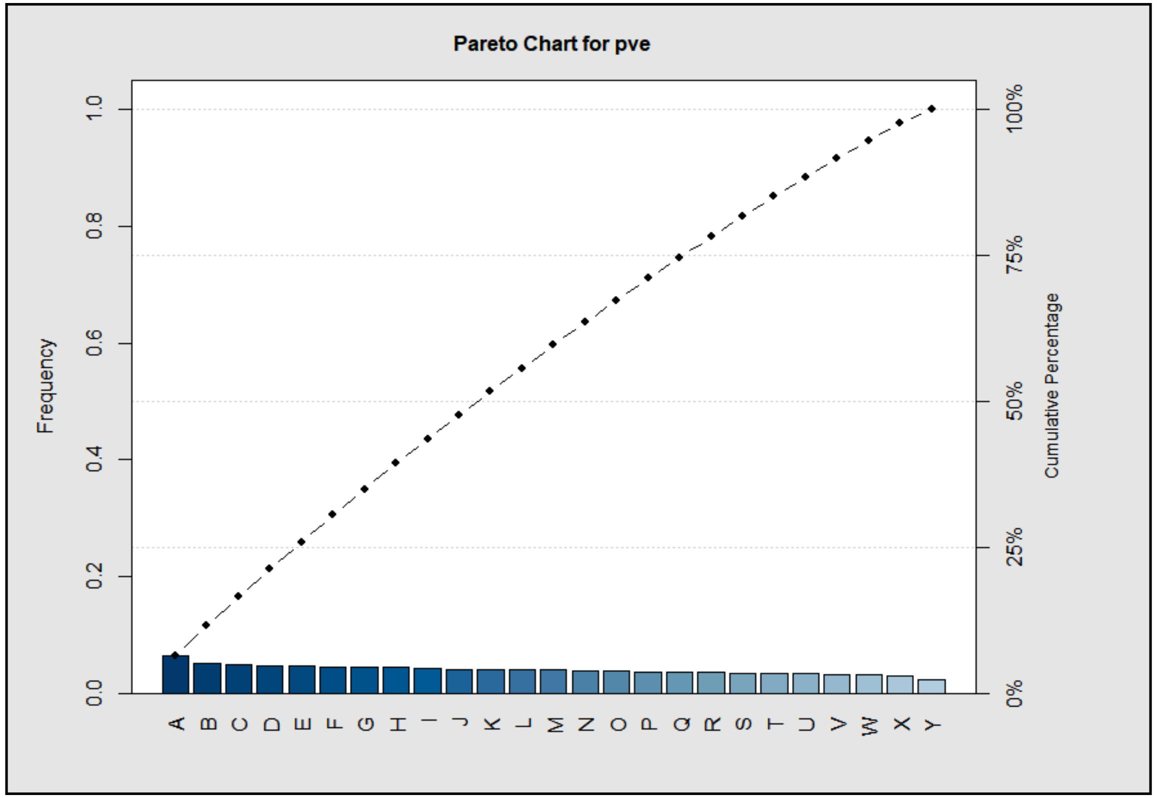
Control Charting

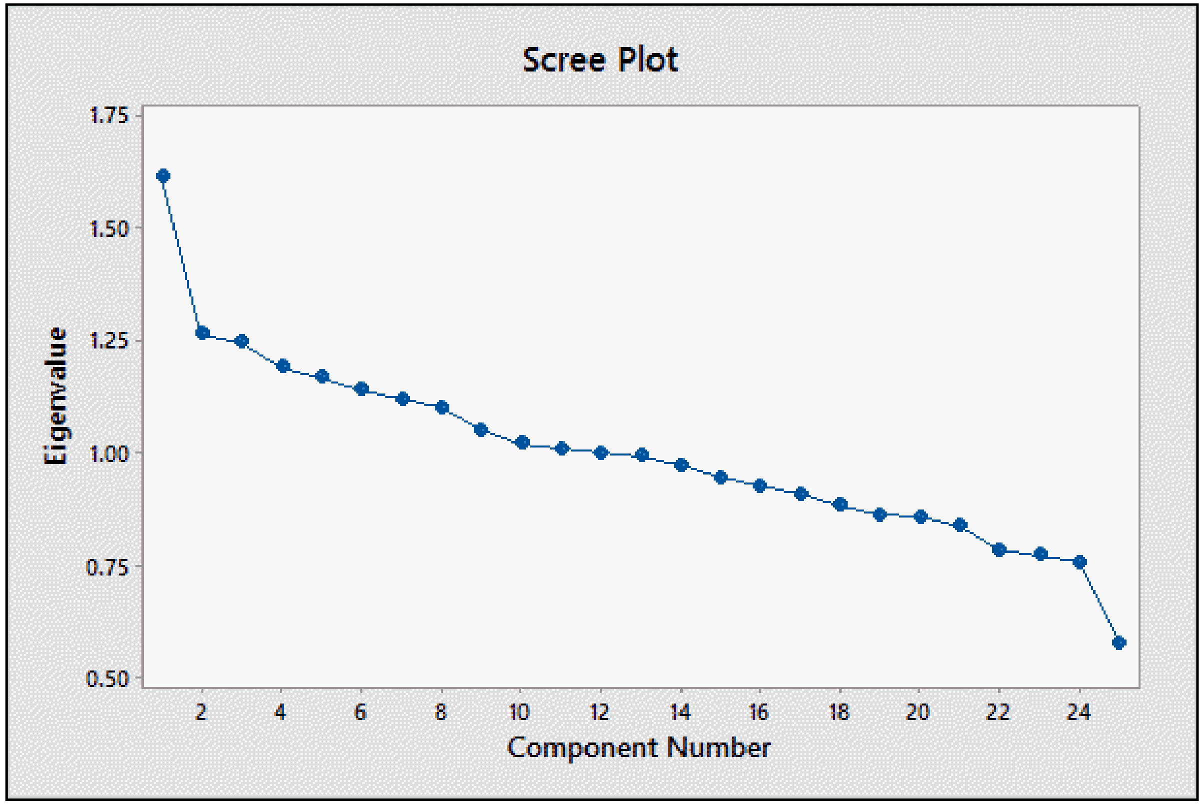
- Perform Hotelling T^2 to remove data points associated with spike type change.
- Perform m-CUSUM to remove data points associated with sustained mean shift type change.

Nature of change

- After removing all the outliers using T^2 and m-CUSUM, to determine where the change occur and type of change like mean shift change, variance change or covariance change.
