ISEN 614 Advanced Quality Control

Fall 2016 Project

Dwarkanath Prabhu
Venkata Kartik Mutya

Industrial & Systems Engineering
Texas A&M University

Understanding the Problem

The problem at hand has 552 samples, each with 209 data points.

- n = 552
- p = 209

This can be denoted as $\{x_j\}$, j = 1,…, 552 and each x_j is a 209 x 1 vector. The task at hand is to identify in-control and out-of-control samples.

The μ_0 and Σ_0 for this data are not known. Hence, this is a Phase I analysis with a sample size of 1. We will use $\bar{\boldsymbol{x}}$ and S to estimate μ_0 and Σ_0.

Since the number of dimensions is very high, we will first reduce data using principal component analysis and then use the Hotelling chart to isolate in-control data.

Data Reduction (1/3)

For principal component analysis, we need $\bar{\boldsymbol{x}}$ and **S** for the sample.

Sample statistics:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$S = \frac{1}{n - 1}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$$

We will calculate eigenvalues and eigenvectors of S to find the reduced dimension. These eigenvectors will be used to form principal components from the original data

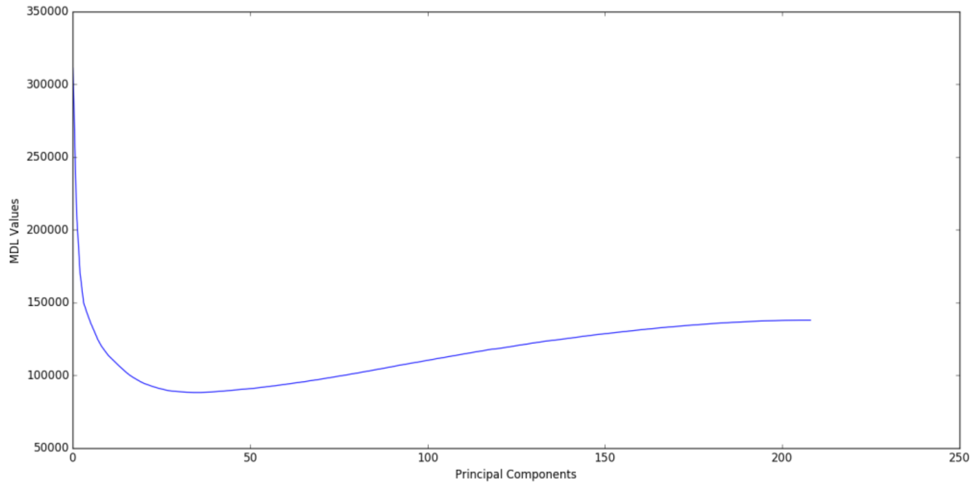
Data Reduction (2/3)

For the S matrix, we calculate the eigenvalues and arrange them in descending order.

We plot a graph using the formula:

$$\text{MDL (l)} = \text{n(p-l)log}(a_l / g_l) + \text{l(2p – l)log(n)/2}$$

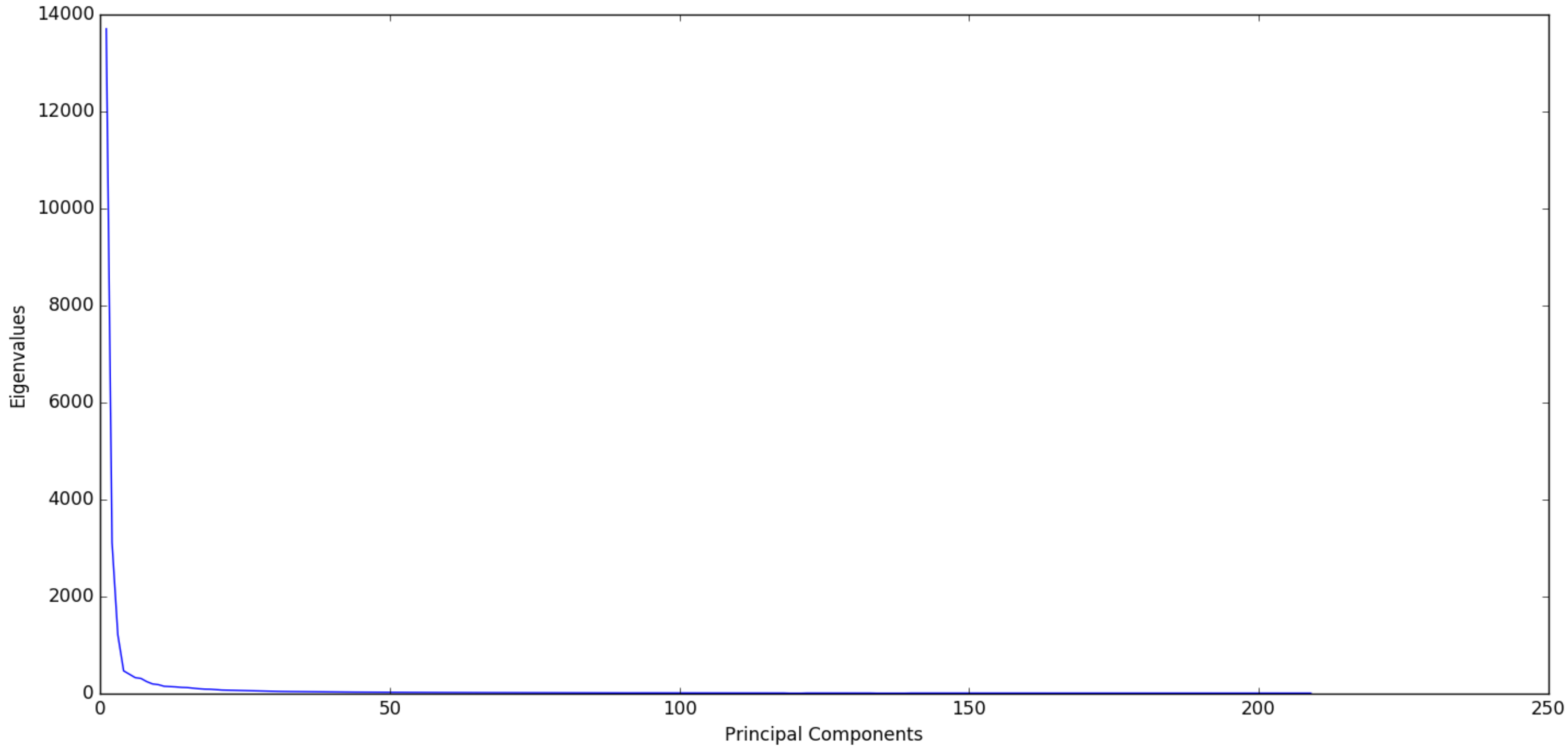
Where a_l, g_l are the arithmetic and geometric means respectively of the smallest (p – l) eigenvalues.



From the graph of MDL values plotted against l, the value of l for which MDL is minimum is 35.

Data Reduction (3/3)

35 is still a very large number. Hence we look at the scree plot i.e. the plot of eigenvalues against the number of principal components



From the scree plot, it can be seen that there is a bend where the x-axis value is 4. We will thus choose only the first 4 principal components for our analysis.

Principal Component Analysis

For Principal Component Analysis (PCA), we calculate the vector **y**, such that

$$y_i = e_i^T x$$

Where i = 1,..,4 and $e_i$is the i^{th} eigenvector of S.

As there are n (= 552) samples, there are n such y vectors of length 4 each.

We will now perform Phase I analysis on y.

