

ROBUST DESIGN WITH BINARY RESPONSE USING MAHALANOBIS
TAGUCHI SYSTEM

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

BARIŞ YENİDÜNYA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
INDUSTRIAL ENGINEERING

JUNE 2009

Approval of the thesis:

ROBUST DESIGN WITH BINARY RESPONSE USING MAHALANOBIS TAGUCHI SYSTEM

submitted by **BARIŞ YENİDÜNYA** in partial fulfillment of the requirements for the degree of
Master of Science in Industrial Engineering Department, Middle East Technical University
by,

Prof. Dr. Canan Özgen

Dean, Graduate School of Natural and Applied Sciences

Prof. Dr. Nur Evin ÖZDEMİREL

Head of Department, Industrial Engineering Dept., METU

Prof. Dr. Gülser Köksal

Supervisor, Industrial Engineering Dept., METU

Examining Committee Members:

Prof. Dr. Sinan Kayalığıl

Industrial Engineering Dept., METU

Prof. Dr. Gülser Köksal

Industrial Engineering Dept., METU

Assoc. Prof. Yasemin Serin

Industrial Engineering Dept., METU

Assoc. Prof. Murat Caner Testik

Industrial Engineering Dept., Hacettepe University

Assist. Prof. Serhan Duran

Industrial Engineering Dept., METU

Date: 16.06.2009

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: BARIŞ YENİDÜNYA

Signature :

ABSTRACT

ROBUST DESIGN WITH BINARY RESPONSE USING MAHALANOBIS TAGUCHI SYSTEM

Yenidünya, Barış

M.S., Department of Industrial Engineering

Supervisor : Prof. Dr. Gülser Köksal

June 2009, 129 pages

In industrial quality improvement and design studies, an important aim is to improve the product or process quality by determining factor levels that would result in satisfactory quality results. In these studies, quality characteristics that are qualitative are often encountered. Although there are many effective methods proposed for parameter optimization (robust design) with continuous responses, the methods available for qualitative responses are limited. In this study, a parameter optimization method for solving binary response robust design problems is proposed. The proposed method uses Mahalanobis Taguchi System to form a classification model that provides a distance function to separate the two response classes. Then, it finds the product/process variable settings that minimize the distance from the desired response class using quadratic programming.

The proposed method is applied on two cases previously studied using Logistic Regression. The classification models are formed and the parameter optimization is conducted using the formed MTS models. The results are compared with those of the Logistic Regression. Conclusions and suggestions for future work are given.

Keywords: Robust Design, Parameter Optimization, Mahalanobis Taguchi System, binary quality characteristics (response)

ÖZ

MAHALANOBIS TAGUCHI SİSTEMİ KULLANARAK İKİLİ ÇIKTI İLE ROBUST TASARIM

Yenidünya, Barış

Yüksek Lisans, Endüstri Mühendisliği Bölümü

Tez Yöneticisi : Prof. Dr. Gülser Köksal

Haziran 2009, 129 sayfa

Endüstriyel kalite iyileştirme ve tasarım çalışmalarında önemli amaçlardan biri ürün/süreç kalitesinin hedeflenen kalite sonuçlarına ulaşmasını sağlayacak faktör seviyelerinin belirlenmesidir. Bu çalışmalarda nicel kalite özellikleri sık karşılaşılan bir durumdur. Sürekli çıktılarda parametre optimizasyonu (robust tasarım) için çok sayıda yöntem olmasına rağmen, nicel sonuçlar için geliştirilmiş metotlar kısıtlıdır. Bu çalışmada, ikili çıktılı robust tasarım problemleri için bir parametre optimizasyon metodu önerilmektedir. Önerilen metod Mahalanobis Taguchi Sistemi kullanarak iki sonuç sınıfını ayıracak uzaklık fonksiyonunu sağlayan bir sınıflandırma modeli kurmakta, daha sonra da istenilen sonuç sınıfına uzaklığı en aza indirgeyecek ürün/süreç değişken seviyelerini kuadratik programlama kullanarak bulmaktadır.

Önerilen yöntem daha önce Lojistik Regresyon kullanılarak incelenmiş iki vaka analizi üzerinde çalışılmıştır. Sınıflandırma modelleri kurulmuş ve kurulan modeller üzerinde parametre optimizasyonu uygulanmıştır. Sonuçlar Lojistik Regresyon sonuçları ile karşılaştırılmıştır. Sonuçlar ve ileride yapılabilecek çalışmalar için öneriler verilmiştir.

Anahtar Kelimeler: Robust Tasarım, Parametre Optimizasyonu, Mahalanobis Taguchi Sistemi, İkili kalite özellikleri (çıkıtı)

To my family

ACKNOWLEDGEMENTS

I would like to express my thanks and gratitude to Prof. Dr. Gülser Köksal for her valuable guidance, supervision and understanding throughout the research.

I would like to thank my managers Kıymet Uraz and Hasan Yıldız for their patience and understanding.

I would also like to thank my friend Dilber Ayhan for her support throughout the research.

Finally, I want to express my sincere gratitude to my family for their patience and emotional support they have provided in my masters study.

TABLE OF CONTENTS

ABSTRACT	iv
ÖZ.....	v
ACKNOWLEDGEMENTS	vii
LIST OF TABLES.....	x
LIST OF FIGURES.....	xii
CHAPTER	
1. INTRODUCTION	1
2. LITERATURE SURVEY AND BACKGROUND	3
2.1. Mahalanobis Taguchi System	3
2.1.1. Mahalanobis Distance	3
2.1.2. The Steps of Mahalanobis Taguchi System.....	6
2.2. Handling Small and Imbalanced Data in MTS	15
2.2.1. Multicollinearity and Singularity.....	16
2.3. Robust Design.....	18
2.4. Threshold Determination	23
2.5. Performance Measures	26
2.6. Resampling Methods for Small and Unbalanced Data in MTS Classification	30
2.6.1. SMOTE	31
2.7. Methods for Robust Design with Qualitative Responses.....	32
2.7.1. Logistic Regression	32
2.7.2. Accumulation Analysis.....	33
2.7.3. Scoring Schemes.....	35
2.7.4. Weighted Signal-to-Noise Ratio.....	36
3. PROPOSED THRESHOLD DETERMINATION METHOD	38
3.1.1. Wisconsin Breast Cancer Dataset	41
3.1.2. Pima Indian Diabetes Dataset.....	42
3.1.3. Magic Gamma Telescope Dataset	46

4.	PARAMETER OPTIMIZATION USING MTS WITH BINARY RESPONSE	52
4.1.	The Approach	52
4.1.1.	Generation of Mahalanobis Space	58
4.1.2.	Evaluation of Discrimination Power and Selection of Important Variables ...	59
4.1.3.	Threshold Selection	61
4.1.4.	Parameter Optimization using Quadratic Programming	61
4.2.	Applications.....	62
4.2.1.	Casting Defect Problem	62
4.2.2.	Customer Satisfaction from a Driver's Seat Case	83
4.3.	Comparison of the Proposed Method and Other Robust Design Methods	105
5.	CONCLUSION & FUTURE WORK	109
	REFERENCES.....	111
	APPENDICES	
	A. ANOVA TABLE OF CUSTOMER SATISFACTION DATASET.....	115
	B. MAIN EFFECTS PLOT FOR CUSTOMER SATISFACTION DATASET	117
	C. MAHALANOBIS TAGUCHI GRAM-SCHMIDT SYSTEM	118
	D. OTHER METHODS IN HANDLING SMALL DATA IN MTS	128

LIST OF TABLES

Table 1 L8(2 ⁷) Orthogonal Array.....	13
Table 2 Factors and their levels.....	20
Table 3 Designed experiment	21
Table 4 Confusion Matrix	26
Table 5 Dataset Information	39
Table 6 Wisconsin Breast Cancer Dataset Performance Measures	42
Table 7 Performance Measures of the Pima Indian Diabetes Dataset	44
Table 8 Performance Measures of Pima Indian Diabetes Dataset with Normal Group Reduced to %20 of the Dataset.....	46
Table 9 Performance Measures of Magic Gamma Telescope Dataset	48
Table 10 Performance Measures of Threshold Methods in Imbalanced Magic Gamma Telescope Dataset.....	48
Table 11 Mean values of the variables for the normal and abnormal groups.....	63
Table 12 Standard deviations of the normal and abnormal groups	64
Table 13 Minimum and maximum values of the variables.....	65
Table 14 Correlation Matrix	67
Table 15 Basic statistics of the MTS model using all variables	68
Table 16 Performance measures of the MTS model with all variables	69
Table 17 Parameter levels that result in the maximum S/N.....	70
Table 18 Basic statistics of the optimized MTS model	70
Table 19 Performance Measures of the Optimized Model	71
Table 20 Mean Values of the Variables.....	73
Table 21 Covariance Matrix of the Variables	73
Table 22 Production Limits of the Parameters	74
Table 23 Optimal Parameter levels of the MTS model.....	75
Table 24 Comparison of the proposed factor levels.....	77
Table 25 Confusion Table of the MTS and LR models	79
Table 26 Comparison of the Original MTS and the MTS formed using LR variables.....	80
Table 27 Confusion table for the MTS (using LR variables) and LR Models	80
Table 28 Optimal parameter levels of the LR model	81
Table 29 Optimal Parameter Levels of MTS using LR variables	82
Table 30 Optimal Parameter Levels of LR using MTS variables	82
Table 31 Comparison of the Optimal Solutions of MTS and LR	83
Table 32 Variables of the Customer Satisfaction Dataset	85
Table 33 Basic Statistics of the Normal/Abnormal Groups	88
Table 34 Performance Measures of the MTS Model Using All Variables	89
Table 35 Basic Statistics of the two MTS models	90

Table 36 Comparison of two MTS models.....	91
Table 37 Basic Statistics of the MTS model with 34 Variables.....	92
Table 38 Performance Measures of the MTS model with 34 variables.....	92
Table 39 Mean Values of the Variables.....	95
Table 40 Optimal Parameter Levels of the MTS model.....	98
Table 41 Parameter Levels for Maximum Distance From the Normal Group.....	98
Table 42 Performance Measure of the LR and the MTS models.....	100
Table 43 Confusion matrix for the LR model.....	100
Table 44 Optimal Parameter Levels of the LR model.....	101
Table 45 Optimal parameter levels of the MTS model using optimal LR solution.....	103
Table 46 Performance measures of the MTS model formed with LR variables.....	103
Table 47 Optimal parameter levels for the MTS model formed using LR variables.....	104

LIST OF FIGURES

Figure 1 P Diagram	19
Figure 2 ROC curve (Bradley 1997).....	29
Figure 3 Pseudocode of SMOTE (Chawla et al.(2002))	32
Figure 4 Wisconsin Breast Cancer Dataset Performance Measures.....	43
Figure 5 Performance Measures of the Pima Indian Diabetes Dataset.....	45
Figure 6 Performance Measures of Imbalanced Pima Indian Diabetes Dataset.....	47
Figure 7 Performance Measures in Magic Gamma Telescope Dataset	49
Figure 8 Performance Measures of Imbalanced Magic Gamma Telescope Dataset	51
Figure 9 Proposed Approach.....	54
Figure 10 Mean values of the variables for the normal and abnormal groups	64
Figure 11 Standard Deviations of the Normal and Abnormal Groups	65
Figure 12 Minimum and maximum values of the variables	66
Figure 13 Correlation graphs of the variables	68
Figure 14 Mahalanobis Distances using All Variables.....	72
Figure 15 Mahalanobis Distances using Selected Variables	76
Figure 16 MD vs. $P(Y=1)$ graph of the samples.....	79
Figure 17 Mean Values of the Questions	86
Figure 18 Standard Deviations of the Questions.....	87
Figure 19 Mean Values of the Anthropometric Measurements.....	87
Figure 20 Standard Deviations of the Anthropometric Measurements	88
Figure 21 Mahalanobis Distances Using 41 Variables	93
Figure 22 Mahalanobis Distances Using 23 Variables	96
Figure 23 Mahalanobis Distances Using 34 Variables	97
Figure 24 $P(Y=1)$ vs. MD graph for the observations.....	102
Figure 25 $P(Y=1)$ vs. MD graph for the observations.....	104

CHAPTER 1

INTRODUCTION

In industrial quality improvement and design studies, an important aim is to improve the product or process quality by determining controllable design parameter levels that would result in satisfactory quality results. Robust design is an approach used to design products/processes which are insensitive to uncontrollable factors.

Industrial quality experiments often result in quality characteristics that are qualitative. Common reasons for this are the technical difficulties and economical infeasibilities in accurately measuring factors. A common example to a qualitative response is the experiment output classified as defective or non-defective.

Although there are many effective methods proposed for parameter optimization with quantitative response variables, the methods available for qualitative responses are limited. In this study, a parameter optimization method for design problems with a binary response is proposed. The aim is to find the controllable design parameter settings (target values) that yield desired (non-defective) outcomes no matter how noise factors behave in the system. The proposed approach uses Mahalanobis Taguchi System (MTS) to model the relationship between the control factors and the response. This model actually measures a distance of a product/process realization (in terms of controllable parameter values) from the mean of normal (non-defective) class of observations. Then it finds the optimal design parameter levels by minimizing this distance equation.

Mahalanobis Taguchi System is a multivariate classification method that defines a distance function to assign new observations to the most appropriate classes. It is commonly used

for binary classes (responses). It uses Mahalanobis Distance as a measurement scale in combination with Taguchi's robust design methods to select the variables that significantly contribute to the distance measurement.

After the MTS model is formed and verified, the model function is used inversely, going from the response to the control factors to determine the optimum factor levels. This inverse problem is solved using quadratic programming.

In Chapter 2, Mahalanobis Taguchi System is explained in detail. Data collection, model forming and model evaluation steps of Mahalanobis Taguchi System are explained in detail as well as the advantages and disadvantages of the method. In addition, robust design is explained and some other methods used for robust design with qualitative responses are presented.

In Chapter 3, a threshold determination method for MTS is proposed. The proposed threshold method takes into consideration the performance measures, aiming to maximize the classification performance.

In Chapter 4, use of Mahalanobis Taguchi System for parameter optimization with a binary response is described. A classification model based on the data collected for optimization is formed, and this model is solved using mathematical programming to find the optimal parameter levels that would result in the desired output. The desired output is typically producing more of the outputs belonging to the desired class.

The proposed method is applied on two cases previously studied with Logistic Regression. The classification models are formed and the parameter optimization is conducted using the formed MTS model. The results are compared with the results of the Logistic Regression.

In Section 4.3, the proposed method is compared with some others known in the literature to solve similar robust design problems.

Finally, in Chapter 5 concluding remarks and future research directions are provided.

CHAPTER 2

LITERATURE SURVEY AND BACKGROUND

Even though there are effective methods developed optimization with quantitative response variables (Wu and Yeh 2006) (Taguchi 1991), the methods available for qualitative responses are limited. This study aims to use Mahalanobis Taguchi System for parameter optimization in cases with binary responses. The proposed approach aims to make use of Mahalanobis Taguchi System as a mean to understand the relationship between the control factors (decision variables) and the response.

2.1. Mahalanobis Taguchi System

Mahalanobis Taguchi System (MTS) is a multivariate diagnosis and forecasting method. MTS is a combination of Mahalanobis Distance (MD) and Taguchi's robust design methodologies. MTS uses MD as a distance metric to form a measurement scale that determines the distance of a unit to a particular reference group.

Taguchi and Jugulum (2000) state the objective of MTS as "to develop and optimize a diagnostic system with a measurement scale that measures abnormality.". However, MTS is used not only for diagnosis but also for classification and prediction in the literature.

2.1.1. Mahalanobis Distance

In one dimension, the Euclidean distance between two points $\mathbf{y} = (y_1)$ and $\mathbf{w} = (w_1)$ is

$$D = \sqrt{(y_1 - w_1)^2} \quad [2.1]$$

In order to measure how many standard deviations two points are apart from each other, Euclidean distance can be divided to the standard deviation:

$$SD = \sqrt{\frac{(y_1 - w_1)^2}{\sigma^2}} \quad [2.2]$$

This distance is called the statistical (or standardized) Euclidean distance.

In R^n , the distance between points $\mathbf{y} = (y_1, y_2, \dots, y_n)$ and $\mathbf{w} = (w_1, w_2, \dots, w_n)$ is

$$D = \sqrt{(y_1 - w_1)^2 + (y_2 - w_2)^2 + \dots + (y_n - w_n)^2} \quad [2.3]$$

which can be written as

$$D = \sqrt{(\mathbf{y} - \mathbf{w})^T (\mathbf{y} - \mathbf{w})} \quad [2.4]$$

If \mathbf{w} is the origin, the equation becomes

$$D = \sqrt{(y_1 - 0)^2 + (y_2 - 0)^2 + \dots + (y_n - 0)^2} \quad [2.5]$$

or

$$(y_1)^2 + (y_2)^2 + \dots + (y_n)^2 = D^2 \quad [2.6]$$

which is the equation of a spheroid.

In the multivariate case, the standardized Euclidean Distance becomes

$$SD = \sqrt{\sum_{j=1}^p \left(\frac{y_{ij} - w_{kj}}{s_j} \right)^2} \quad [2.7]$$

where s_j is the standard deviation of the j^{th} variable.

This equation can also be written as:

$$SD^2 = (\mathbf{y} - \mathbf{w})^T \mathbf{S}^{-1} (\mathbf{y} - \mathbf{w}) \quad [2.8]$$

where \mathbf{S} is the diagonal matrix $\mathbf{S} = (s_1, s_2, \dots, s_n)$.