- Estimate in control distribution parameters (μ_0 & Σ_0) which can be used for phase 2 analysis.

PRINCIPAL COMPONENT ANALYSIS



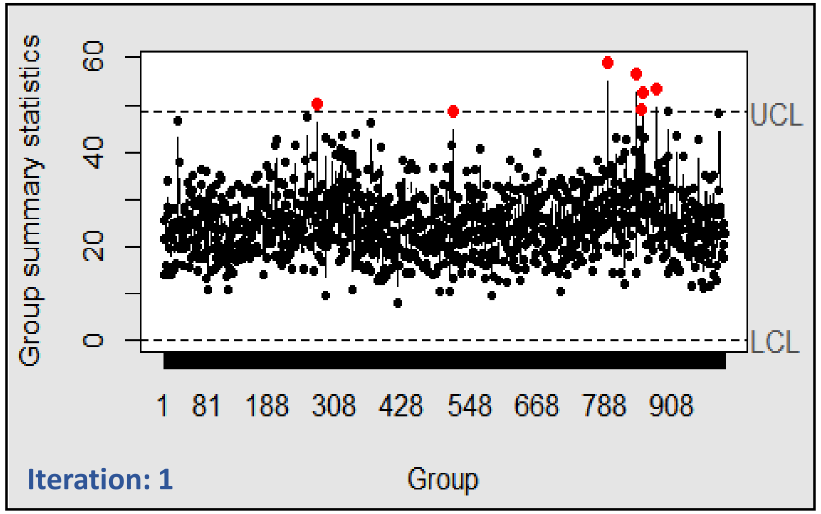


- The Principal Component Analysis is performed in R. After that we have plotted pareto chart and scree chart to identify vital few points.

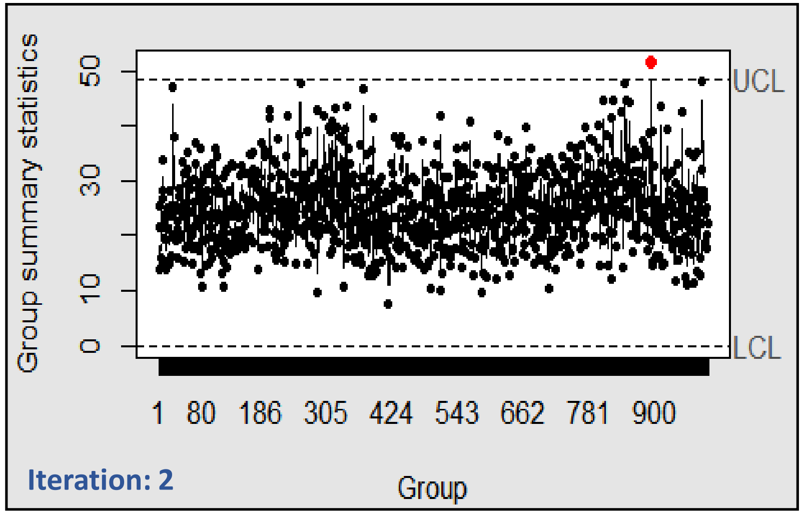
- In scree plot, elbow occurs at point 2 but in pareto chart percentage variance explained by point 2 is approximately 12%.