Phase I Analysis

For Phase I analysis of y, we approximate the upper control limit using

$$UCL = \chi^2_{1-\alpha}(\mathrm{p})$$

Here, we have chosen α = 0.05. p is the reduced dimension, hence p = 4. This value comes to 9.49

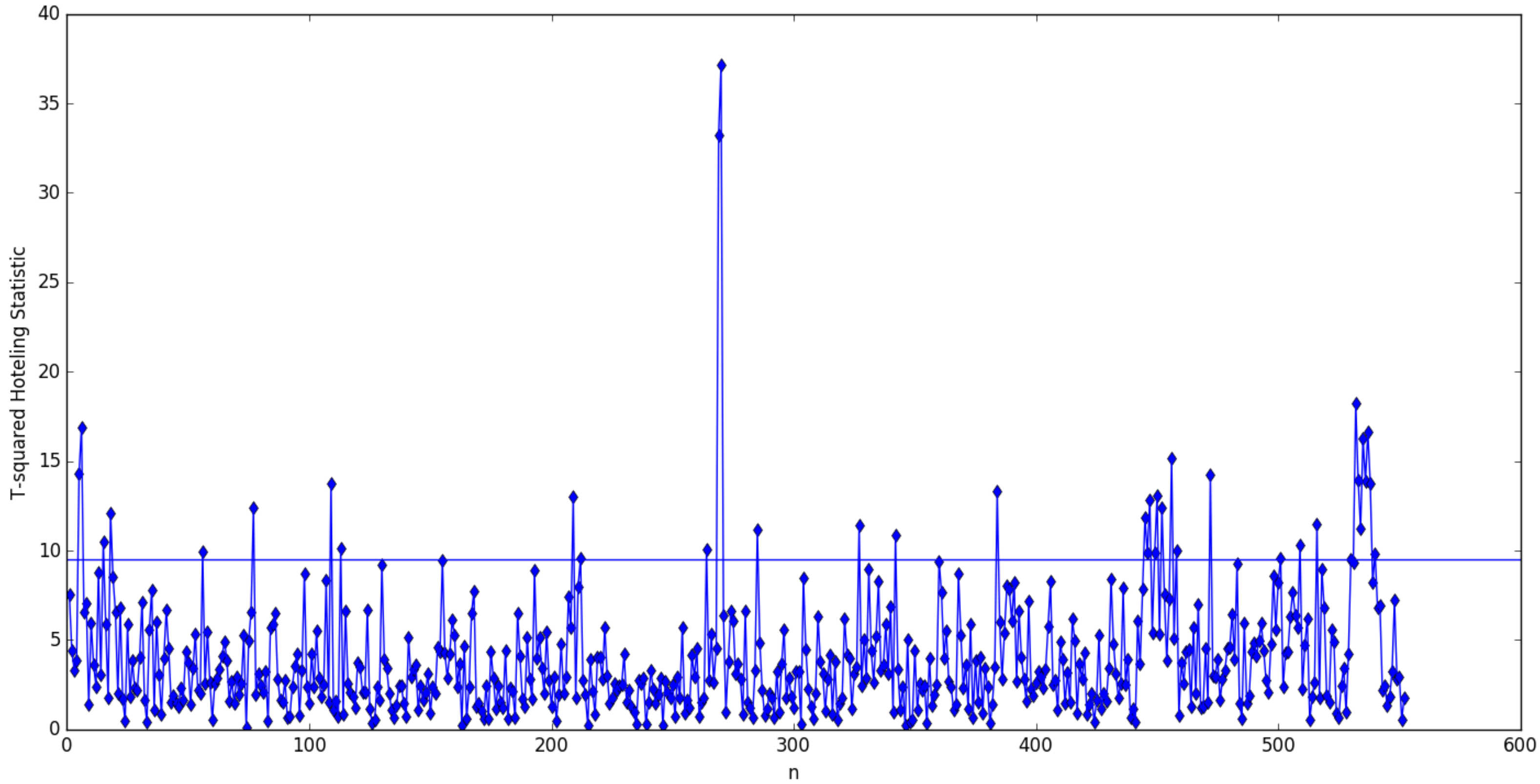
We will now plot the Hotelling T^2statistic for each sample. To isolate in-control data, we will remove out-of-control samples and recalculate the T^2 statistic till we are left with only in-control samples.