Although Standardized Euclidean Distance calculates how many standard deviations two points are apart, it does not take the correlations between the variables into consideration. If the correlations between the variables are also to be taken into consideration, S is replaced by the covariance matrix Σ .

$$MD^2 = (\mathbf{y} - \mathbf{w})^T \Sigma^{-1} (\mathbf{y} - \mathbf{w}) \quad [2.9]$$

The inverse of the covariance matrix in the equation standardizes all variables to the same variance and removes the correlations between the variables. If the variables are not correlated, then the Mahalanobis distance is the same as the standardized Euclidean distance. (Rousseeuw and van Zomeren 1990)

This distance measure is known as Mahalanobis distance (MD). P C Mahalanobis developed Mahalanobis distance, a statistical distance measure based on the multivariate normal distribution. (Mahalanobis 1936)

Originally, MD was developed to measure the distance between two multivariate normal distributions having different means but the same covariance matrix: $N(\mu_1, \Sigma)$ and $N(\mu_2, \Sigma)$. (Dasgupta 1993)

Probability Density Function of a multivariate normal distribution at x_i is

$$f(x_i) = \frac{1}{|\Sigma|^{\frac{1}{2}} (2\pi)^{p/2}} e^{-\frac{1}{2} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)} \quad [2.10]$$

If the probability density function $f_x(x_1, x_2, \dots, x_p)$ is graphed in two dimensions, it can be seen that it has the shape of an ellipse. In R^n , points with equal densities can be found by taking the natural logarithm of $f(x)$. After arranging, the result is:

$$(\mathbf{x} - \mathbf{y})^T \Sigma^{-1} (\mathbf{x} - \mathbf{y}) = c^2 \quad [2.11]$$

which is the equation of the Mahalanobis Distance. Hence, points with equal density have the same Mahalanobis Distance.

2.1.2. The Steps of Mahalanobis Taguchi System

1. Generation of the Normal Space

Mahalanobis Taguchi System defines a reference group to construct a measurement scale that measures the distance of the samples to that particular group. This reference group is called the "normal group", and the samples that are outside this group are considered to be "abnormal" samples.

The first step in MTS is to construct a basis (Mahalanobis Space) for a normal group. This constructed basis will be used as a reference to measure the normality of other observations.

To construct the basis, the p variables that are considered to be important in the definition of normality are identified. Due to the effects of the number of variables and the number of observations in the data explained throughout in this study, the choice of important variables is an important step and it depends on the experience of the experimenters doing the experiment. (Taguchi and Jugulum 2000)

The variables can either be continuous, discrete, or categorical. Discrete, and continuous variables can be used as they are without any modification. For categorical variables, it is suggested that they are converted into binary variables (For example, male=0 and female=1) If there are more than two levels in the variable, a series of dummy variables, each defining the presence of a level, can be used. (Taguchi, Chowdhury and Yuin 2001)

Although the Mahalanobis Distance was developed with the assumption of multivariate normal distribution, Taguchi state that MTS is a data analytic method with no distribution assumptions. However, this have been criticized in by Hawkins (2003) and Abraham and Variyath (2003), stating that normality assumption have to be checked, and transformations are made whenever required.

2. Collection of Samples To Construct the Normal Group

In the second step, n samples that will be used to construct the normal group are collected. Taguchi and Jugulum (2000) state that the number of samples must be greater than the number of variables, which is a requirement to obtain a nonsingular correlation matrix of the variables. Still, this may not be sufficient. In cases where the correlation matrix of the population is not known and the sample correlation matrix is to be used, there is a lower limit on the number of samples in order to estimate a correlation matrix in a multivariate normal case. Several authors have provided different lower limits on the number of samples.

In their work, Hirohisa et al. (2006) state that the MTS method "empirically requires that the number of samples is greater than three-times the number of variables.". Foley (1972) had also worked on this issue and had shown that "the ratio of the number of samples to the number of variables has to be greater than 3 in order to minimize the misclassifications in determining whether the samples are from the same multivariate distribution or not. Decreasing the sample size is shown to increase false negatives".

Young (1978) has shown that although it is minimally required that the number of samples are as many as the number of variables in order to have a nonsingular correlation matrix, there is also a lower limit to the number of samples needed to "characterize a multivariate Gaussian process with n samples". If the sample size is small, difficulties arise in determining whether the samples are originated from the same multivariate normal distribution or not.

Later in their work, Leese and Main (1994) has also shown that when number of samples forming the normal group is small, these samples have a large effect on the Mahalanobis Space formed. When the number of samples is small, adding or removing a new sample to the normal group may have significant effects.

If in this step, the samples that would form the normal group cannot be clearly identified, Mahalanobis Space can be constructed using all samples believed to be normal and their MDs can be calculated. Then in an iterative process, outlier samples with high MDs can be

moved from the normal group to the abnormal group, and the MDs are calculated again. This should continue until no outlier is left in the normal group. (Jugulum 2000)

3. Calculation of the Mahalanobis Distance for Each Sample in the Normal Group

Let $Y \in \text{Normal Group}$

$$E(Y) = \mu$$

$$V(Y) = \Sigma$$

The formula of Mahalanobis Distance of a sample X to a group with mean μ and covariance Σ is

$$MD^2 = \|X, Y\|^M = (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad [2.12]$$

Σ : the covariance matrix for the group.

Taguchi and Jugulum (2000) use an alternative notation in their work, by first standardizing the variables and then using the correlation matrix. This is possible, as the correlation matrix is the covariance matrix divided by the products of standard deviations.

Also, in order to obtain a mean close to 1 regardless of the number of decision variables, the MD is scaled by dividing by the number of decision variables, p .

In the first step the samples in the normal group are standardized using the mean \bar{x}_i and standard deviation σ_j for each variable.

$$Z_i = \left(\frac{x_{i1} - \bar{x}_1}{\sigma_1}, \dots, \frac{x_{ip} - \bar{x}_p}{\sigma_p} \right) \quad [2.13]$$

$\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$: Vector corresponding to sample $i, i=1, \dots, n$

\bar{x}_j : Mean of variable $j, j=1, \dots, p, \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$

σ_j : Sample standard deviation of variable j , $j=1,\dots,p$

Scaled MD for each sample from the normal group is computed using:

$$MD^2 = \frac{1}{p} Z^T Q^{-1} Z \quad [2.14]$$

Q = Correlation matrix of p variables.

For the sake of simplicity, scaled MD has been named as MD throughout this study.

Jugulum (2000) provides a proof that the scaled MDs of the normal group [2.14] has a mean of 1 and a variance of $2/p$, p being the number of variables. This proof is based on the fact that Mahalanobis Distances follow a chi square distribution with p degrees of freedom when the sample size n is large and all variables follow a normal distribution. Using this proof, Jugulum (2000) state the advantage of using scaled Mahalanobis Distance as being able to generalize the Mahalanobis Space to any number of variables with the expected mean MD being always 1. However, later on in his thesis, he is urged to comment on each case study where the average of MD is not one, by saying that the average MD "approach to unity".

Woodall et al. (2003) explain that the proof provided by Jugulum(2000) is actually an approximation based on chi-square distribution with p degrees of freedom. They also explain that "this is the probability distribution of $(1/p)MD$, provided that the sampling is from a multivariate normal distribution, the mean vector and variance-covariance matrix are assumed to be known and are used in the calculations instead of their estimates."

In the case where it is assumed that the samples are from a multivariate distribution but the estimates of the mean vector and the covariance matrix are used, Woodall et al. (2003) explain that the marginal distribution of MD_j is from a beta distribution and actually has a mean of $(n-1)/n$ instead of 1. The mean of MD_j is also $(n-1)/n$ if the n observations of the normal group represent the entire population of the normal items. The average MD value for the samples in the normal group is always exactly, $(n-1)/n$.

In their later response, Taguchi and Jugulum (2003) accepted that $(n-1)/n$ is the correct formula instead of unity. This formula can also be checked in the examples provided in their work (Jugulum 2000) where the mean MDs are in fact $(n-1)/n$.

The correlation matrix method of calculating MD has some shortcomings that arise in calculating the inverse of the correlation matrix. It is clear that in order to calculate the inverse of the correlation matrix, the correlation matrix has to be invertible. Two reasons for not having an invertible matrix are the multicollinearity and singularity between the variables, and the number of samples. These issues are detailed in Section 2.2

In their work, Taguchi and Jugulum (2000) suggested the use of Mahalanobis Taguchi Gram-Schmidt Method (MTGS) in order to overcome such issues. MTGS is explained in Section 0.

4. Removal of the Outliers from the Normal Group

After calculating the MD of the samples, the samples having higher MDs are found to be outliers should be removed from the normal group, and the MDs should be recalculated using the remaining samples in the normal group. This process is repeated until a suitable MS with no outliers is generated. There are different methods regarding outlier detection.

Mahalanobis Distance is actually a widely used method for multivariate outlier detection. In that context, Mahalanobis distance to the centroid follows a χ^2 distribution with degree of freedom equal to number of variables, assuming the samples are from a multivariate distribution. If Mahalanobis Distance is greater than $\chi^2(d.f.)$ at $\alpha=0.001$, then it can be considered to be a multivariate outlier. (Tabachnick and Fidell 2001)

Hirohisa et al. (2006) propose that the samples having MD larger than $1+2\sigma_{MD}$ (where σ_{MD} is the standard deviation of the MDs of the normal group) are to be considered as outliers and must be removed from the normal group. However, they also state that the $n>3p$ rule must be followed and new samples must be collected if required.

Russeeuw and van Zomeren (1990) also emphasize the n/p ratio in outlier detection. They state in their article that multivariate outliers can be masked where the ratio is low and that

there must be at least five observations per variable, that is $n/p > 5$ in order to check the outliers.

5. Gathering of r Samples from Outside the Normal Group

In order to check whether the model formed is able to determine whether a sample is from outside the normal group or not, samples that are known to be outside the normal group must be collected.

Woodall et al. (2003) found the explanation of this step unclear. They state that a more clear definition of whether the abnormal samples should be selected from a statistical population or from extreme cases is unclear.

In response, Jugulum and Taguchi (2003) explain that samples from different abnormality conditions must be collected and each sample outside the normal group is considered to be unique, not as a group.

From the case studies provided by Taguchi et al. (2001) it is seen that the authors consider each abnormal sample to be from a different experiment. This point was also found vague by Woodall et al. (2003); as in other cases provided, examination of abnormal with different levels of severities is shown. Woodall et al. (2003) also express that "Because of the lack of clarity regarding the fundamental issues of sampling in MTS, we find its implementation ill advised."

Taguchi and Jugulum (2000) explain that the abnormal samples should be collected in such a way that the MDs of the abnormal samples are "higher" than the normal samples, but the meaning behind the word "higher" is not clearly defined according to Woodall et al. (2003). If the MD of the abnormal samples should be greater than all the normal samples, this would limit the usefulness of MTS, and Type I and Type II errors should be taken into consideration. (Woodall, et al. 2003)

6. Calculation of the MDs from outside the normal group.

MD of each abnormal sample is calculated using the formula

$$MD^2 = \frac{1}{p} Z^T \Sigma^{-1} Z \quad [2.15]$$

where Z is the abnormal sample standardized using the means and standard deviations of the normal group. Σ is the correlation matrix of the normal group, calculated in Step 3.

7. Defining the Optimal System by Identifying Useful Variables

After forming the Mahalanobis Space, the model formed will be optimized in such a way that the discrimination between the normal and abnormal samples is the highest possible. As Mahalanobis Distance is a measurement scale that does not contain information about the direction, discrimination power is measured by the largeness of the MDs of the abnormal samples. Jugulum (2000) states that the MDs of the abnormal samples have to be significantly higher than the MDs of the normal samples. However, this has been criticized by Woodall et al. (2003) as there is no clear definition of how large the abnormal items' MDs have to be.

In order to select significantly useful variables, Taguchi's robust design methods are used. Orthogonal arrays and Signal to Noise (S/N) ratios are used in this step to examine the effects of the variables on the response. Although it is preferred that a full factorial design is used to examine all variable combinations, this usually requires a significant computational effort. The use of Orthogonal Arrays is suggested as a means to examine the effect of each variable with less number of runs.

Orthogonal arrays (OA) are used to examine the effects of forming models with different variable combinations. Each variable is assigned to a column of a two level orthogonal array where level 1 is to use a variable, and level 0, is not to use variable. A row in the orthogonal array represents a variable combination.

Table 1 L8(2⁷) Orthogonal Array

Run	X1	X2	X3	X4	X5	X6	X7
1	1	1	1	1	1	1	1
2	1	1	1	0	0	0	0
3	1	0	0	1	1	0	0
4	1	0	0	0	0	1	1
5	0	1	0	1	0	1	0
6	0	1	0	0	1	0	1
7	0	0	1	1	0	0	1
8	0	0	1	0	1	1	0

Table 1 shows an L8(2⁷) OA. There are 8 different variable combinations seen in the table. For example, the first row shows the case where all of the variables are included. In the second row, only the first 3 variables are included.

For each row of the orthogonal array, the previous steps of MTS are redone and the MDs of the abnormal samples are recalculated. MDs of the normal samples do not have to be calculated again, as only the abnormal samples are used in calculating the Signal to Noise ratios.

After calculating the MDs for the abnormal samples using different combinations of the variables, S/N ratios are calculated for each row of the OA. Signal to Noise ratios (S/N) are used both to determine which variables are useful in discriminating and to evaluate the performance of the constructed scale.

Taguchi and Jugulum (2000) define S/N as "the measure of accuracy of the measurement scale for predicting the abnormal conditions in decibel (dB) units". As the error decrease, S/N ratio does increase.

Using the calculated S/N ratios, the gain obtained from each variable is calculated. The gain is the difference between the average S/N ratio of the runs that include the variable and the average S/N ratio of the runs that do not include the variable. If including a variable increases the S/N ratio, that variable is included into the model.

Among different S/N ratios, the use of dynamic S/N ratio is suggested for it would give more accurate results. (Jugulum 2000) However, dynamic S/N requires prior knowledge about the severity of the abnormality – the actual distance of the abnormal samples to the normal group. In many cases, it is not possible to provide this information. In cases where the response is binary or ordered categorical (where the response is defective/not defective etc.), it is not possible to determine the actual distance of the samples to the normal group. Taguchi and Jugulum (2000) state that larger-the-better type can also be used in such situations.

Although the use of OA's and S/N ratios are suggested, there are some reported disadvantages of the use of OA's and S/N ratios to determine important variables.

The first drawback is that the use of OA's does not always give the optimal combination of variables. Evaluating all possible combinations of variables is a more guaranteed way. For the medical case provided by Taguchi and Jugulum (2000), it was shown that by evaluating all possible combinations of variables, a better solution in the sense of S/N ratio was obtained. (Woodall, et al. 2003)

If the number of variables is large, it may actually not be computationally feasible to try all possible combinations of variables. For example, if 50 variables are used, the number of runs required to try all possible combinations would be $2^{50} = 1,1259E+15$. As an alternative, Abraham and Variyath (2003) proposed a forward selection procedure where the variables are added one by one. The first variable is selected by looking at each single variable and taking the one that has the maximum S/N ratio. The second variable is the one that gives the maximum S/N ratio together with the previously selected variable. The process continues until further improvement is not possible. If the number of variables is needed to be further reduced, variables with very small S/N gains may also be removed from the model.

The second and the major drawback come from the way variables are allocated to orthogonal arrays. When the variables are not as many as to fill all the columns of an OA, they are assigned to the first k columns. This way of use comes from the assumption that there are no interactions between the variables, and that the variables can be allocated in any possible way. However, it was shown that changing the order the variables are

allocated to the OAs do change the selection of useful variables. (Jugulum 2000) When the variables are allocated to different columns, the resulting main effects change. The selection of items are done using the main effects without any statistical tests, As a result, the selection of useful variables differ when dynamic S/N is used, but they remain the same when using larger-the-better S/N ratio. (Abraham and Variyath 2003)

Larger-the-Better (S/N) Ratio

If the actual distance of the abnormal samples to the normal group is not known or they cannot be separated into severity groups, the combination of variables can be chosen in such a way that the discrimination is the highest. This could be ensured by obtaining MDs of the abnormal samples that are as large as possible.

Let $MD_j, j=1, \dots, t$ be the MD values for the abnormal samples. The larger-the-better S/N ratio is calculated by:

$$S/N \text{ Ratio} = -10 \log \left[\left(\frac{1}{t} \right) \sum_{j=1}^t \left(\frac{1}{MD_j} \right)^2 \right] \quad [2.16]$$

2.2. Handling Small and Imbalanced Data in MTS

It was explained in Section 2.1.2 that one drawback of the MD calculation in MTS arises in the calculation of the inverse correlation matrix present in the distance formula.

$$MD^2 = (\mathbf{x} - \mathbf{y})^T \Sigma^{-1} (\mathbf{x} - \mathbf{y}) \quad [2.17]$$

Calculation of the Mahalanobis Distance requires taking the inverse of the covariance matrix. If the covariance matrix is not invertible, Mahalanobis Distance cannot be

calculated. An $n \times n$ matrix is invertible only if it is a full rank matrix, that is, it is square and the rank of the matrix is equal to n .

There are two reasons for not having an invertible covariance matrix in MD calculation. The first reason can be the high correlation between the variables. High correlation results in multicollinearity or singularity.

The second reason is insufficient number of samples in forming the Mahalanobis Space, where the number of samples is less than the number of variables. This was examined in Section 2.1.2.

2.2.1. Multicollinearity and Singularity

High correlation between the variables causes multicollinearity or singularity. Multicollinearity is present when the variables are very highly correlated (>0.90). Multicollinearity often occurs when cross-products or powers of variables are included in the analysis together with the original variables.

Singularity, on the other hand, is seen if one of the variables is a combination of two or more of the other variables. Such variables are actually redundant variables.

Thus, if there is multicollinearity or singularity between the samples forming the normal group, there are actually fewer variables than it appears, and the correlation matrix is not of full rank because there are not actually as many variables as columns.