- So, It is meaningless to consider only 2 principal components for further analysis, as they do not explain the total variability associated with the original data set.

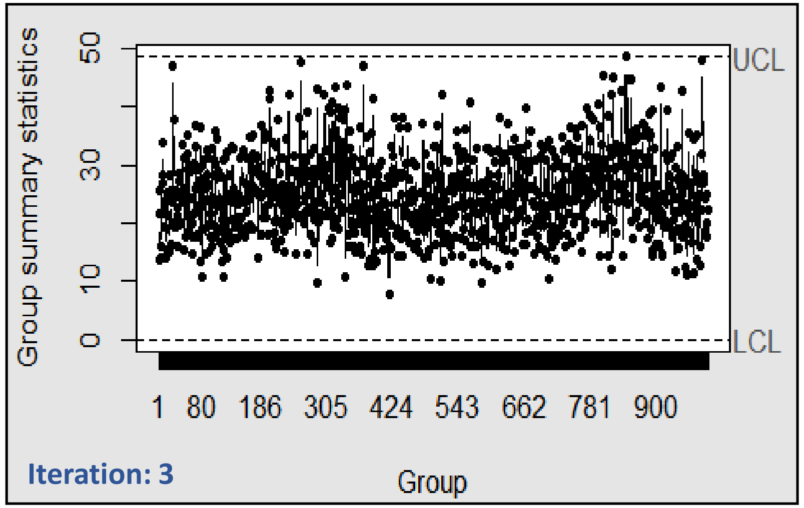
CONTROL CHART: HOTELLING T^2



Number of samples : 1000
UCL = 48.57
Out of control = 7



Number of samples : 993
UCL = 48.57
Out of control = 1



Number of samples : 992
UCL = 48.57
Out of control = 0

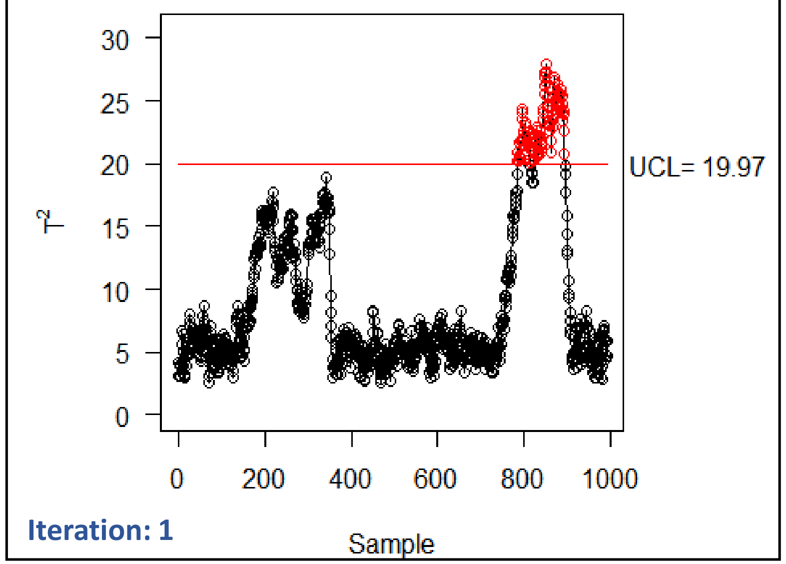
- We have considered the case iii (a) for generating T^2 chart as sample size n =1 is given.
- For case iii (a) UCL is decided by BETA distribution.
- UCL corresponds to β value for α = 0.0027. So, UCL = 48.57.

- In 1st iteration, there are 7 points out of control.
- In 2nd iteration, there is 1 point of control with number of sample 993.
- In 3rd iteration, all the points are in control with number of sample 992.