To calculate T^2 statistic, we use:

$$T^2 = \left(y_j - \bar{y}\right)^T S^{-1}(y_j - \bar{y})$$

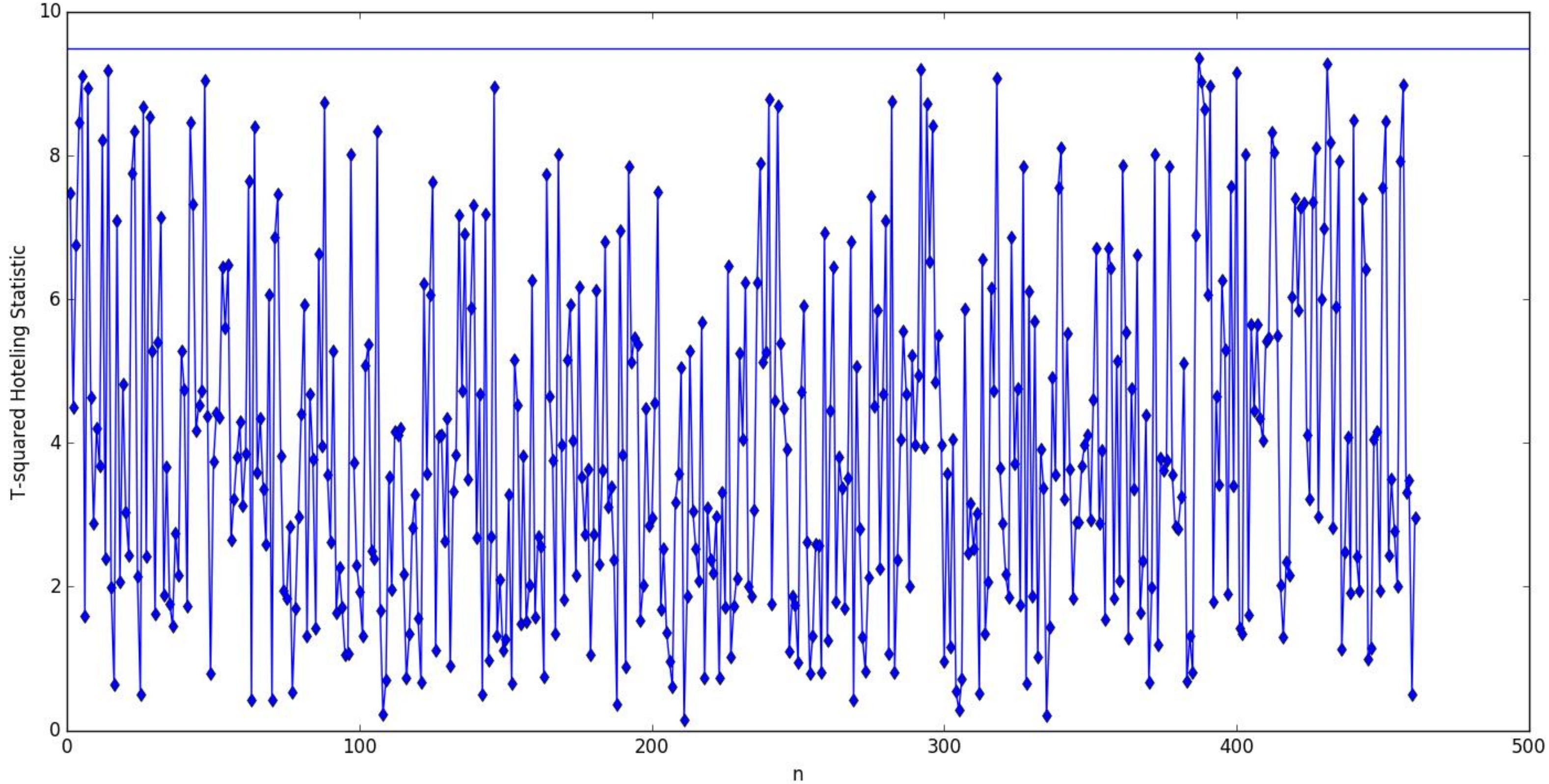
Where $\bar{y}$ is the mean of y, S is the covariance matrix of y and j is the sample number. These can be calculated the same way as we did for x on Slide 3.

Hotelling Statistic First Iteration



In this plot, it can be seen that there are several samples that are out of control

Hotelling Statistic In-Control Samples



In this plot, all samples are in control. In total there are 461 in-control samples.

Thank You