Either bivariate or multivariate correlations can create multicollinearity or singularity. If a bivariate correlation is too high, it can be seen in the correlation matrix as a correlation above 0.90. Deleting the redundant variables will solve the problem. However, if multivariate correlations are high, multivariate statistics have to be used to find the variables that cause the problem. (Tabachnick and Fidell 2001)

Inverse of a matrix can be calculated using the following formula:

$$A^{-1} = \frac{1}{\det A} A_{\text{adj}} \quad [2.18]$$

A_{adj} = adjoint matrix of matrix A

$\det A$ = determinant of matrix A

In the singularity case, the determinant is zero, and thus the inverse matrix cannot be calculated.

In multicollinearity, the determinant is not exactly zero, but it is very close to zero. The matrix is invertible, however the calculations become unstable. Division by a near-zero determinant produces very large and unstable numbers in the inverted matrix. The sizes of numbers in the inverse matrix fluctuate wildly with only minor changes in the sizes of the correlations in A. Use of an unstable inverse matrix in calculations will also make the steps following inverse calculation unstable. (Tabachnick and Fidell 2001).

In multivariate analysis, several methods are proposed to overcome singularity and multicollinearity. For singularity, redundant variables can be removed to solve singularity. For multicollinearity, these methods include transformation, removal of univariate outliers, and removal of multivariate outliers. (Tabachnick and Fidell 2001)

Univariate outliers may be determined by considering a range of 3 standard deviations from the variable mean. If the sample is more than 3 standard deviations from the mean in a variable, it is considered to be a univariate outlier.

For multivariate outliers, Tabachnick and Fidell (2001) suggest the use of Mahalanobis Distance. If a sample is from a multivariate normal distribution, then its MD to the centroid of the normal group follows a χ^2 distribution with degree of freedom equal to number of variables. If Mahalanobis Distance is greater than $\chi^2(\text{d.f.})$ at $\alpha=0.001$, then it is considered to be a multivariate outlier. A chi-square random variable with p degrees of freedom is defined as the sum of squares of p independent standard normal random variables.

(Tabachnick and Fidell 2001) This method, however, can be difficult to apply if the Mahalanobis Distance cannot be calculated at all.

Two solutions were proposed by Jugulum (2000) for the inversion problem. The first solution is the use of orthogonal vectors instead of the original variables via Mahalanobis Taguchi Gram-Schmidt Method (MTGS). The second solution is the use of adjoint matrix instead of the inverse matrix. These are explained in Appendices IV and V respectively.

Aside from solutions provided by Jugulum (2000), it is also possible to use generalized inverses in cases where matrix inversion is not possible. Generalized inverses are explained in Appendix D.

2.3. Robust Design

Robust design is a method developed for design optimization of the products or processes so that the desired outputs are obtained. Robust design aims to make the performance of the product/process insensitive to variations in the production process and the variations present in the materials that are used in the production process. (Phadke 1989)

In robust design, a designed experiment is set up to study the effects of the decision variables on the outputs. For this, orthogonal arrays are used to reduce the number of experiments required. To measure the quality of the output, Signal-to-Noise ratios are used.

A product/process can be represented using the Parameter Diagram shown in Figure 1. As shown in the figure, three different types of parameters influence the response of the product.

The response of the product is denoted by Y . This response, or the output of the product/process function, can be any suitable quality characteristic such as defect level.

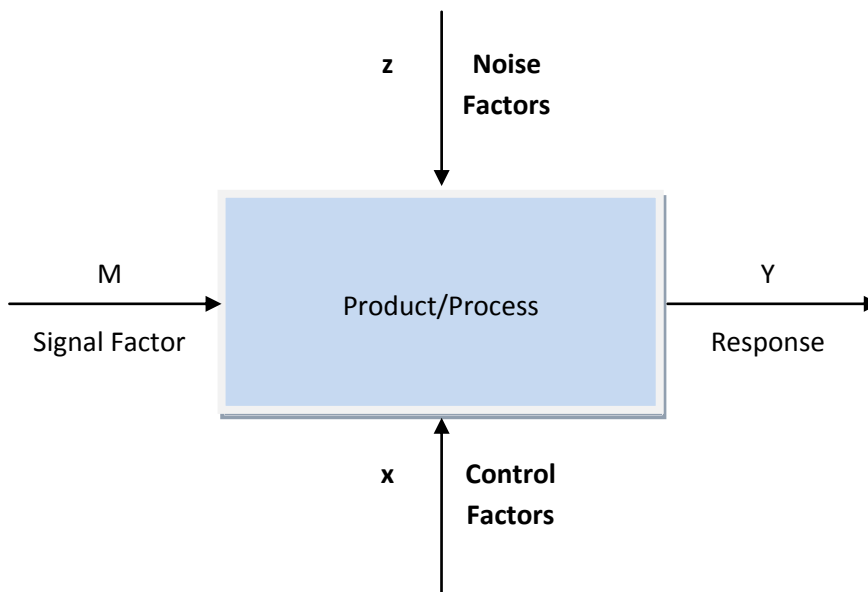


Figure 1 P Diagram

Phadke (1989) defines these factors as follows:

Signal factors (M) are the parameters set by the user or operator of the product to express the intended value for the response of the product. For example, the speed setting on a table fan is a signal factor with resulting in 3-4 different types of output. The signal factors can be one or more, depending on the problem. If the signal factor takes a constant value, such problems are called static problems. Otherwise, the problem is called a dynamic problem.

Noise factors (z) are the parameters that cannot be controlled. The factors that are difficult or expensive to control are also taken as noise factors. As there is no control on the noise factors, they differ in each case, causing the response to deviate from the aimed target.

Noise factors can be related to the environment (Temperature, humidity etc.) or the production process /input material properties.

Control factors (x) are the parameters that can be determined by the decision maker. Each control factor can take different values, or levels. The aim is to determine the best values of these parameters.

In each problem studied, it is important to identify the signal factors, noise factors, control factors and the response. As the relationship between these factors and the response is not known at the beginning, they are observed in a designed experiment.

A designed experiment is a set of experiments where different settings of control factors are examined to understand their effects on the response. Each experiment is called a run or a treatment. Orthogonal arrays are commonly used to set up experiments efficiently with low number of runs.

Table 2 shows an example with four factors and their chosen levels. The starting levels (levels before conducting the experiment) are identified by the subscript 0 in the table. The others are the alternate levels that are tried in the designed experiment.

Table 2 Factors and their levels

Factor	Levels		
	1	2	3
Factor 1	$F1_0-50$	$F1_0$	$F1_0+50$
Factor 2	$F2_0$	$F2_0+5$	$F2_0+10$
Factor 3	$F3_0-30$	$F3_0$	$F3_0+30$
Factor 4	$F4_0$	$F4_0+10$	$F4_0+20$

Table 3 is a designed experiment with nine separate experiments. The rows are the experiments, and the columns show the factors and their levels in consideration. The matrix experiment consists of nine individual experiments corresponding to the nine rows. The estimation of the factor effects is found using the response.

Table 3 Designed experiment

Experiment No	Column Number and Factor Assigned				Response
	Factor 1	Factor 2	Factor 3	Factor 4	
1	1	1	1	1	Y_1
2	1	2	2	2	Y_2
3	1	3	3	3	Y_3
4	2	1	2	3	Y_4
5	2	2	3	1	Y_5
6	2	3	1	2	Y_6
7	3	1	3	2	Y_7
8	3	2	1	3	Y_8
9	3	3	2	1	Y_9

Steps in Robust Design

Having defined the basics of robust design, robust design method can be divided into its steps. These are given by Phadke (1989) as follows:

Planning the experiment

1. Identify the main function, side effects, and failure modes

In this step, the production process is tried to be understood. The production process and the problems associated with it are examined to determine what to focus on.

2. Identify noise factors and the testing conditions for evaluating the quality loss

The factors that cannot be controlled in the production process are identified and the testing conditions that can be used to measure the importance of the noise factors are determined.

3. Identify the quality characteristic to be observed and the objective function to be optimized

The response of the production function that can be measured is determined in this step. Also, the objective function, which is a function of the response measured is determined.

4. Identify the control factors and their alternate levels.

The control factors that can be changed to reach the desired response are identified. Their alternate levels that can be observed in the designed experiment are determined.

5. Design the matrix experiment and define the data analysis procedure.

The matrix experiment (or the designed experiment) is formed. The data analysis procedure, defining how the results of the matrix experiment will be analyzed is formed. Data analysis usually consists of determining the main effects of the variables and performing Analysis of Variance

Performing the experiment

6. Conduct the designed experiment

The experiments are conducted by examining the different combinations of control factors defined in the runs of the designed experiment.

Analyzing and verifying the experiment results

7. Analyze the data, determine optimum levels for the control factors, and predict performance under these levels

The results of the experiments are analyzed, and the effects of the control factors on the response are determined. Using this information, the optimum levels for the control factors are determined.

8. Conduct the verification experiment and plan future actions

The optimal levels of the control factors are tried in a verification experiment to understand whether the results obtained are valid

Mahalanobis Taguchi System can be used in robust design as a mean to understand the function between the control factors and the response.

2.4. Threshold Determination

In Mahalanobis Taguchi System, it is important to decide on a Mahalanobis Distance value that would act as a threshold, separating normal and abnormal groups. Taguchi and Jugulum (2000) state that MD of the abnormal samples is "higher" than the normal samples. However, no clear definition for the word "higher" was provided. The case studies provided in Taguchi and Jugulum (2000) and in Taguchi et al. (2001) portray abnormal samples having very large MDs. However, Woodall et al. (2003) express that such a wide and clear distinction between the normal and abnormal groups is not always possible, contrary to the case studies provided. They state that if the expectation is that MDs of the abnormal samples are greater than all of the normals in all cases, usefulness of MTS would be limited as it is often the case that two classes overlap. Jugulum et al. (2003) answer this criticism by explaining that the statement "higher than" is a relative concept. "Higher than" is relative to the severity of the abnormality and, more important, they state, to the cost of savings due to correctly classifying an observation to be abnormal. They also suggest that selecting a threshold value that balances two types of mistakes is preferable in quality engineering. In cases where the discrimination between the normal and the abnormal samples is not clear, Taguchi et al. (2001) suggest the use of loss function in order to balance misclassification of the members of the two classes. They explain the use of loss function in quality assurance context in the following way:

The samples can be examined using MTS method and be classified as defective or non-defective. Then the samples that are found defective can be inspected to see if they are actually defective or not. Threshold value used to determine whether a sample should be classified as defective and put into a complete examination, denoted by D , is calculated as follows:

$$D = \sqrt{\frac{A}{A_0}} T^* \quad [2.19]$$

A = The cost of completely examining a sample

A_0 = The loss caused by not examining a sample throughout and the sample being abnormal.

T^* = Mean MD of the believed to be abnormal group.

Woodall et al. (2003) argue that the above formula is not clear as choice of the group believed to be abnormal is unclear.

In their later work, Taguchi et al. (2006) suggested the use of adjoint matrix of the correlation matrix instead of the inverse. Although this method is proposed to solve multicollinearity problems, the scale is different from the unit scale as the determinant in the denominator is removed from the inverse formulation to obtain the adjoint matrix. Thus, pronouncing a threshold value that could be used in general is impossible.

Although Taguchi and his colleagues do not take distribution assumptions into consideration, it was elaborated previously that Mahalanobis Distance is based on multivariate normal distribution (Dasgupta 1993). As Mahalanobis Distance is used to detect multivariate outliers, Mahalanobis distance of a sample from the centroid of a known group follows a χ^2 distribution with d.f. equal to number of variables. If Mahalanobis Distance is greater than $\chi^2(d.f.)$ at $\alpha=0.001$, then it is a multivariate outlier (Tabachnick and Fidell 2001). This method can also be used to determine the cut-off point where the samples above will be regarded as abnormal.

In their article, Itagaki et al. (2007) used $\overline{MD} + 2\sigma$ as a threshold with the assumption of normality for the MD of the normal samples. They state that the probability of normal samples being between $\overline{MD} \pm 2\sigma$ is %95.44.

Su and Hsiao (2007) propose Probabilistic Thresholding Method (PTM) as an alternative to Exhaustive Search Method. Exhaustive search is a simple but time consuming method where all possible solutions are systematically tried. Su and Hsiao state that "[...] real applications always use the "exhaustive search method" (ESM) to search the threshold that results in the highest classification accuracy on the training set". The computational effort of searching through a range of thresholds and the probability of overfitting the model to the training set are two underlined disadvantages of Exhaustive Search Methods. Chebyshev's inequality is used to develop PTM where Chebyshev's inequality provides an upper limit on the probability of getting a value larger than mean plus a multiple of standard deviation.

Chebyshev's inequality provides the probability of getting a value that is more than a multiple of standard deviation from the mean.

$$P(\mu_x + r\sigma_x \leq X) \leq \frac{1}{r^2} \quad [2.20]$$

where μ_x and σ_x are the mean and the standard deviation of the random variable X . The probability of getting a value at more than r standard deviations from the mean is at most $\frac{1}{r^2}$

Su and Hsiao use this theorem on the MDs of the normal group. Regarding the MD of a normal sample as a random variable X_{md} coming from a probability distribution with μ_{md} and σ_{md} as the mean and the standard deviation, probability of obtaining a normal MD larger than the threshold is $P(T \leq X_{md})$. Using this probability, and the overlap percentage of the normal and abnormal samples in the training set, the r value and the threshold are determined. This method can be detailed as follows:

1. Removal of the outliers from the abnormal group

Samples with MDs less than three standard deviations from the mean of the MDs of the normal group are removed for being outliers.

2. Compute the percent of non-overlapping normal samples

The percent of samples with MDs smaller than the minimum MD of the abnormal samples are calculated, and denoted by w . As some of the abnormal samples are removed for being outliers, a small parameter λ ($w > \lambda > 0$) is created, in order to adjust w to the actual overlapping percentage. Thus, under the MD boundary covering $w - \lambda$ of the normal distribution, the abnormal samples are predicted perfectly. Maximum false alarm percentage is $1 - (w - \lambda)$ for the normal group.

As a result, the upper bound of the false alarm probability $P(T \leq X_{md}) \leq 1 - (w - \lambda)$.

3. Apply Chebyshev's theorem

If the mean μ_{md} and σ_{md} are known, and w and λ are known,

Let $1-(w-\lambda)=\frac{1}{r^2}$, and $r = \sqrt{\frac{1}{1-(w-\lambda)}}$

Thus,

$$T = \mu_{md} + \sigma_{md} \sqrt{\frac{1}{1-(w-\lambda)}}$$

2.5. Performance Measures

In order to measure how well a machine learning method works, different performance measures have been defined. In this section performance measures that will be used to assess the effectiveness of MTS are defined.

Table 3 below is the confusion matrix for the two class classification problem. The rows show the actual classes, and the columns show the predicted classes. In machine learning literature, the minority class is usually labeled positive (defective), and the majority class is labeled negative (non-defective). True positives are the positive samples classified to be positive. True negative, on the other hand, are the samples that are negative and are classified to be negative. True Positive and True Negative values show the samples that are classified correctly. False Positive samples are the samples that are actually negative but have been classified to be in the positive class. False Negative samples are the samples that are actually positive but are classified to be in the negative class.

Table 4 Confusion Matrix

	Predicted Positive	Predicted Negative
Actual Positive	True Positive	False Negative
Actual Negative	False Positive	True Negative

From the values in the confusion matrix, different performance measures can be calculated:

Precision: Precision is the ratio of True Positives to all samples predicted to be positive. Precision is the correct classification rate for the samples that are classified to be positive.

$$\text{Precision} = \frac{TP}{TP + FP} \quad [3.1]$$

Recall: Recall (or Sensitivity) is the ratio of True Positives to all actual positives. Recall is the classification rate of the positive samples.

$$\text{Recall} = \frac{TP}{TP + FN} \quad [3.2]$$

Specificity: Specificity is the ratio of True Negatives to all actual negatives. Specificity is the classification rate of the negative samples.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad [3.3]$$

G-mean: G-mean is the geometric mean of recall and specificity

$$G - \text{mean} = \sqrt{\text{Specificity} \text{Recall}} \quad [3.4]$$

F-Measure: F-Measure is an accuracy measure, combining precision and recall.

β in the formula is the parameter used to give different weights to precision or recall. Most commonly used β values are 0.5, 1, and 2. $\beta=0.5$ weights precision twice as much as recall, while $\beta=2$ weights recall twice as much as precision. $\beta=1$ gives equal weights to precision and recall. These will be designated as F0.5, F1, and F2 in the sequel.

$$F - \text{measure} = \frac{(1+\beta^2)(\text{Precision} * \text{Recall})}{(\beta^2 * \text{Precision} + \text{Recall})} \quad [3.5]$$

Accuracy: Accuracy (or Percent Correctly Classified), is the ratio of True Positives and True Negatives to all samples. It shows the correct classification rate for the classifier.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad [3.6]$$

Kappa: (Cohen's) Kappa is a measure of the agreement between the two different methods of classification. In this case, Kappa measures the agreement between the actual classes and the predicted classes.

$$\theta_1 = \frac{TP + TN}{N}$$

$$\theta_2 = \left(\frac{TP + FN}{N}\right) \left(\frac{TP + FP}{N}\right) + \left(\frac{TN + FN}{N}\right) \left(\frac{FP + TN}{N}\right)$$

$$\text{Kappa} = \frac{\theta_1 - \theta_2}{1 - \theta_1} \quad [3.7]$$

In imbalanced datasets, accuracy can be a misleading performance measure. Even if all of the minority class samples are misclassified, the correct classification rate would still be high as a result of the imbalance. For this reason it would be more appropriate to measure the separate classification rates of both classes. Thus it can be said that in imbalance datasets, the main aim would be to improve the classification rate of the minority class (recall) without decreasing the classification rate of the majority class (precision). These two goals often cannot be achieved together, as precision and recall affects each other negatively. For this reason, F-measure is often used as a single measure combining precision and recall. If more importance is given to one class than the other, β can be changed to give different weights to precision and recall. (Chawla 2005)

Receiver Operating Characteristic (ROC) curve have been widely used in summarizing the classifier performance in imbalanced datasets. In the ROC curve, True Positive Rate (Sensitivity) is plotted against False Positive Rate (1-Specificity). ROC curve displays the performance of the classifier over a range of trade-offs between the TPR and FPR by changing the threshold. The point (0,1) is the best point in the graph, where all the positive samples are correctly classified, and none of the negative samples are incorrectly classified. The line $y=x$ shows the case where classification is made randomly. A point on the graph can be chosen considering the trade-offs between true positive rate and false positive rate. (Bradley 1997).