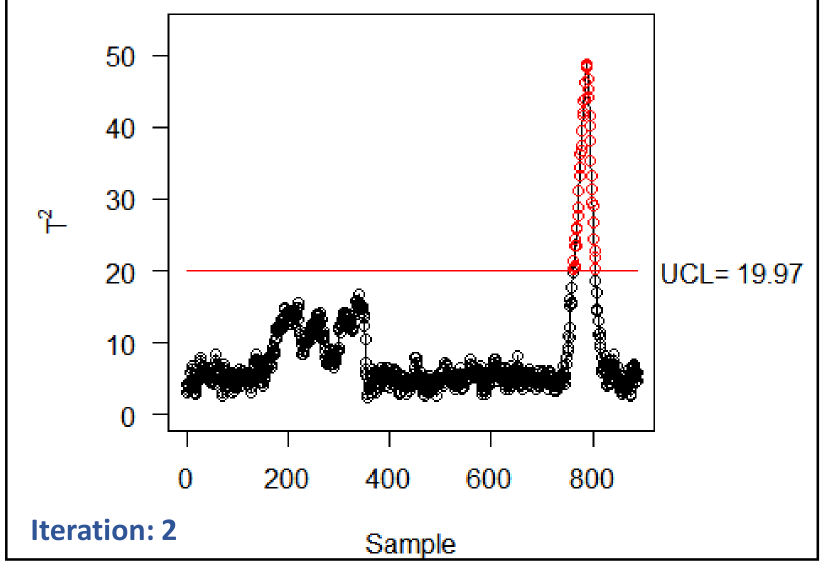
- By using T^2 chart, We removed all the data points causing large spike.

- For removing the points causing sustained mean shift, we can use either m-EWMA or m-CUSUM analysis.

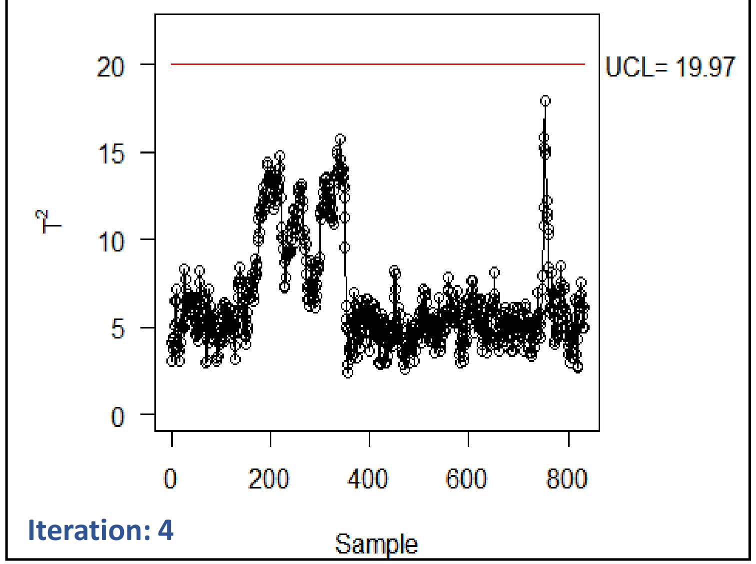
CONTROL CHART: m-CUSUM



Iteration: 1



Iteration: 2



Iteration: 4

Number of samples : 992
UCL = 19.97
Out of control = 105

Number of samples : 887
UCL = 19.97
Out of control = 45

Number of samples : 832
UCL = 19.97
Out of control = 0

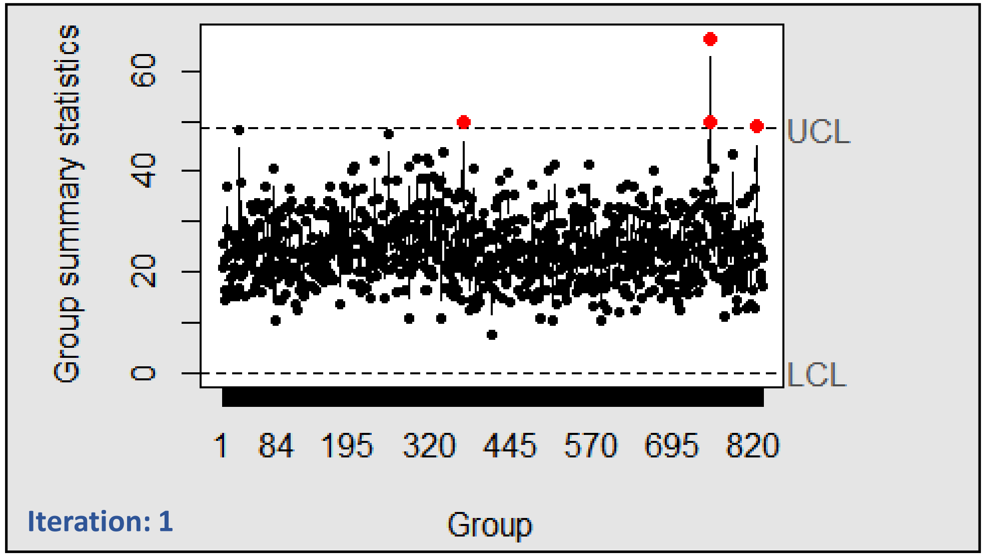
- This approach is to eliminate mean shift of statistical distance 3.

- UCL is calculated by extrapolation from literature data. So, UCL = 19.97.

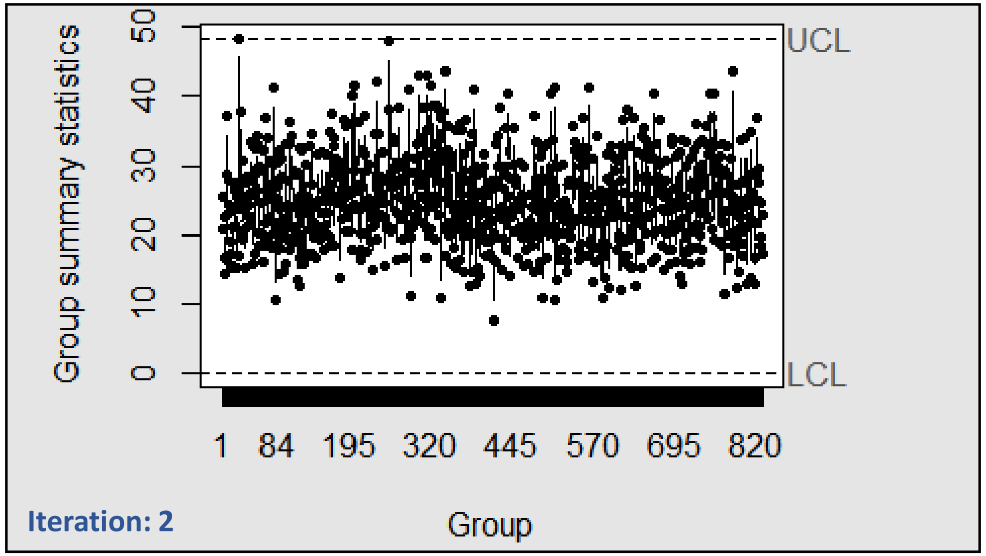
- In 1st iteration, there are 105 points out of control.
- In 2nd iteration, there are 45 points out of control with number of samples 887.
- In 3rd iteration, there are 10 points out of control with number of samples 842.
- In 4th iteration, all the points are in control with number of samples 832.

- By using m-CUSUM, we eliminated all the data points responsible for sustained mean shift.

CONTROL CHART: HOTELLING T^2 (After m-CUSUM)



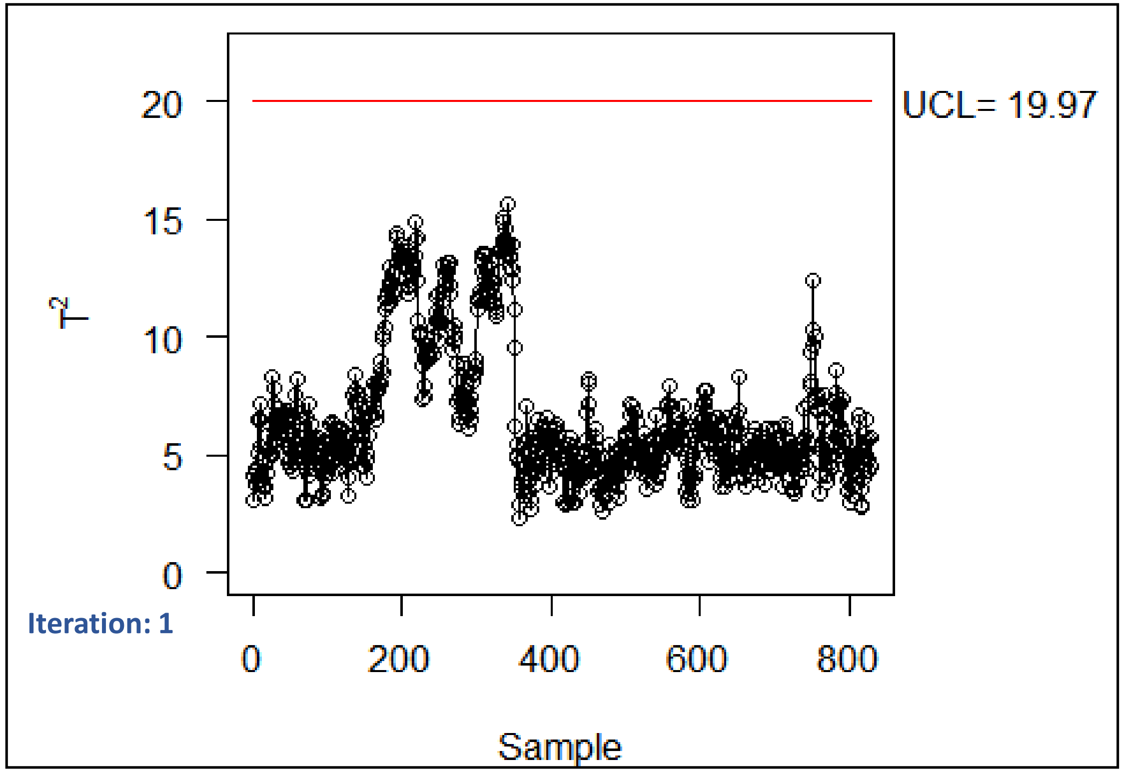
Number of samples : 832
UCL = 48.45
Out of control = 4