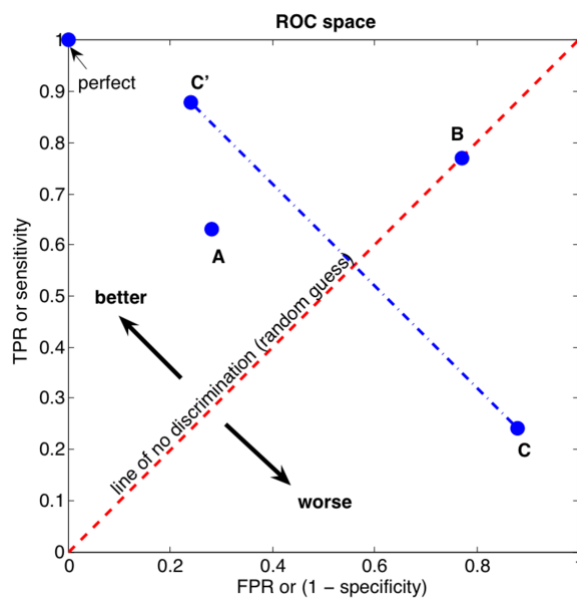


Figure 2 ROC curve (Bradley 1997)

Area Under ROC Curve (AUC) is a measure that represents how well the model classifies considering all the threshold values. It is a commonly used way to measure the classifier performance of a method. As it does not depend on the decision threshold chosen, it is

commonly used to compare the different classification methods. It is a preferred method when a single performance measure indicator is required or when a threshold has not been determined. (Bradley 1997)

2.6. Resampling Methods for Small and Unbalanced Data in MTS Classification

A data set is called imbalanced if one class has more examples than the other classes. In imbalanced data sets, the class having more examples is called the majority classes and the class having fewer examples is called the minority class. The use of machine learning methods in quality problems aims to predict the defectives, which usually make up the minority class. The main question when dealing with imbalanced datasets is whether the imbalance between the classes affects the prediction performance or not.

There has been a lot of research done on the effects of class imbalancedness on machine learning methods. Contrary to the idea that the natural distribution of the dataset should be kept while modeling the classification model, Weiss and Foster (2001) show that if the dataset is imbalanced, the natural distribution is not the best choice for modeling. According to their work, correct classification of the minority class is much lower than that of the majority class. By modifying the class distribution, it has been shown that correct classification rate and AUC improved.

In MTS, imbalance in the dataset has different effects. It was explained previously that in order to form a model from the normal group, normal group must have as many samples as the number of variables in order for the correlation matrix to be nonsingular. However, research shows that although this is a minimal requirement for calculating the inverse of the correlation matrix, there is also a lower limit to the number of samples needed to "characterize a multivariate Gaussian process with n samples" (Foley 1972) (Young 1978) This issue is one of the reasons that oversampling in the normal sample is desired. In addition, Jugulum (2000) states that as the sample size increases, the variance of the MD in the normal group approaches to one, and for this reason it is desired to have as many normal samples as possible

A resampling method was as a joint study with Berna Bakır and Dilber Ayhan. This method was proposed in order to study the effects of class imbalance in classification have been developed. The method is based on SMOTE and the aim is to study the effects of class ratio and sample size on MTS classification performance. This approach generates datasets with different ratios and sample sizes, using the original dataset. Oversampling is done using SMOTE method. The details of this study can be found in Ayhan (2009).

2.6.1. SMOTE

Chawla et al. (2002) developed an oversampling approach for imbalanced datasets. In this method new "synthetic" samples from the samples in the minority class are used to oversample the minority class. Synthetic samples are formed along the line segments joining any/all of the k minority class nearest neighbors. The parameter k is chosen according to the number of synthetic samples to be created. If the number of synthetic samples will be twice the number of minority class, two neighbors from the k nearest neighbors are chosen randomly.

Synthetic samples are generated using the pseudo code provided by Chawla et al. (2002):

```
Dataset : The data that will be oversampled
T : The number of samples in the dataset
N : Percent of oversampling
k : Number of neighbors that will be used in oversampling

numattrs : the number of attributes

N=N/100

for each sample in the Dataset
Find k nearest neighbors for each sample
    for u=1:N
        Select a random neighbor from the determined k neighbors
            for attr=1:numattrs
```

```

    dif = (Attribute value of the neighbor)-(Attribute value
of the sample)

    gap = random number between 0 and 1

    synthetic_sample_attribute = (Attribute value of the
sample) + gap*dif

end

end

```

Figure 3 Pseudocode of SMOTE (Chawla et al.(2002))

2.7. Methods for Robust Design with Qualitative Responses

In quality experiments it is usually desired that the response is quantitative, as a continuous, measurable response would allow better evaluation of the effects of the parameters. However, obtaining a quantitative response may not be possible because of the technical or economical difficulties in accurate measurement of a continuous response.

Qualitative, or categorical, variables can be binary (defective/non-defective), ordered categorical (low defect/mild defect/ high defect) or nominal (no ordering of the response). Different methods have been developed for modeling and optimizing data with qualitative response.

2.7.1. Logistic Regression

Logistic Regression is a type of generalized linear models commonly used for predicting a discrete response. The independent variables used in Logistic Regression can be binary, ordered categorical or nominal. Although there are applications for ordered categorical and nominal cases, the dependent variable in logistic regression is usually binary where the dependent variable takes the value of 1 with a probability p , or the value of 0 with a probability $1-p$. (Tabachnick and Fidell, 2001) ; (Erdural,2006)

In logistic regression, the relations between the independent variables and the dependent variable are described by the logistic function:

$$f(z) = \frac{e^z}{1+e^z} \quad [4.1]$$

Using the logistic function, the dependent variable takes a value between 0 and 1, regardless of the values of the independent variables.

The probability of being a member of class 1, $P(Y=1)$ is calculated using the formula

$$E(Y) = P(Y = 1) = \frac{e^z}{1+e^z} \quad [4.2]$$

where the variable z is

$$z = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_jx_j$$

β_0 is the constant, and β_i is the coefficient of the independent variable x_i

One advantage of logistic regression is that there is no assumption about the distribution of the independent variables.

2.7.2. Accumulation Analysis

Accumulation Analysis (AA) is a method introduced by Taguchi (1991) for the analysis of ordered categorical data. In AA, the cumulative response categories are determined. Cumulative response categories are separate response ranges (For example, $Y < 5, Y < 10$ etc.). The effects of factor levels on the cumulative probability distributions by the categories are calculated using an ANOVA-like procedure. Using this information, the optimal factor levels required for a desired cumulative category is calculated. (Erdural 2006) (Nair 1986) (Wu and Yeh 2006)

Consider a one-factor experiment with factor A with I levels. For each level of A , there are n observations. These observations are separated into K ordered categories.

Let us denote cumulative frequencies by

$$C_{ik} = \sum_{j=1}^k Y_{ij} \quad [4.3]$$

where Y_{ik} is the observed cumulative frequency of level i in category k , $k=1, \dots, K$ and $i= 1, \dots, I$.

The averages of the cumulative frequencies across the factor levels are

$$C_k = \sum_{i=1}^I \frac{C_{ik}}{I} \quad [4.4]$$

The sum of squares for factor A is calculated as

$$SS_A = n \sum_{k=1}^{K-1} \sum_{i=1}^I \frac{(C_{ik} - C_k)^2}{[C_k(n - C_k)]} \quad [4.5]$$

and the sum of squares for error term is

$$SS_e = n \sum_{k=1}^{K-1} \sum_{i=1}^I \frac{[C_{ik}(n - C_{ik})]}{[C_k(n - C_k)]}$$

The expectation of SS_e is $[n(n - 1)I^2(K - 1)/(In - 1)] \approx I(n - 1)(K - 1)$. As a result of this, Taguchi uses $I(n - 1)(K - 1)$ as degree of freedom for SS_e and using the statistic

$$F_A = \frac{MS_A}{MS_e}$$

to test the effect of factor A.

Nair (1986) and Box and Jones (1986) stated that in AA, SS_e depends on the effect of the other factors. Therefore, the sum of squares are not independent as it is in ANOVA. Similarly, it was also noted that the effects of the factors are also dependent to other factors.

Accumulation Analysis have been widely criticized by Nair (1986) and Hamada and Wu (1986) for being inefficient and unnecessarily complicated. They suggest that scoring methods are simpler and more efficient for ordered categorical data.

2.7.3. Scoring Schemes

Nair (1986) presented two scoring schemes to identify the location and dispersion effects separately. An investigation recommended using the mean square to identify a prominent effect. The optimal factor settings of location and dispersion effects can be obtained by the contribution of each control factor on both effects. The final optimal factor settings are obtained by adjusting between the location and dispersion effects.

Hamada and Wu (1986) explained the scoring methods using a case with J categories. Let q_j be the overall proportion of observations in category j , $j=1,2,\dots,J$. The mid-rank of category j , τ_j is defined as

$$\tau_j = \sum_{k=1}^{j-1} q_k + q_j/2$$

So that τ_j is proportional to the mid-rank for category j .

Location scores are calculated as

$$I_j = \tilde{\tau}_j / \left[\sum_{j=1}^J q_j \tilde{\tau}_j^2 \right]^{1/2} \quad j=1,2,\dots,J$$

where $\tilde{\tau}_j = \tau_j - \sum_{j=1}^J q_j \tau_j = \tau_j - 0.5$

And the dispersion scores are calculated as

$$d_j = e_j / \left[\sum_{j=1}^J q_j e_j^2 \right]^{1/2} \quad j=1,2,\dots,J$$

where

$$e_j = I_j \left(I_j - \sum_{j=1}^J q_j I_j^3 \right) - 1$$

Using the location and the dispersion scores, ordered categorical data is analyzed. For the i th run, frequencies of each category are multiplied by these scores and summed for each category to obtain location/dispersion pseudo-observations, L_i and D_i

$$L_i = \sum_{j=1}^J f_{ij} l_j, i = 1, 2, \dots, n$$

$$D_i = \sum_{j=1}^J f_{ij} d_j, i = 1, 2, \dots, n$$

where f_{ij} is the frequency of category j at the i th run.

Location sum of squares for factor A at I levels is calculated by

$$SSA(I) = \sum_{i=1}^I L_i^2 / n_i$$

Dispersion sum of squares is calculated by

$$SSA(I) = \sum_{i=1}^I D_i^2 / n_i$$

where n_i is the total number of observations at the i th level of factor A.

To determine the factor level that results in the minimum dispersion, Nair calculated the mean of dispersion pseudo-observations for each level of that factor and selected the level with the minimum dispersion pseudo-observation as the optimal.

2.7.4. Weighted Signal-to-Noise Ratio

Weighted Signal-to-Noise Ratio (WSNR) is a method developed by Taguchi. In this method, a weight proportional to the quality loss is assigned to each response category. The categorical response is used as it is continuous and smaller-the-better SNR is used. The optimal parameter levels are the ones that results in the maximum SNR.

For each response category, three weights are determined: lower bound, middle bound, and upper bound. If there are J categories and f_{ij} , $j=1,2,\dots,J$ are the observed frequency of the jth category at the ith run, the smaller-the-better SNR is calculated using the following formula:

$$n_i = -10 * \log \left(\frac{1}{n} \sum_{j=1}^J w_j^2 f_{ij} \right)$$

Using the SNR, factor effects are calculated, and the factor levels that result in the highest SNR are chosen as optimal levels.

CHAPTER 3

PROPOSED THRESHOLD DETERMINATION METHOD

As the threshold determination in MTS is not an extensively studied area, it is proposed that Exhaustive Search Method (ESM) is used to determine a threshold where all possible thresholds are systematically tried. It is clear that a single, static threshold value cannot be used for every MTS model. A new threshold must be determined in each case. This threshold should be chosen depending on the desired classification properties. It is proposed that the threshold value is chosen so that the selected performance measure, such as correct classification rate, reaches its best value. In this method, a performance measure is selected, and the range of thresholds is searched incrementally to determine the threshold that maximizes the selected performance measure. Depending upon the importance given to a performance measure, the determined threshold will change. For example, it may be desired to determine defective samples (abnormal samples) in a production environment to the best degree for the classification of non-defective samples (normal samples) defective can be much less costly than classifying defective samples non-defective. In this case, a conservative threshold can be chosen by aiming the maximization of the classification rate of the defective samples.

In majority of the MTS cases found in literature, the dataset is used as a whole, and is not separated into training and testing groups. However, using the proposed method, this would cause overestimation of the performance measures as the threshold is chosen so that the performance measure is maximized. In our case, the data is separated into 3 groups using stratified cross validation. The threshold is determined on the training set and the obtained threshold value is applied to the testing set.

For this study, 3 different well known datasets are selected from UCI Machine Learning repository (2009). Used datasets are Pima Indian Diabetes, Wisconsin Breast Cancer and Magic Gamma Telescope. Table 5 lists some information about the used datasets.

Table 5 Dataset Information

Dataset	# of Classes	Total Items	# of attributes	# of class 0 (-)	# of class 1 (+)	Ratio of Minority Class to All Dataset
Pima Indian Diabetes	2	768	8	500	268	0.35
Wisconsin Breast Cancer (Diagnostic)	2	569	30	357	212	0.37
Magic Gamma Telescope	2	19020	10	12332	6688	0.35

It can be seen that all three datasets have two classes with approximate minority ratios of 0.35. In these cases, it would be difficult to label these datasets as imbalanced. In these datasets, the threshold that maximizes the selected performance measure is selected by searching through a threshold range 0-10. As a result of the properties of the Mahalanobis Distance detailed in Section 2, it is very uncommon for a sample in the normal group to take a MD value larger than 10. Thus, maximum threshold to be tried was set to be 10.

The threshold obtained from the training set is tested on the testing set and the performance measures of PCC, F05, F1, F2, and G-mean (See Section 2.5) are calculated. For each performance measure, the threshold is found by selecting the one that maximizes the selected performance measure.

In addition, two more methods have been used.

Euclidean: It was explained that the best point in the ROC graph is the (0,1) point where the false classification rate is 0 and correct classification rate is 1. As explained in Section 2.5, ROC graph shows the tradeoffs between correct and false classification rates, where each point is the corresponding correct and false classification rates for the selected threshold. This method tries to find the threshold that minimizes the Euclidean distance of the point

corresponding to the threshold point in the ROC curve to the best point in the ROC graph, where correct classification rate is 1 and false classification rate is 0.

Mean + 3 Standard Deviation: This method sets the threshold to 3 standard deviations of MDs of the normal group away from the mean MDs of the normal group. Aside from the fact that this distance is used commonly to detect outliers, it was also suggested by Watabe (2005)

PTM: Probabilistic Threshold Method (PTM) was developed by Su and Hsiao (2007). The threshold is determined using PTM method.

The methods are experimented on 3 replications where the groups are reformed randomly stratified in each replication. Using 3-fold stratified cross validation in 3 replications, 9 different results from the test groups for each performance measure are obtained. The performance measures explained in Section 2.5 were used to assess the performance of the models. The results below show the averages of these results for each threshold method and dataset.

The proposed threshold determination method can be listed as follows:

1. Separate the data into training and testing sets

In classification studies, the dataset is usually divided into training and testing sets. The model is formed using the training set and tested on the testing set. In this case, n-fold stratified cross validation is used, n changing depending on the sample size. Also, in order to overcome the effects of the random assignments of the samples to the groups, whole process can be examined in different number of replications. In this work, 3 replications are prepared.

2. Form the MTS model

The MTS model is formed using the training sets. As the response of the process, the MDs of each sample are obtained as a result.

3. Select the performance measure to be maximized

Depending on the classification performance that is most important, a performance measure is chosen. For example, if overall classification rate of the samples is important, PCC is chosen to be the important performance measure. As another example, if detecting the defective samples is important, the classification rate of the abnormal class (recall) can be chosen.

4. Try all possible threshold values on the calculated MDs. Calculate the performance measures.

Starting from 0, threshold values are tried one by one incrementally on the MDs of the training set. For each threshold, the selected performance measure is calculated.

As there is a limit to the MDs normal samples can take (defined by the mean and the variance of the MDs of the normal group), the trial can be continued until a large MD value (such as 10, which is a very high MD for a normal group sample)

5. Choose the threshold that maximizes the selected threshold

After the performance measures are calculated for a range of thresholds, select the threshold resulting in the highest performance measure value for each training set.

The thresholds determined for each training set are averaged to find the threshold.

6. Verify the threshold

MTS models formed for each threshold is applied to the respective testing sets, and the threshold value is used to find the performance of the model on the testing set.

3.1.1. Wisconsin Breast Cancer Dataset

Wisconsin Breast Cancer Dataset is also a widely used dataset in machine learning literature. Healthy patients form the majority group with 357 samples and sick patients form the abnormal group with 212 samples.

In this dataset, the mean values of the correct classification rates are close to each other in each method. Mean+3sigma acts poorly in determining the minority class (see: Table 6). As expected, F05 and F2 predicted the classes that these performance measures give more weight to better. Figure 4 shows the average of the performance measures over replications for each threshold method, using the best threshold chosen. For example, For the G-mean performance measure, the average of the total of 9 best thresholds (3 replications with 3 folds) is 1.944 (See Table 6). Using the best threshold in each case, the averages of the performance measures are calculated.