Number of samples : 828
UCL = 48.45
Out of control = 0

- After removing the points using m-CUSUM, we performed T^2 chart again to verify that all points are in control or not.
- As shown in graph, In 1st iteration there are 4 points out of control in 832 samples. So, we removed them.
- In 2nd iteration, all the points are in control with 828 samples.
- By performing this, all the points responsible for spike change have been removed.

CONTROL CHART: m-CUSUM



Number of samples : 828
UCL = 19.97
Out of control = 0

- To ensure that all points are in control, we performed m-CUSUM chart one more time.
- As shown in graph, we can say that all the points are now in control.
- So, total 172 points are out of control in which 12 points are responsible for spike change and 160 are responsible for sustained mean shift.
- Neither T^2 nor m-CUSUM detects spike of individual quality characteristic.
- But we can not use multiple univariate chart as it has some drawbacks which are discussed on next slide.

REASONS FOR NOT USING INDIVIDUAL X BAR CHART

Reasons for not using individual X bar chart:

- We have not selected the individual charts for principal components because a process may be out of control even when the individual charts are in control.

- There are 25 individual X bar chart in our case.

- One of the major drawback of using multiple univariate control chart is either α or β error could be seriously inflated and this is true when the number of variables are more than 10. In our case they are 25. So, it could be a problem.

- It is difficult to adjust errors when we have more than 10 variables and this is known as "curse of dimensionality".

- Another drawback of using multiple univariate chart is it is unable to detect change in correlation among variables.

- During our pre exploratory data analysis, we found that column7 and column21 contains out of control points due to either mean shift or covariance change. We can not surely say that this points are out of control because individual column, as this OOC points might be cumulative effect of all the variables.

IN CONTROL DISTRIBUTION PARAMETERS (μ_0 & $\sum_0$)

Covariance Matrix ($\sum_0$)