Table 6 Wisconsin Breast Cancer Dataset Performance Measures

	Average of Best Thresholds	PCC	KAPPA	Precision	Recall	Specificity	G-mean	F05	F1	F2	AUC	Stability
F05	3.3	0.873	0.866	0.901	0.742	0.951	0.840	0.862	0.811	0.768	0.956	0.016
F1	2.111	0.888	0.881	0.834	0.879	0.893	0.886	0.841	0.854	0.868	0.956	0.017
F2	1.533	0.861	0.853	0.752	0.946	0.81	0.875	0.783	0.837	0.899	0.956	0.018
Euclidean	1.889	0.882	0.875	0.81	0.899	0.872	0.885	0.826	0.851	0.879	0.956	0.018
PCC	2.133	0.887	0.88	0.835	0.874	0.894	0.884	0.842	0.853	0.865	0.956	0.018
G-mean	1.944	0.883	0.876	0.812	0.898	0.874	0.886	0.827	0.852	0.879	0.956	0.019
Mean+3Std	3.093	0.875	0.868	0.899	0.747	0.951	0.843	0.863	0.815	0.772	0.956	0.011
PTM	1.857	0.886	0.879	0.811	0.909	0.873	0.891	0.828	0.857	0.887	0.956	0.014

3.1.2. Pima Indian Diabetes Dataset

In Pima Indian Diabetes data set, normal (healthy) group form the majority group. The table 6 and Figure 2 show the average of the performance measure values of the testing sets for the selected thresholds.. For example, when F05 is used for the threshold determination, the threshold is found to be 1.267. This threshold resulted with an average correct classification rate of %67.7 in the testing sets. It is also observed that in the testing set, the best F05 mean value is not obtained by using this threshold. The best F05 results (0.548) are actually obtained by using the threshold that maximizes the average g-mean or the Euclidean distance of the point corresponding to that threshold to the best point of the ROC graph.

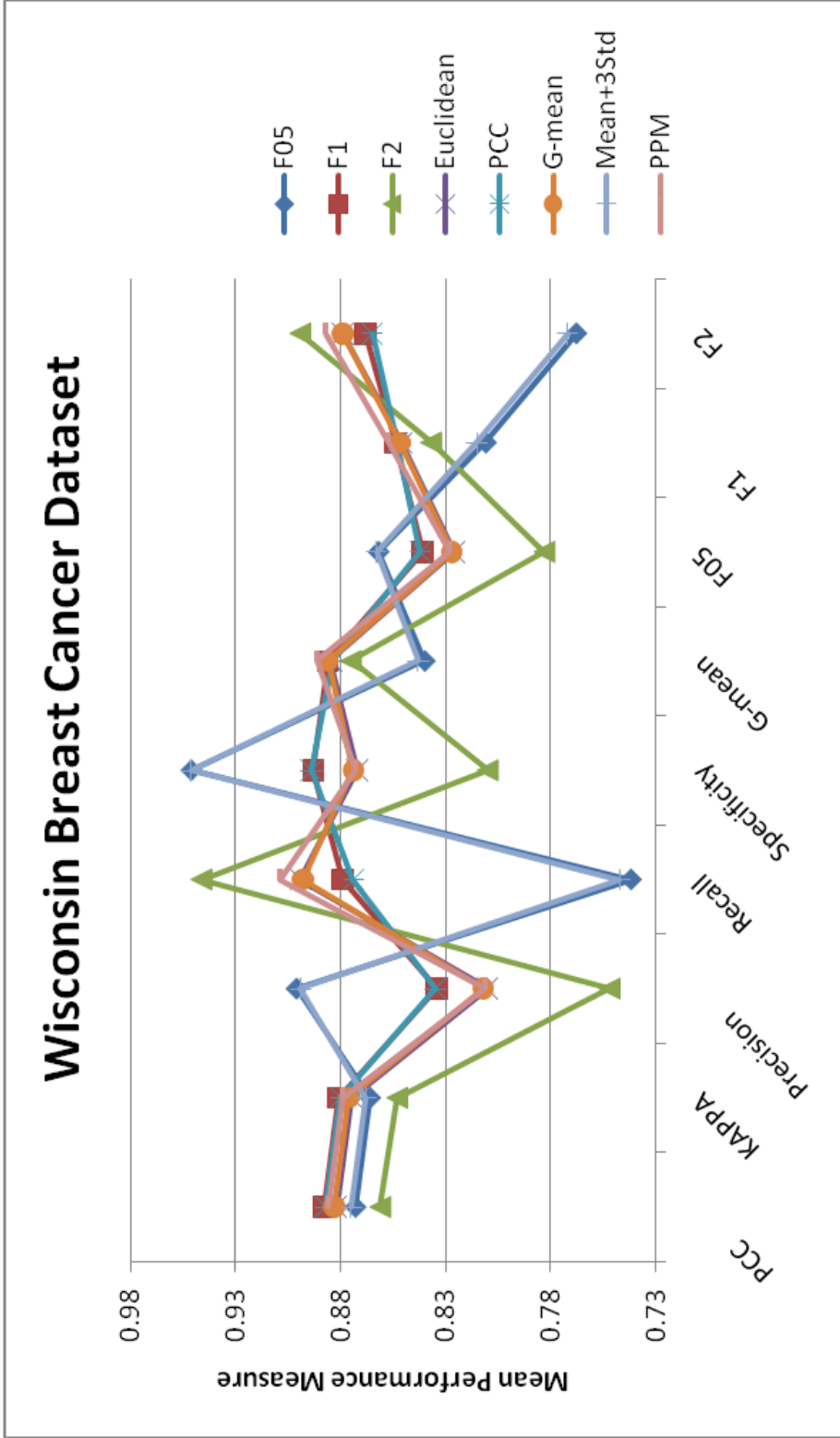


Figure 4 Wisconsin Breast Cancer Dataset Performance Measures

Table 7 Performance Measures of the Pima Indian Diabetes Dataset

	Average of Best Thresholds	PCC	KAPPA	Precision	Recall	Specificity	G-mean	F05	F1	F2	AUC	Stability
F05	1.267	0.677	0.659	0.538	0.563	0.738	0.645	0.540	0.545	0.555	0.723	0.014
F1	0.822	0.646	0.625	0.497	0.771	0.579	0.668	0.534	0.603	0.693	0.723	0.008
F2	0.411	0.486	0.468	0.400	0.940	0.242	0.477	0.452	0.561	0.740	0.723	0.012
Euclidean	0.944	0.668	0.649	0.519	0.711	0.645	0.677	0.548	0.600	0.662	0.723	0.008
PCC	1.600	0.680	0.665	0.549	0.447	0.804	0.599	0.520	0.487	0.461	0.723	0.015
Mean+3Std	3.813	0.665	0.662	0.651	0.085	0.976	0.288	0.276	0.149	0.102	0.723	0.004
G-mean	0.944	0.668	0.649	0.519	0.711	0.645	0.677	0.548	0.600	0.662	0.723	0.008
PPM	1.918	0.683	0.670	0.577	0.356	0.859	0.553	0.512	0.439	0.385	0.723	0.004

Probabilistic Threshold Method results in high correct classification rates but its recall (classification rate of the minority group) is the second lowest. Although the majority class is predicted well, minority class performed poorly.

Although setting the threshold to mean+3*standard deviation results in the best classification rate for the majority group, this also results in a very high threshold of 3.815, placing %91.5 of the minority samples into the wrong class. Looking at the performance measures, it is seen that F1, Euclidean, and G-mean yield similarly good results in all performance measures.

Pima Indian Diabetes dataset is not a clearly imbalanced dataset and many of the threshold methods have resulted in similarly good results except for F2 and Mean+3 Std. Dev.

In this dataset, the majority group was the healthy patients. As MTS model is formed using the normal group, it was expected that the larger the group used to form the model, the better the model is. In order to observe the results when the normal group is the minority, the dataset was made imbalanced by randomly removing samples from it until the normal group is %20 percent of all data. The results are given in Table 8 and Figure 6.

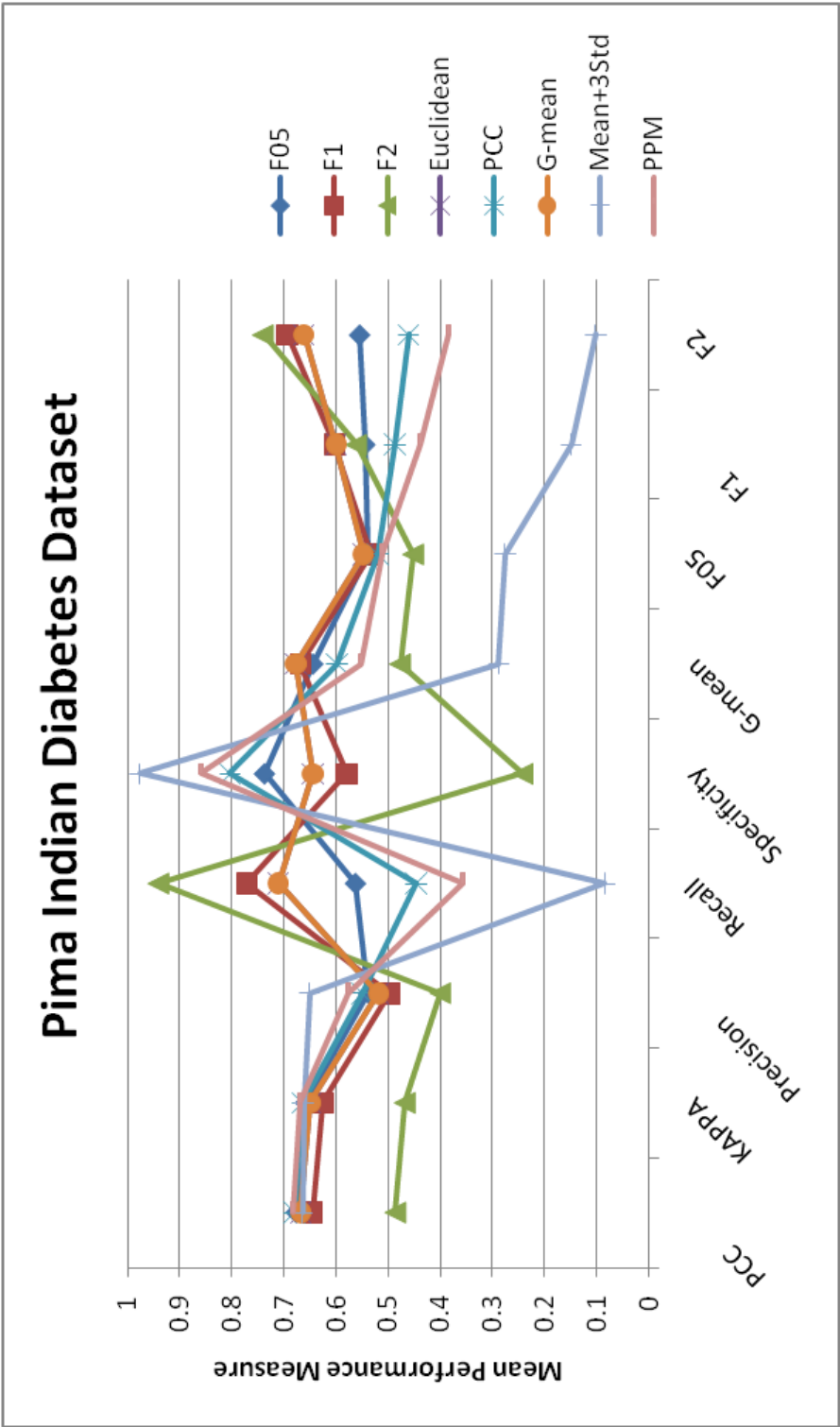


Figure 5 Performance Measures of the Pima Indian Diabetes Dataset

Table 8 Performance Measures of Pima Indian Diabetes Dataset with Normal Group Reduced to %20 of the Dataset

	Average of Best Thresholds	PCC	KAPPA	Precision	Recall	Specificity	G-mean	F05	F1	F2	AUC	Stability
F05	0.7	0.718	0.707	0.826	0.797	0.471	0.613	0.818	0.809	0.801	0.699	0.03
F1	0.244	0.741	0.739	0.765	0.952	0.086	0.286	0.796	0.847	0.906	0.699	0.02
F2	0.078	0.758	0.758	0.759	0.998	0.012	0.109	0.797	0.862	0.938	0.699	0.001
Euclidean	0.922	0.685	0.67	0.855	0.705	0.621	0.662	0.819	0.771	0.73	0.699	0.019
PCC	0.411	0.74	0.735	0.785	0.908	0.216	0.443	0.806	0.84	0.879	0.699	0.025
G-mean	0.911	0.684	0.67	0.852	0.708	0.61	0.657	0.817	0.771	0.732	0.699	0.022
Mean+3Std	3.598	0.318	0.307	0.863	0.118	0.942	0.333	0.376	0.206	0.142	0.699	0.016
PTM	1.838	0.496	0.475	0.879	0.389	0.83	0.568	0.701	0.539	0.438	0.699	0.035

When the dataset is imbalanced, F1, F2, and PCC perform poorly in determining the normal (and also minority) class. PTM on the other hand, performs better in determining the normal class but does not classify the samples in the abnormal group correctly. From these results, it is seen that the best results on the average are given by maximizing the Euclidean Distance and G-mean measures in the training set. Although they do not result in the best average PCC values, they act equally well for both classes.

1.1.1. Magic Gamma Telescope Dataset

Magic Gamma Telescope Dataset contains 12332 normal (no gamma ray reflected from the atmosphere) and 6688 abnormal (detected gamma ray reflection from the atmosphere) observations. From the results, it is seen that all threshold methods except F2 give equal average correct classification rates. PTM and Mean+3sigma give very good results in average precision but they result in low average recall values. The best measure for determining the abnormal samples is F2, but that measure performed poorly on the other measures. F1, Euclidean and G-mean give the most balanced average results in all performance measures.

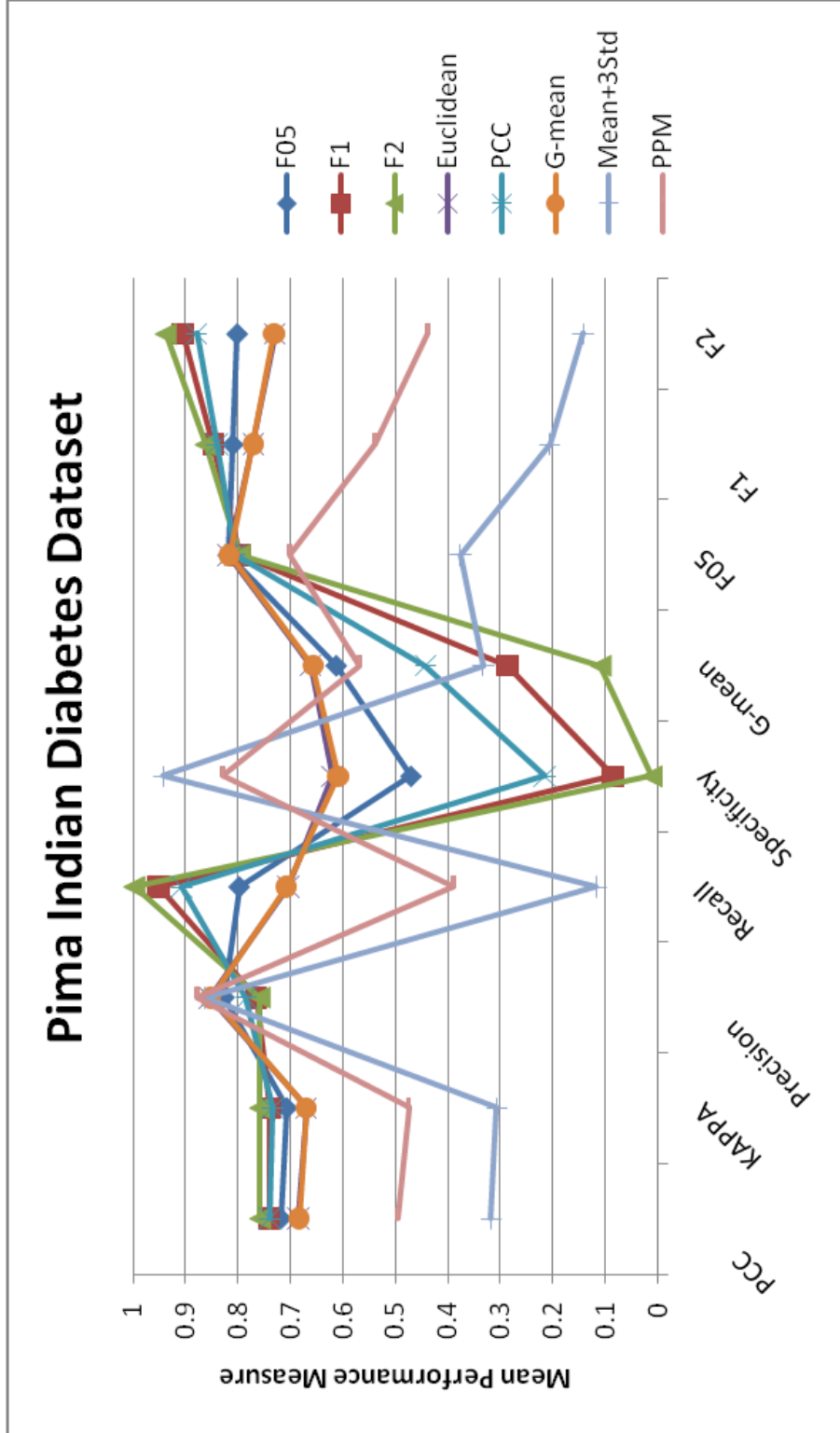


Figure 6 Performance Measures of Imbalanced Pima Indian Diabetes Dataset

Table 9 Performance Measures of Magic Gamma Telescope Dataset

	Average of Best Thresholds	PCC	KAPPA	Precision	Recall	Specificity	G-mean	F05	F1	F2	AUC	Stability
F05	2.067	0.774	0.765	0.79	0.487	0.929	0.673	0.702	0.602	0.527	0.818	0.001
F1	1.167	0.74	0.724	0.61	0.725	0.748	0.736	0.629	0.662	0.698	0.818	0.001
F2	0.6	0.585	0.565	0.456	0.922	0.403	0.610	0.507	0.61	0.765	0.818	0
Euclidean	1.167	0.74	0.724	0.61	0.725	0.748	0.736	0.629	0.662	0.698	0.818	0.001
PCC	1.9	0.775	0.766	0.764	0.525	0.911	0.692	0.699	0.622	0.56	0.818	0.001
G-mean	1.167	0.74	0.724	0.61	0.725	0.748	0.736	0.629	0.662	0.698	0.818	0.001
Mean+3Std	5.899	0.721	0.716	0.953	0.217	0.994	0.464	0.567	0.353	0.256	0.818	0
PTM	2.595	0.764	0.756	0.851	0.398	0.962	0.619	0.693	0.542	0.446	0.818	-0.001

This dataset was made imbalanced by randomly removing samples from the majority group. In this case, a very large imbalance was created by removing all but 67 samples in the normal group. This way, the abnormal class is now 100 times more than the normal class.

Table 10 Performance Measures of Threshold Methods in Imbalanced Magic Gamma Telescope Dataset

	Average of Best Thresholds	PCC	KAPPA	Precision	Recall	Specificity	G-mean	F05	F1	F2	AUC	Stability
F05	0	0.987	0.987	0.987	1	0	0.000	0.99	0.993	0.997	0.769	0
F1	0	0.987	0.987	0.987	1	0	0.000	0.99	0.993	0.997	0.769	0
F2	0	0.987	0.987	0.987	1	0	0.000	0.99	0.993	0.997	0.769	0
Euclidean	1.267	0.683	0.682	0.995	0.682	0.748	0.714	0.91	0.808	0.727	0.769	-0.001
PCC	0	0.987	0.987	0.987	1	0	0.000	0.99	0.993	0.997	0.769	0
G-mean	1.333	0.666	0.665	0.995	0.665	0.748	0.705	0.903	0.795	0.711	0.769	-0.003
Mean+3Std	3.244	0.344	0.342	0.997	0.336	0.92	0.556	0.711	0.501	0.387	0.769	-0.006
PTM	1.718	0.566	0.565	0.997	0.562	0.862	0.696	0.861	0.717	0.615	0.769	0.001

Magic Gamma Telescope Dataset

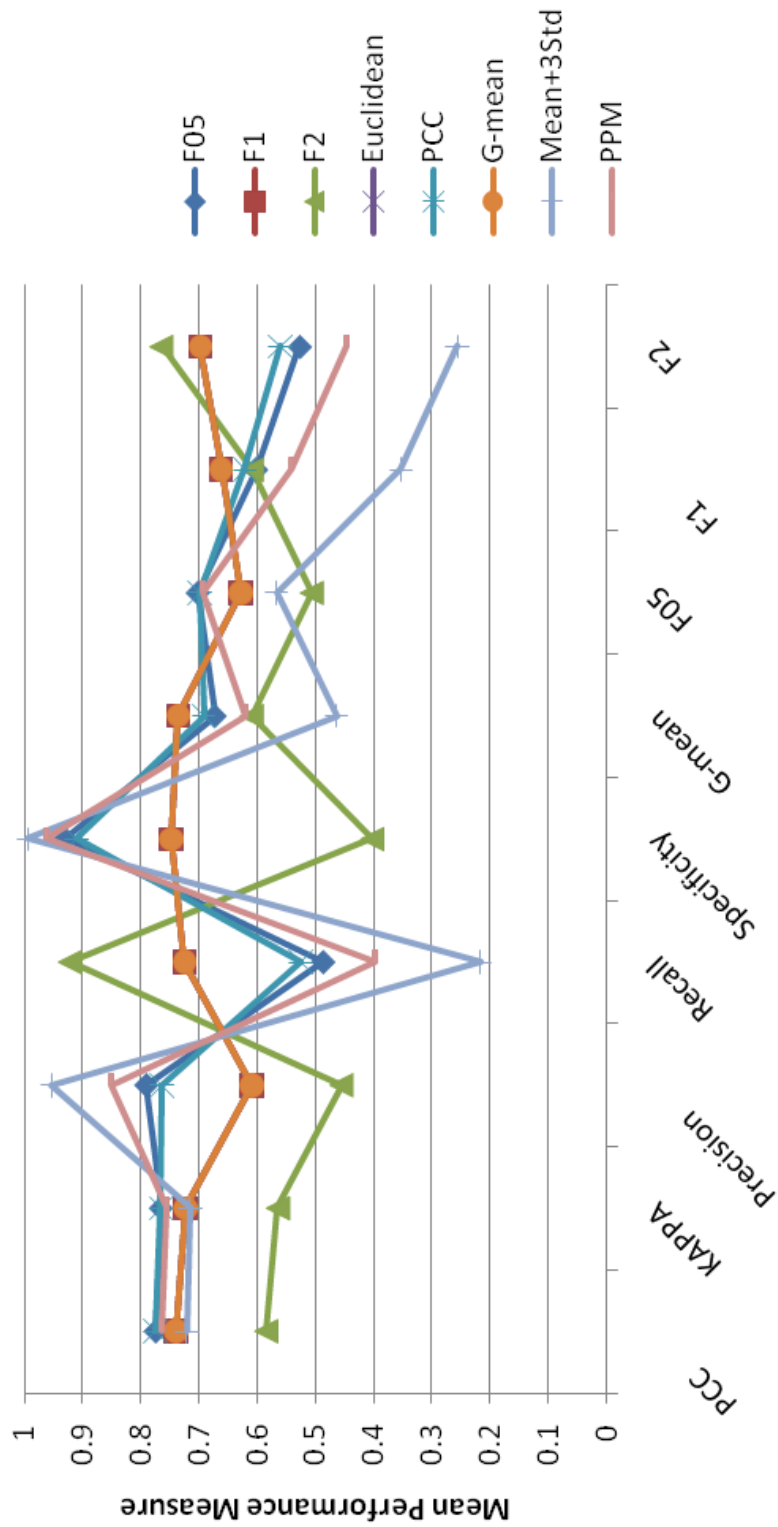


Figure 7 Performance Measures in Magic Gamma Telescope Dataset

The effect of such a significant imbalance can be seen in the average performance measures. By setting the threshold to 0, a correct classification rate of %98.7 is obtained. The recall is 1, meaning that all the samples in the abnormal group are correctly classified. However, Specificity is 0, indicating that all the normal samples are put into the abnormal class. In this highly imbalanced case, the best measures are observed to be G-Mean and Euclidean.

Depending on the importance of classifying each class, a different method can be used. In imbalanced datasets, measures such as PCC can be misleading. It is observed above that misclassification of the minority group does not affect PCC in highly imbalanced cases. Recall and Specificity are better measures to observe in imbalanced datasets. In order to obtain a threshold that would act well for both classes, measures such as Euclidean or G-mean can be used. G-mean is a commonly used performance indicator in imbalanced cases, and setting a threshold that maximizes G-mean did in fact result in good results in the testing sets. Thus, it is suggested that for cases where classification performance of each class is important, the threshold value that maximizes g-mean is used.

In the cases studied in this thesis, G-mean will be used to determine the threshold as the datasets in the quality context are often imbalanced.

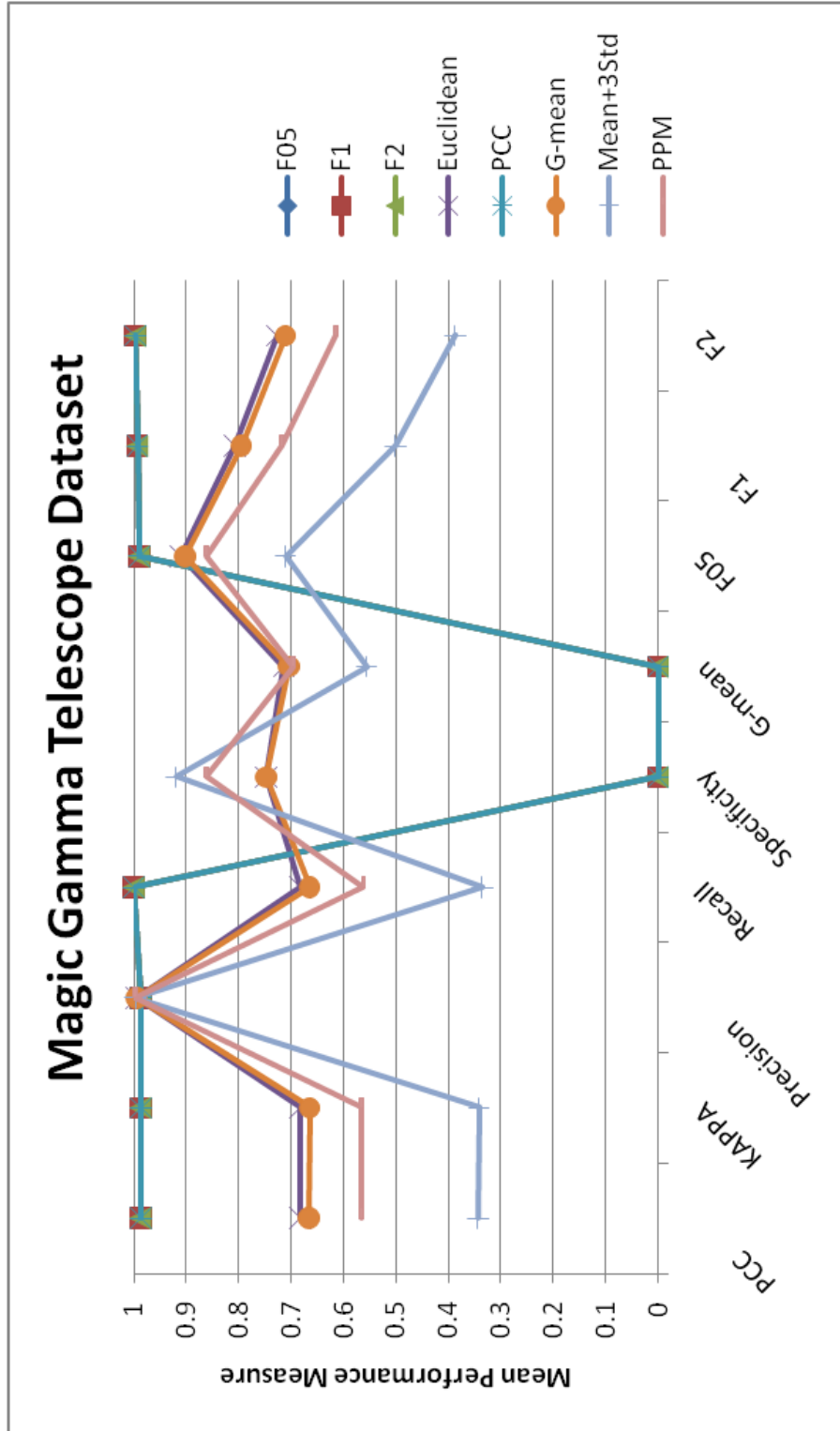


Figure 8 Performance Measures of Imbalanced Magic Gamma Telescope Dataset

CHAPTER 4

PARAMETER OPTIMIZATION USING MTS WITH BINARY RESPONSE

In this thesis, the use of Mahalanobis Taguchi System is proposed for parameter optimization for the first time to the best of our knowledge. The aim of parameter optimization, or robust design, as explained in Section 2.3 is to determine the levels of parameters that optimize a defined response for quality and to minimize the variation caused by noise factors.

The proposed approach aims to make use of Mahalanobis Taguchi System as a mean to understand the relationship between the control factors (\mathbf{x}) and the response (y). After the MTS model is formed and verified to be representing the problem, the model function can be solved inversely, going back from the response to the control factors to determine the optimum factor levels.

This inverse problem is solved and optimized using quadratic modeling. The approach is demonstrated on quality problems, where the desired response is the targeted quality level, such as non-defective product, or high customer satisfaction.

4.1. The Approach

In the proposed approach Mahalanobis Taguchi System (MTS) is used to build a classification model that would be representing the relationship between the control factors and the response. Since MTS is a method developed for classification problems with

binary responses, the two response categories can be normal/abnormal, non-defective/defective, conforming/nonconforming, and so on.

Figure 1 show the parameter diagram commonly used in robust design problems. In our study, static problems with only one signal factor level are considered.

The proposed approach is presented in the Figure 9. The first 5 steps of the approach are in accordance with the robust design steps provided in Phadke (1989). 6th and 7th steps are the steps related to forming the classification model. Starting from the 8th step, the mathematical programming model is formed and optimized. The last step is the verification of the results. If the classification model does not perform as expected or if the optimal solution found is away from the desired class, the process have to be started from the start, checking the variables inserted into the model.

The proposed approach starts with understanding the production function and defining the factors that can or cannot be controlled to change the output of the process (controllable and noise factors, respectively). In the second step, the separation between the noise factors and the control factors should be made clear. Only control factors (\mathbf{x}) should be included in the MTS model, noise factors should not be included.

After the factors are defined, the response that will be measured to observe the quality level is selected.

It is important that the samples that are collected before starting the building of the MTS model be a result of a designed experiment. The levels of the control factors are determined and combinations of these levels are experimented using an appropriate matrix experiment. The levels that are experimented have to be the nominal levels of the control factors, that is, they are not the levels that are measured but the levels that are set by the experimenter. The levels set by the experimenter can differ due to noise factors during the experiments.

After a matrix experiment is designed and is used to collect samples, the responses of these samples are measured. These responses are used in the first step of MTS, defining the reference class. The reference class is used to develop the classification model that will be used to determine whether the other observations are from the reference class or not. If

the response of the process is continuous or ordered categorical, it must first be converted into a two class response.

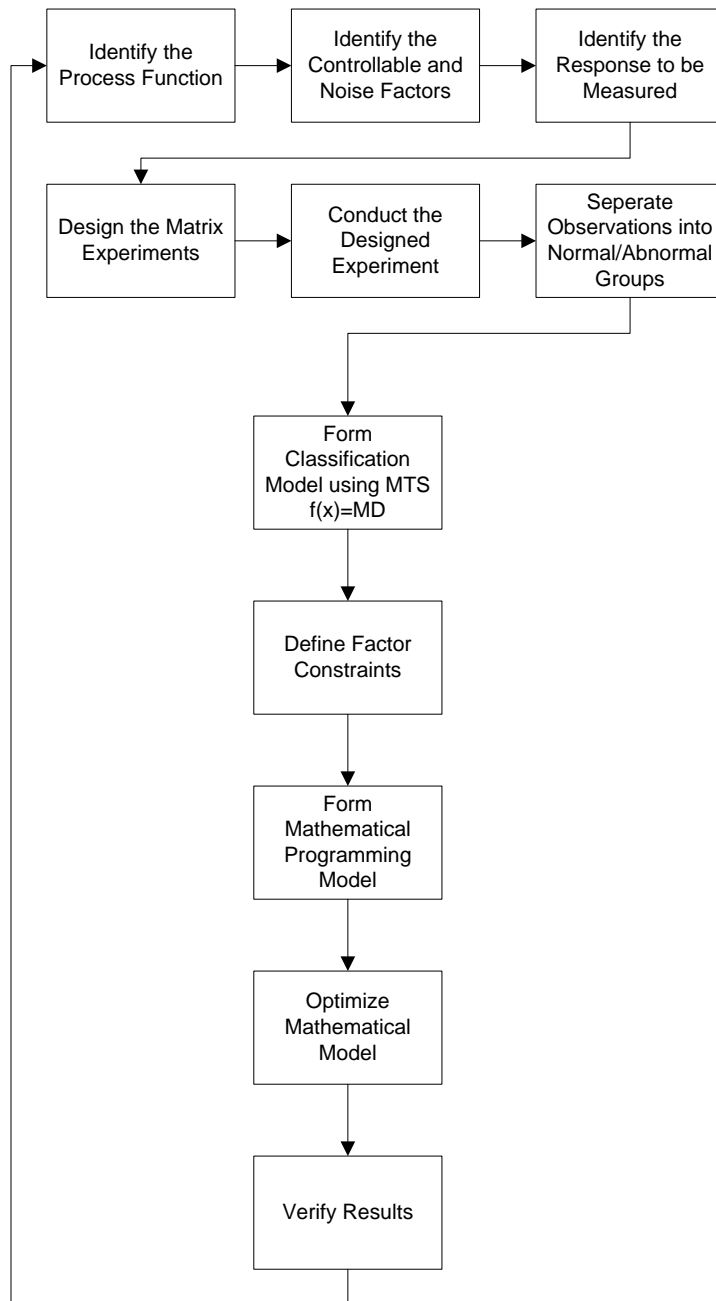


Figure 9 Proposed Approach

The classification model formed using MTS can be used to determine the distance of any observation to the reference class. This model, a function of the control factors, can then be solved inversely, going back from the desired response to the optimal control factor levels. As the response of the MD function is the distance to the desired output (the reference class), the objective is to minimize the distance to the center of the reference (or normality) region, while satisfying any constraints placed on variables. The constraints may include manufacturing or technology limitations, boundaries on the variables and integer variables.