	X__1	X__2	X__3	X__4	X__5	X__6	X__7	X__8	X__9	X__10	X__11	X__12	X__13	X__14	X__15	X__16	X__17	X__18	X__19	X__20	X__21	X__22	X__23	X__24	X__25
X__1	0.494606	0.022718	-0.02474	0.016734	0.002761	3.19E-05	0.013663	4.13E-05	0.005872	0.000458	0.00831	-0.01033	0.020823	0.005173	0.036378	0.00292	0.015691	-0.00298	0.014459	0.019257	0.024555	-0.00804	-0.01946	0.002063	0.000222
X__2	0.022718	0.550013	-0.01978	0.020819	0.011919	0.000359	-0.03492	-0.02477	-0.01811	-0.01071	0.010539	0.017468	-0.01621	0.038818	0.015228	-0.02588	-0.02234	0.027778	-0.00607	0.03478	0.025308	0.002073	0.002508	-0.01429	0.024346
X__3	-0.02474	-0.01978	0.495576	0.022383	0.005564	-0.03615	-0.05357	-0.02593	0.004637	0.014774	0.008062	0.005316	0.020932	-0.00826	-0.00371	-0.00488	-0.00383	0.003719	-0.01061	-0.02018	-0.0139	-0.01167	0.008154	0.013721	0.003329
X__4	0.016734	0.020819	0.022383	0.522341	-0.00244	-0.09269	-0.16604	-0.00874	0.03164	0.032318	0.007953	0.01633	0.009642	0.004025	0.001053	-0.02531	0.011539	0.000525	0.017556	0.032673	0.016029	0.012114	0.012718	0.017893	0.025796
X__5	0.002761	0.011919	0.005564	-0.00244	0.481303	-0.03903	-0.04049	-0.00124	0.016659	-0.00627	0.008263	-0.00229	0.017572	0.027921	0.00246	0.050444	-0.02305	0.016136	0.04483	0.025608	-0.00262	0.012624	-0.00917	0.001797	-0.0196
X__6	3.19E-05	0.000359	-0.03615	-0.09269	-0.03903	0.743086	0.375458	-0.02187	-0.03148	-0.02435	0.013267	0.009519	0.018814	-0.01551	-0.02671	0.027527	0.010761	-0.00078	-0.02603	-0.04097	-0.00378	-0.03082	-0.02249	-0.05031	-0.00328
X__7	0.013663	-0.03492	-0.05357	-0.16604	-0.04049	0.375458	1.145372	-0.05635	-0.0371	-0.04892	0.013164	0.047821	-0.02302	-0.00465	-0.02424	0.005891	0.011838	-0.03399	0.001363	-0.02916	0.025709	-0.0229	-0.03249	-0.04502	0.008507
X__8	4.13E-05	-0.02477	-0.02593	-0.00874	-0.00124	-0.02187	-0.05635	0.501394	-0.00896	0.039334	-0.00758	0.029421	-0.01456	-0.01922	0.030611	0.002855	0.024113	-0.01827	0.020186	-0.00704	0.020452	0.037798	0.020387	0.007772	0.009836
X__9	0.005872	-0.01811	0.004637	0.03164	0.016659	-0.03148	-0.0371	-0.00896	0.477147	0.024619	0.000839	-0.01818	0.007764	-0.02633	0.011385	-0.01686	-0.00529	-0.01295	0.001388	-0.01051	-0.01311	-0.0162	0.004033	0.016859	0.016632
X__10	0.000458	-0.01071	0.014774	0.032318	-0.00627	-0.02435	-0.04892	0.039334	0.024619	0.520492	-0.01902	0.006434	-0.04207	-0.01983	-0.00217	0.010247	-0.00535	-0.01074	0.021558	0.004344	0.009736	0.002804	-0.00169	0.001701	0.001881
X__11	0.00831	0.010539	0.008062	0.007953	0.008263	0.013267	0.013164	-0.00758	0.000839	-0.01902	0.507645	-0.00977	-0.00921	-0.01481	-0.00153	-0.00154	0.001306	0.005632	0.000595	0.01438	0.0427	0.011762	0.007868	0.014781	0.02919
X__12	-0.01033	0.017468	0.005316	0.01633	-0.00229	0.009519	0.047821	0.029421	-0.01818	0.006434	-0.00977	0.490679	0.009512	0.000987	0.005377	-0.00966	0.003083	0.032069	-0.01305	0.005159	-0.0337	-0.01851	-0.01407	0.026981	0.01662
X__13	0.020823	-0.01621	0.020932	0.009642	0.017572	0.018814	-0.02302	-0.01456	0.007764	-0.04207	-0.00921	0.009512	0.492244	0.026026	0.049175	-0.00207	-0.00938	-0.02019	-0.02063	0.020334	-0.01118	-0.01971	-0.03561	-0.01409	-0.01772
X__14	0.005173	0.038818	-0.00826	0.004025	0.027921	-0.01551	-0.00465	-0.01922	-0.02633	-0.01983	-0.01481	0.000987	0.026026	0.539256	0.039422	-0.00427	-0.02188	0.003006	0.000273	0.050499	-0.00834	-0.00102	-0.00028	-0.00859	-0.0059
X__15	0.036378	0.015228	-0.00371	0.001053	0.00246	-0.02671	-0.02424	0.030611	0.011385	-0.00217	-0.00153	0.005377	0.049175	0.039422	0.510233	0.015347	-0.00208	-0.01051	0.002791	0.02243	0.008436	-0.02837	0.009787	0.010575	-0.00917