The Mahalanobis Distance formula is used as the objective function of the optimization problem. The Mahalanobis Distance from an observation $\mathbf{x}=\{x_1, x_2, \dots, x_p\}$ to the center of the group $\mathbf{m}=\{\mu_1, \mu_2, \dots, \mu_n\}$ was defined as

$$MD^2 = (\mathbf{x} - \mathbf{m})^T \Sigma^{-1} (\mathbf{x} - \mathbf{m})$$

where Σ is the covariance matrix of the variables.

From the Mahalanobis Distance formula it is seen that the lowest Mahalanobis distance is obtained when the variables are equal to the mean values of the variables in the normal group. When $x_i=\mu_i$ for all $i=1, \dots, p$, MD is equal to 0. However, if there are constraints on the variables, this may not be allowed. If the variables are integer or binary, this creates constraints on the variables, where it may not be able to assign these variables their respective means. The other reason may be the mean value of the normal group being out of the specification limits. In these situations, nonlinear optimization methods can be used to find the optimal values of the decision variables.

The objective function of the problem, the MD formula, is a quadratic function. If the objective function is quadratic, but the constraints are linear, the optimization problem becomes a special case of nonlinear optimization problems, a quadratic program (QP).

The quadratic program is defined as:

$$\text{Minimize (or Maximize) } f(\mathbf{x}) = \mathbf{c}\mathbf{x} + \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}$$

Subject to

$$\mathbf{A}\mathbf{x} \leq \mathbf{b}$$

$$\mathbf{x} \geq 0$$

where $\mathbf{c}=\{c_1, c_2, \dots, c_p\}$ are the coefficients of the linear terms in the objective function, and \mathbf{Q} is an $p \times p$ symmetric matrix describing the coefficients of the quadratic terms. The variables are denoted by the p -dimensional column vector \mathbf{x} , and the constraints are defined by an $(m \times p)$ \mathbf{A} matrix and an m -dimensional column vector \mathbf{b} .

When the objective function $f(\mathbf{x})$ is strictly convex for all feasible points, the problem has a unique local minimum which is also the global minimum. A sufficient condition to guarantee strictly convexity is for \mathbf{Q} to be positive definite.

The difficulty of solving the quadratic programming problem depends largely on the nature of the matrix \mathbf{Q} . In convex quadratic programs, which are relatively easy to solve, the matrix \mathbf{Q} is positive semidefinite. If \mathbf{Q} has negative eigenvalues - nonconvex quadratic programming - then the objective function may have more than one local minimizer. For positive definite \mathbf{Q} , the problem can be solved in polynomial time.

An $n \times n$ real symmetric matrix \mathbf{Q} is positive definite if $\mathbf{x}'\mathbf{Q}\mathbf{x} > 0$ for all non-zero vectors \mathbf{x} with real numbers. Matrix \mathbf{Q} is positive definite if and only if all eigenvalues are positive.

Three properties of positive definite matrices are also important to mention:

- a. Every positive definite matrix is invertible and its inverse is also positive definite
- b. If \mathbf{Q} is positive definite and $r > 0$ is a real number, then $r\mathbf{Q}$ is positive definite
- c. Since the product of eigenvalues of a matrix equals its determinant, the determinant of a positive definite matrix \mathbf{Q} is positive.

The quadratic programming model in the proposed method can be defined as

$$\text{Minimize } f(x) = \mathbf{c}\mathbf{z} + \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} \quad [4.1]$$

where $\mathbf{c}=\mathbf{0}$, $\mathbf{z} = (\mathbf{x} - \mathbf{m})$, $\mathbf{Q} = \Sigma^{-1}$

subject to

$\mathbf{x} \leq$ upper limits

$\mathbf{x} \geq$ lower limits

and Σ is the covariance matrix.

In order to claim that the problem has a unique global optimum, the covariance matrix has to be positive-definite. From the definition of covariance matrix, it is known that covariance matrices are positive-semidefinite.

Covariance matrix is calculated by the following:

$$\Sigma = E[(\mathbf{x} - E(\mathbf{x}))'(\mathbf{x} - E(\mathbf{x}))]$$

In order for Σ to be positive-semidefinite,

$$\mathbf{a}'[(\mathbf{x} - E(\mathbf{x}))'(\mathbf{x} - E(\mathbf{x}))]\mathbf{a} \geq 0 \text{ for any } \mathbf{a} \neq 0$$

or

$$((\mathbf{x} - E(\mathbf{x}))\mathbf{a})((\mathbf{x} - E(\mathbf{x}))\mathbf{a}) \geq 0$$

which is always correct.

4.1.1. Generation of Mahalanobis Space

The most important step in MTS is the generation of Mahalanobis Space. Mahalanobis Space is the reference (or "normality") region defined by p variables. Generation of Mahalanobis Space includes determination of variables, and collection of samples from the normal group. If there are any outliers, they will also have to be removed. There are different methods to determine the outliers as explained in Section 2.1. In this study, samples with MDs larger than 3 standard deviations from the mean MD of the normal group are regarded as outliers and are removed.

In the first step, n observations that best represent the normal class are selected. The "normal" group must be the group that is expected to be reached and also the group which is as homogeneous as possible. As the means and the covariance matrix of this group is used in model forming, the samples forming the normal group have to be sharing some common properties.

In quality problems, the normal group must be the non-defective/conforming group. This group is expected to be more homogenous than the defective group, as different types of defects can be present among the defectives. In addition, the optimal parameter levels found are the levels that result in an optimal point that is as close to the selected normal group as possible.

It was explained in Section 2.2.1 that size of the normal group must be significantly larger than number of variables p . If the number of samples is less than the number of variables, a singular correlation matrix will appear, rendering the calculation of MD impossible.

In cases where the data is already present and it is not possible to collect any more samples, it must be assured that the number of samples are at least as many as the number of variables. In this step, a resampling method can be used in order to satisfy the sample size requirement and also to have more stable estimation of the correlation matrix. Among many resampling methods, SMOTE is selected to be a suitable oversampling method. SMOTE, as explained in SECTION 5.1, is an oversampling method that creates "synthetic" samples from the actual samples. Using SMOTE, new samples are created without actually growing the normality region determined by the original samples.

In the next step, the MDs of the samples in the normal group are calculated. If new synthetic samples are created and inserted into the normal group, their MDs should be calculated as well. Using the calculated distances, samples that make the normal group should be checked for outliers. If the samples that have large MDs are outliers, they have to be removed from the normal group. This prevents samples that are actually away from the group from enlarging the normality region. For this reason, the samples that are more than 3 standard deviations away from the mean of the normal group MDs can be removed in an iterative process. After each step, Mahalanobis Distances are recalculated with the new variable means and covariance matrix. This removal is continued until no outlier is left.

It is also important that the sample size limitations are satisfied while removing outliers. If the number of samples becomes less than the number of variables, new samples have to be collected and inserted into the normal group. If it is not possible to collect new samples, SMOTE can be used to create new synthetic samples.

Having removed outliers, the samples left make up the reference group that are used in the proceeding steps.

4.1.2. Evaluation of Discrimination Power and Selection of Important Variables

After determining the normal samples used as the reference, the discrimination power of MTS model is optimized using Taguchi's robust design methodologies. In order to do this, samples that are known to be outside the normal group are used. Although the normal group members should be from the same population and uniform, abnormal samples can be from any condition outside the normality conditions. The abnormal samples are used to select the important variables that result in best discrimination.

In order to select significantly useful variables, orthogonal arrays and Signal to Noise (S/N) ratios are used in this step to examine the effects of the variables on the MDs.

Orthogonal arrays (OA) are used to examine the effects of forming models with different variable combinations. Each variable is assigned to a column of a two level orthogonal array

where level 1 is to use a variable, and level 0, is not to use variable. Each row in the orthogonal array represents a variable combination. For each variable combination, the distances of the abnormal samples to the normal group mean are recalculated. MDs of the normal samples do not have to be calculated again, as the model is based on the assumption that they are normal and form a uniform normality region.

The use of Orthogonal Arrays is suggested as means to examine the effect of each variable with less number of runs. However, if the variables are not as many as to forbid trial of each variable combination, it is suggested that a full factorial design is used.

After calculating the MDs for the abnormal samples for different combinations of the variables, Signal to Noise ratios (S/N) are used both to determine which variables are useful in discriminating and to evaluate the performance of the constructed scale. For each row of the orthogonal array, S/N ratio is computed for these two purposes.

Among different S/N ratios, larger-the-better type is selected as it is desired that the formed model makes the distances of the abnormal samples to the Mahalanobis Space as high as possible.

In the next step, S/N ratios of all variable combinations in OA are calculated and Analysis of Variance is performed. The null hypothesis is whether including a variable changes the mean S/N ratio or not. α is selected as 0.10. For the variables that are determined using ANOVA to significantly affect S/N ratios, the gains obtained are calculated. The gain is the difference between the average S/N ratio of the runs that include the variable and the average S/N ratio of the runs that do not include the variable. If including a variable increases the S/N ratio, it is included into the model.

The variables found using S/N gains are used to form the MTS model. Mahalanobis Distances of the normal and the abnormal samples are recalculated using the selected variables. Although this step is not directly required for optimization, it is important in order to compare the performance measures of the original and reduced models.

4.1.3. Threshold Selection

In 4.1.2, the MTS model that best separates the abnormal samples from the normal samples in the overall are determined. To decide on whether a sample is in the normality region or not, a threshold must be determined after forming the MTS model and calculating the MDs of the normal and the abnormal samples. Samples having lower MDs than the threshold will be classified normal, while the others will be classified as abnormal. For threshold determination, it was proposed in Section 3 that an exhaustive search method is used, with the objective of maximizing the selected threshold. It was determined that for imbalanced datasets, the threshold is chosen so that g-means measure is maximized.

Although this step is not a requirement for using the MTS model in parameter optimization, assessing the classification performance gives insight about the usability of the MTS model in parameter optimization. If the classification performance is poor, different actions can be taken to ensure that the model describes the case studied as best as possible. The variables used in modeling can be questioned, the models can be reformed etc. to increase the classification performance.

4.1.4. Parameter Optimization using Quadratic Programming

In 4.1.2, the MTS model that best classifies the normal and abnormal samples is formed. Using the mean, and the covariance matrix from the normal group, the quadratic program is constructed with the objective of minimizing the Mahalanobis Distance. The solution to this quadratic model [4.1] gives the parameter values that result in the smallest Mahalanobis Distance possible to the centroid of the normal group, subject to parameter constraints.

4.2. Applications

4.2.1. Casting Defect Problem

The proposed approach is applied to a real life dataset, originated from a Six-Sigma project performed in an Iron-Casting company of reducing defects in an iron-casting process. Details about this case can be found in İşkol et al. (2005). This dataset was also studied by Erdural (2006) in his thesis for developing a method for "robust design of products or processes with categorical response".

The data used includes 7 process variables. The observations in this problem are not collected using a designed experiment. 138 different combinations of the 7 variables have been experimented for a total of 12417 times each times producing a casting. Thus, there are 138 rows in the dataset. It is assumed that the levels of the 7 variables are not the actually observed levels, but are the nominal levels.

The number of trials (items produced) for each combination and the number of defectives produced are available as the response.

The variable combinations were observed in the range of 7 to 180, depending on how many items are produced under each combination with the defects occurring at most 8 times in these combinations. For each row, the total produced and the defectives observed are the two responses from these experiments.

The first step of MTS is to determine the samples that will be forming the normal group. As this dataset contains repeated measurements in 138 variable levels, there is not a binary response that can be used directly. The two attributes that can be used to select the normal samples are the total number produced in that variable combination and the number of defectives produced as a result of this production.

The acceptable defect rate was defined as %5 by the process owners, İşkol et al. (2005). Therefore, the samples meeting this criterion have to be used to form the normal group. However, this is not a designed experiment, and variable combinations (rows) have been observed in different quantities. For example, the samples 6 and 128 both have zero

defects. Looking at the total produced, sample 6 have been produced 160 times while the sample 128 has been produced only 8 times. Selecting both of these samples to be in the normal group by looking at the defective ratio may be misleading. In order to take the sample size into consideration, hypothesis testing is used.

It is desired to test the hypothesis that in n samples, the defect rate p is smaller than the acceptable defect rate $p_0 = 0.05$.

$$H_0: p \geq p_0$$

$$H_1: p < p_0$$

To test this hypothesis, the following statistic is used:

$$Z_0 = \begin{cases} \frac{(x + 0.5) - np_0}{\sqrt{np_0(1 - p_0)}} & \text{if } x < np_0 \\ \frac{(x - 0.5) - np_0}{\sqrt{np_0(1 - p_0)}} & \text{if } x > np_0 \end{cases}$$

As the hypothesis is one-sided, the null hypothesis is rejected if $Z_0 \leq -Z_\alpha$ (Montgomery 2005)

The samples in which it is not possible to reject the null hypothesis are included into the normal group. For $\alpha = 0.10$, the samples are rejected if $Z_0 < -1.28$

Following these rules, 75 rows are selected to be producing non-defective products. These will be used as the normal group. The remaining 63 samples are regarded as abnormal.

Prior to modeling the data with MTS, it is possible to examine the variables in the dataset. Examining the means of the normal and abnormal groups, it is seen that for variables X1, X2, X3, X4 and X6, the mean of the normal group is larger than the abnormal group. For X7, the abnormal group has a higher mean than the normal group. In X5, both groups have very close means.

Table 11 Mean values of the variables for the normal and abnormal groups

	X1	X2	X3	X4	X5	X6	X7
Normal	-0.267	-0.243	-0.147	-0.138	0.009	-0.214	0.243
Abnormal	0.318	0.289	0.175	0.165	-0.010	0.251	-0.289

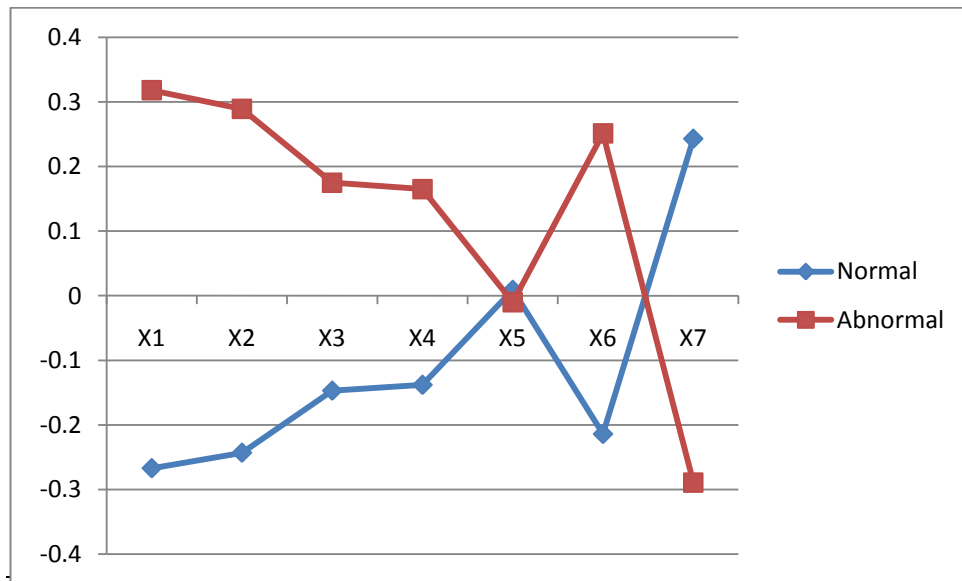


Figure 10 Mean values of the variables for the normal and abnormal groups

In the same way, the standard deviations of the groups are provided in Table 12.

Table 12 Standard deviations of the normal and abnormal groups

	X1	X2	X3	X4	X5	X6	X7
Normal	0.773	0.747	0.736	0.769	0.902	0.811	1.072
Abnormal	1.136	1.171	1.188	1.164	1.113	1.141	0.825

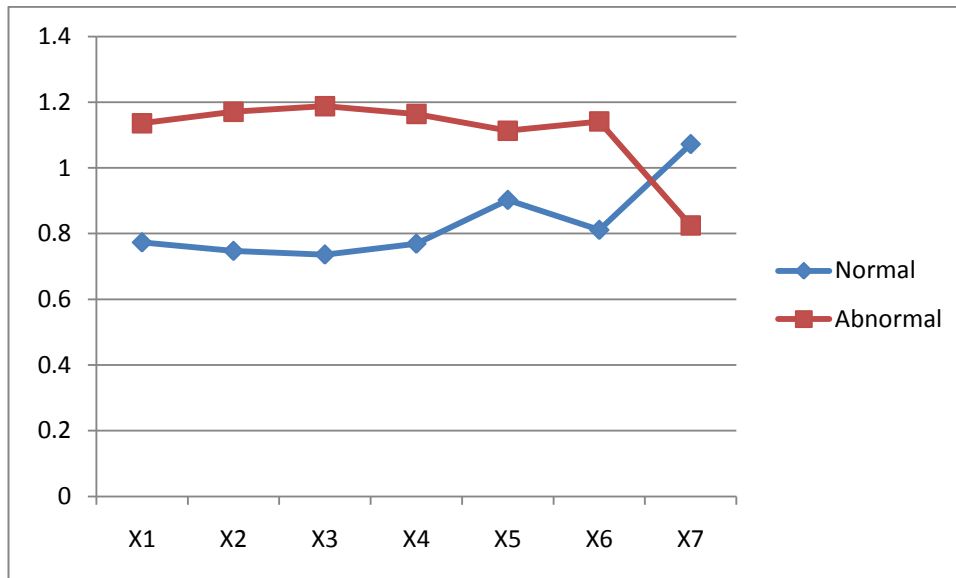


Figure 11 Standard Deviations of the Normal and Abnormal Groups

The standard deviations of the variables in the abnormal groups are higher than the variables in the normal group, except the variable X7. It can also be seen that the standard deviations are much higher than the variable means.