X__16	0.00292	-0.02588	-0.00488	-0.02531	0.050444	0.027527	0.005891	0.002855	-0.01686	0.010247	-0.00154	-0.00966	-0.00207	-0.00427	0.015347	0.461019	-0.01631	-0.00044	0.010617	0.002549	0.003773	0.003788	-0.02506	0.019566	-0.00148
X__17	0.015691	-0.02234	-0.00383	0.011539	-0.02305	0.010761	0.011838	0.024113	-0.00529	-0.00535	0.001306	0.003083	-0.00938	-0.02188	-0.00208	-0.01631	0.459091	-0.01151	0.041749	0.008645	-0.00644	0.012773	0.037416	0.002189	-0.02503
X__18	-0.00298	0.027778	0.003719	0.000525	0.016136	-0.00078	-0.03399	-0.01827	-0.01295	-0.01074	0.005632	0.032069	-0.02019	0.003006	-0.01051	-0.00044	-0.01151	0.53793	-0.04272	-0.00606	-0.00592	0.017968	-0.01107	-0.00142	0.02561
X__19	0.014459	-0.00607	-0.01061	0.017556	0.04483	-0.02603	0.001363	0.020186	0.001388	0.021558	0.000595	-0.01305	-0.02063	0.000273	0.002791	0.010617	0.041749	-0.04272	0.503655	-0.00643	0.009676	0.01442	-0.00555	0.021747	-0.00917
X__20	0.019257	0.03478	-0.02018	0.032673	0.025608	-0.04097	-0.02916	-0.00704	-0.01051	0.004344	0.01438	0.005159	0.020334	0.050499	0.02243	0.002549	0.008645	-0.00606	-0.00643	0.495607	0.005096	0.005676	-0.0125	0.030442	-0.00013
X__21	0.024555	0.025308	-0.0139	0.016029	-0.00262	-0.00378	0.025709	0.020452	-0.01311	0.009736	0.0427	-0.0337	-0.01118	-0.00834	0.008436	0.003773	-0.00644	-0.00592	0.009676	0.005096	0.53898	-0.00862	0.008702	0.005045	0.026658
X__22	-0.00804	0.002073	-0.01167	0.012114	0.012624	-0.03082	-0.0229	0.037798	-0.0162	0.002804	0.011762	-0.01851	-0.01971	-0.00102	-0.02837	0.003788	0.012773	0.017968	0.01442	0.005676	-0.00862	0.499867	-0.00356	-0.02327	0.016056
X__23	-0.01946	0.002508	0.008154	0.012718	-0.00917	-0.02249	-0.03249	0.020387	0.004033	-0.00169	0.007868	-0.01407	-0.03561	-0.00028	0.009787	-0.02506	0.037416	-0.01107	-0.00555	-0.0125	0.008702	-0.00356	0.546366	-0.00391	0.005647
X__24	0.002063	-0.01429	0.013721	0.017893	0.001797	-0.05031	-0.04502	0.007772	0.016859	0.001701	0.014781	0.026981	-0.01409	-0.00859	0.010575	0.019566	0.002189	-0.00142	0.021747	0.030442	0.005045	-0.02327	-0.00391	0.475736	-0.0243
X__25	0.000222	0.024346	0.003329	0.025796	-0.0196	-0.00328	0.008507	0.009836	0.016632	0.001881	0.02919	0.01662	-0.01772	-0.0059	-0.00917	-0.00148	-0.02503	0.02561	-0.00917	-0.00013	0.026658	0.016056	0.005647	-0.0243	0.463888

Mean Vector (μ_0^T)

-0.00747	0.016091	-0.04801	-0.11786	-0.04541	0.245868	0.465908	0.008263	-0.01063	0.026361	-0.05583	-0.01261	0.003922	-0.00069	0.014528	-0.01709	-0.00971	-0.00973	0.019387	0.008737	0.004765	-0.02061	-0.00158	-0.02137	-0.00512

KEY TAKEAWAYS

- Due to high dimensionality of data, it is difficult to select a particular approach for setting up a process control detection method.

- To reduce the data dimensions, PCA is used but in our case it does not solve the purpose as principal components do not include total variability associated with original data. . So, we have to consider all the variables given in original data.

- We have plotted hotelling T^2 chart to detect large spike type change and removed all the data points responsible for this change.

- T^2 chart is unable to detect sustained mean shift type change. So, we have to use either m-CUSUM or m-EWMA chart to detect the sustained mean shift type change.

- Due to high popularity of m-CUSUM chart we have used it for detection of sustained mean shift and removed all the data points responsible for this change.

- After removing outlier from the T^2 and m-CUSUM, we get all the data points are in control.

- We have estimated distribution parameters μ_0 & $\sum_0$ from in control data which can be used as a baseline for future quality conrol analysis(Phase ii analysis).