The minimum and the maximum values for the variables are shown in Table 13.

Table 13 Minimum and maximum values of the variables

	X1	X2	X3	X4	X5	X6	X7
Min. Normal	-1.704	-1.419	-1.610	-1.597	-2.002	-1.416	-0.778
Min. Abnormal	-1.704	-1.419	-1.681	-1.673	-2.397	-1.947	-1.328
Max. Normal	0.809	1.127	2.036	1.841	1.946	0.796	1.973
Max. Abnormal	3.044	2.747	3.608	3.904	3.130	2.478	2.083

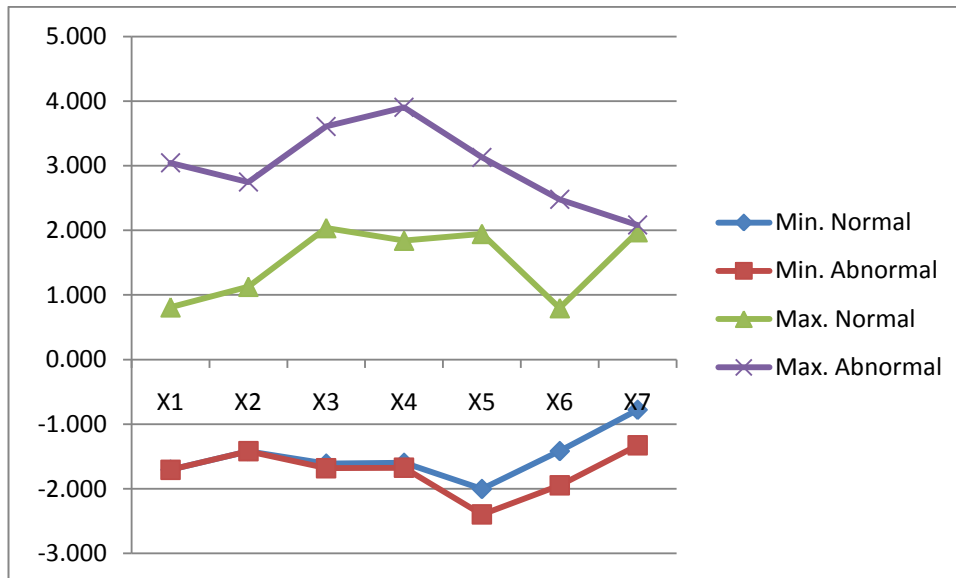


Figure 12 Minimum and maximum values of the variables

It is seen that although the minimum values in each group are very close to each other, the maximum values are higher in the abnormal group.

From the information about the variables, even though there are differences between the groups when each information is checked separately, there is actually no clear distinction between the two groups.

As the normal group will be used as a reference to understand the correlations between the variables, the correlation matrix can also be examined. If there are variables that are highly correlated to other variables, they can cause multicollinearity, or even singularity. As such a case prevents from calculating the Mahalanobis Distances; the variables causing this problem have to be removed before forming the model. The correlation matrix of the variables for the normal groups is given in Table 14.

Table 14 Correlation Matrix

	X1	X2	X3	X4	X5	X6	X7
X1	1.000	0.865	-0.323	-0.214	0.006	0.866	-0.782
X2	0.865	1.000	-0.413	-0.228	0.017	0.894	-0.809
X3	-0.323	-0.413	1.000	0.599	-0.140	-0.435	0.407
X4	-0.214	-0.228	0.599	1.000	-0.078	-0.288	0.226
X5	0.006	0.017	-0.140	-0.078	1.000	0.223	-0.417
X6	0.866	0.894	-0.435	-0.288	0.223	1.000	-0.930
X7	-0.782	-0.809	0.407	0.226	-0.417	-0.930	1.000

It is seen that there is a high positive correlation between the variables X1, X2, and X6 . Also, X7 has a high negative correlation with variables X1, X2, and X6. Finally, X3 has high positive correlation with X4. The correlations of the variables can be seen in Figure 13.

There are 75 normal and 63 abnormal samples in this case. The first step is to remove the outliers from the normal group. To do this, the Mahalanobis distances of the samples are calculated. The samples that have distances more than 3 standard deviations away from the average MD of the normal group are discarded for being outliers. After removal of outliers, Mahalanobis Distances of the samples are recalculated, and outliers are searched in an iterative process until no outlier is left.

Out of the 75 samples in the normal group, 1 sample is removed for being an outlier. The remaining 74 samples are used as the normal group.

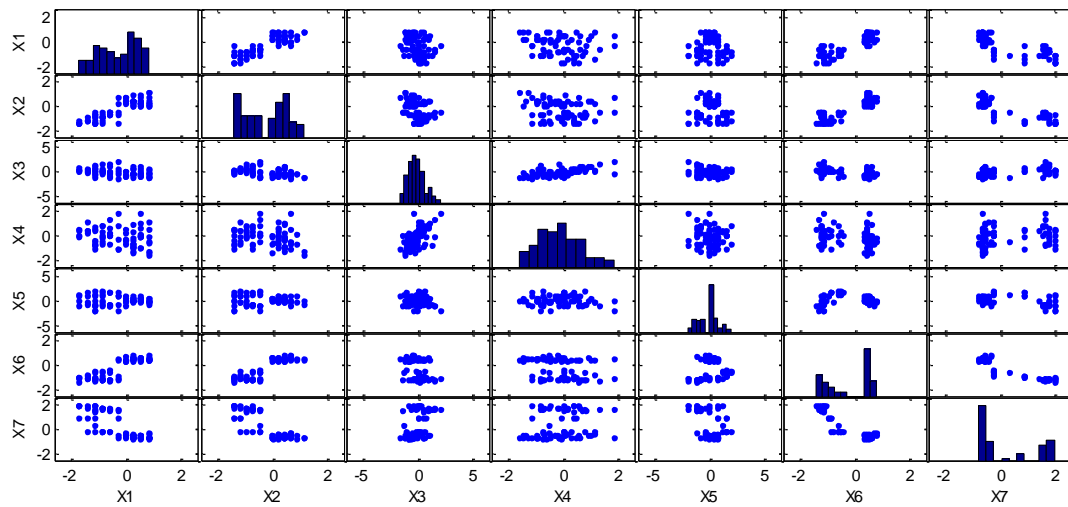


Figure 13 Correlation graphs of the variables

Having defined the samples in the normal group, the MD of the normal and the abnormal samples can be calculated.

Using all variables, the resulting MDs are given in Table 14. The mean of the abnormal samples is 7.387, a much higher value than the mean of the normal group. However, the standard deviation of the abnormal group MDs is also large. The minimum MD value of the abnormal group is 0.205, which is actually lower than the minimum MD of the normal group samples.

Table 15 Basic statistics of the MTS model using all variables

	Normal	Abnormal
Mean	0.986	7.387
Standard Deviation	0.549	11.777
Minimum	0.260	0.205
Maximum	2.508	46.715

The resulting performance measures are seen in Table 15. From Figure 12 it was seen that there is an overlapping of both classes. There is no clear distinction between the classes. As the threshold was chosen so that the g-mean is maximized (See Section 2.4), it is important

first to check the g-mean. The g-mean value is 0.635. Recall and Specificity shows that the normal samples are classified better (0.524 vs. 0.747) than the abnormal samples.

Table 16 Performance measures of the MTS model with all variables

Threshold	1.300
PCC	0.645
KAPPA	0.623
G-mean	0.635
Precision	0.625
Recall	0.524
Specificity	0.747
F05	0.609
F1	0.574
F2	0.543
AUC	0.665

Looking at the overall prediction rate, Percent Correctly Classified (PCC) is 0.645. KAPPA measure of 0.623 shows that there is a moderate agreement between the actual classes and the predicted classes.

After the results are observed using all variables, useful variables that actually increase the discrimination between the classes have to be determined. Originally, S/N ratios and Orthogonal Arrays are used to find the variables that are increasing the discrimination between the two classes significantly. To do this, calculation of MDs will be redone for different combinations of variables. The reason OAs are used is to reduce the number of runs. In this problem, the number of variables is 7. Thus it is computationally feasible to try all possible combinations of variables. $2^7-1 = 127$ variable combinations are tried.

For each combination, the larger-the-better S/N ratios are calculated using the normal samples. To observe the direction of the change in S/N ratios, main effects of the variables are calculated. The variables that actually contribute positively to the discrimination power when included into the model.

Alternatively, as all possible variable combinations are tried, the variable combination that maximizes the S/N ratio can be selected. In this case, the maximum SNR is obtained using the variables X1, X2, X3, X5, and X7. These variables result in a S/N ratio of 117.

Table 17 Parameter levels that result in the maximum S/N

X1	X2	X3	X4	X5	X6	X7
1	1	1	0	1	0	1

Table 18 shows some basic statistics about the MDs calculated. Comparing to the all variable case, it is now seen that the mean of the abnormal group decreased from 7.387 to 4.9519. Also, the maximum MS have decreased from 46.715 to 25.9905.

Table 18 Basic statistics of the optimized MTS model

	Normal	Abnormal
Mean	0.9865	4.9519
Standard Deviation	0.5950	7.2003
Minimum	0.1385	0.2796
Maximum	2.8804	25.9905

Although the average Mahalanobis Distances are observed to decrease, the performance measures given in the table below show that performance of the model have in fact increased as seen in Table 19.

It is seen that the classification rate of both classes increased, therefore increasing the g-mean value from 0.635 to 0.694. The overall classification rate also increased from 0.645 to 0.681. Aside from there mentioned, all other performance measures increased by removing the variables X4 and X6.

Looking at the classification performance of the model, it is seen that the defective samples were predicted poorly. Classification results can also be seen in Figure 15.

Table 19 Performance Measures of the Optimized Model

Threshold	1.410
PCC	0.681
KAPPA	0.662
G-mean	0.694
Precision	0.657
Recall	0.540
Specificity	0.800
F05	0.656
F1	0.607
F2	0.565
AUC	0.697

The model built using MTS can then be used to determine whether new samples are defective or not. The Mahalanobis Distance can be calculated using the covariance matrix and the means of the variables for the normal group.

The covariance matrix is given in Table 20.

For the normal group, the mean of the variables are provided in Table 21.

This information is sufficient to determine the values the parameters should take in order to produce non-defective samples. Looking at the Mahalanobis Distance formula, it is evident that the optimal parameter values are equal to the means of the normal group. However, there are production limitations that may prevent us from setting the parameters to the mean values. The limitations for this case are provided by Erdural (2006). It was also added that these production limitations are based on economical rationales. These production limits are given in the Table 22.

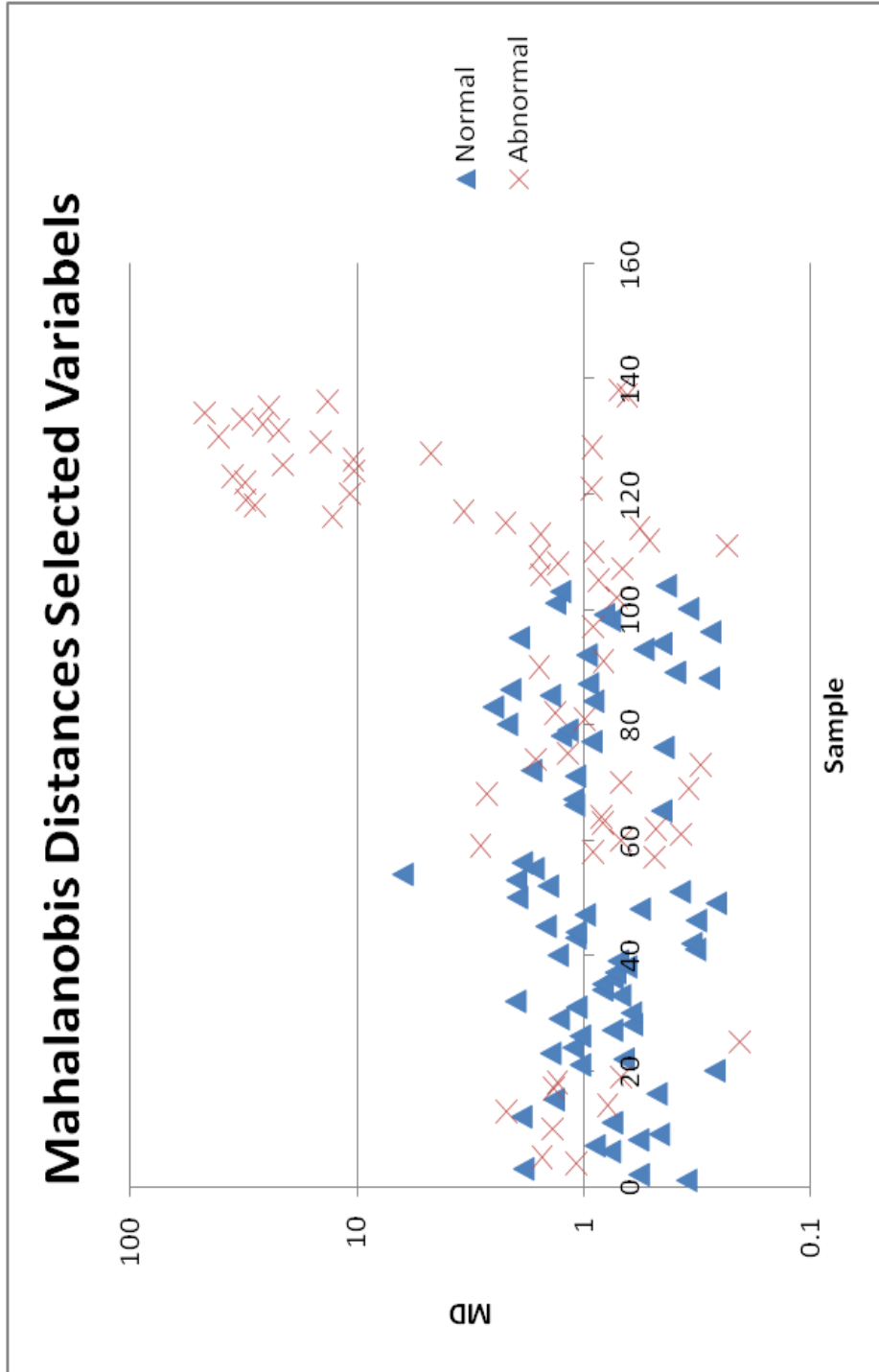


Figure 14 Mahalanobis Distances using All Variables

Table 20 Mean Values of the Variables

Variable	Mean
X1	-0.251
X2	-0.230
X3	-0.150
X4	-0.156
X5	0.014
X6	-0.206
X7	0.249

Table 21 Covariance Matrix of the Variables

	X1	X2	X3	X4	X5	X6	X7
X1	8.851	-3.101	-0.761	0.210	1.202	-4.223	0.875
X2	-3.101	12.342	1.149	-0.743	1.744	-6.492	1.055
X3	-0.761	1.149	3.419	-1.722	0.056	0.065	-0.392
X4	0.210	-0.743	-1.722	2.770	0.112	1.531	0.849
X5	1.202	1.744	0.056	0.112	2.569	0.704	3.024
X6	-4.223	-6.492	0.065	1.531	0.704	23.463	10.436
X7	0.875	1.055	-0.392	0.849	3.024	10.436	10.328

The limits for the variables X3 and X4 were not provided by Erdural (2006) as they were not present in the LR models. For these variables the minimum and maximum values for these variables observed from the dataset are used.

Looking at the production limits on the variables, it is seen that the mean of X6 and X7 are outside their production limits. X7, being in the model, is outside the specified limits if the production is made using the mean values of the normal group. As the samples had to be discretized to be used in MTS, the means of the defined normal group did fall outside the specified limits. Even though discretization of the samples into two classes does change the means of the variables, there are only 17 samples with X7 within the specifications, out of

138 samples. Thus, it is possible that the while doing the experiments, the production limits were not followed.

Table 22 Production Limits of the Parameters

	Lower Limit	Upper Limit
X1	-0.86	0.25
X2	-1.19	-0.03
X3	-1.68	3.6
X4	-1.68	3.9
X5	-1.6	0.37
X6	-1.68	-0.97
X7	1.53	1.97

This limitation prevents using mean values for any of the variables because of the correlations between the variables. If a variable cannot be set to the mean, then there may be points closer to the centroid of the normal group where every variable takes different values.

This information about the normal region will then be used to find the point closest to the center of the normal space while satisfying the constraints. Excel Premium Solver with XPRESS solver engine was used to solve the problem. XPRESS Solver Engine uses primal-dual-path following Interior Point and Newton-Barrier method, based on Mehrotra predictor-corrector method.

The optimization model is

$$\text{Minimize } f(x) = \frac{1}{p} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

subject to

$$-0.86 \leq x_1 \leq 0.25$$

$$-1.19 \leq x_2 \leq -0.03$$

$$-1.68 \leq x_3 \leq 3.6$$

$$-1.68 \leq x_4 \leq 3.9$$

$$-1.6 \leq x_5 \leq 0.37$$

$$-1.68 \leq x_6 \leq -0.97$$

$$1.53 \leq x_7 \leq 1.97$$

$$\mu = \begin{pmatrix} -0.251 \\ -0.230 \\ -0.150 \\ -0.156 \\ 0.014 \\ -0.206 \\ 0.249 \end{pmatrix}$$

Σ^{-1} =inverse of the covariance matrix.

Solving this quadratic model, the optimal parameter values are found are shown in Table 23.

Table 23 Optimal Parameter levels of the MTS model

X1	X2	X3	X5	X7
-0.860	-0.874	-0.710	-0.847	1.530

It is seen that X1 and X7 are on their lower limits. X2, X3 and X5 are closer to their lower limits.

These parameters result in a MD of 0.309, which is lower than the threshold of 1.41. According to this model, these parameters result in a non-defective solution that is close to the centroid of the normal group.

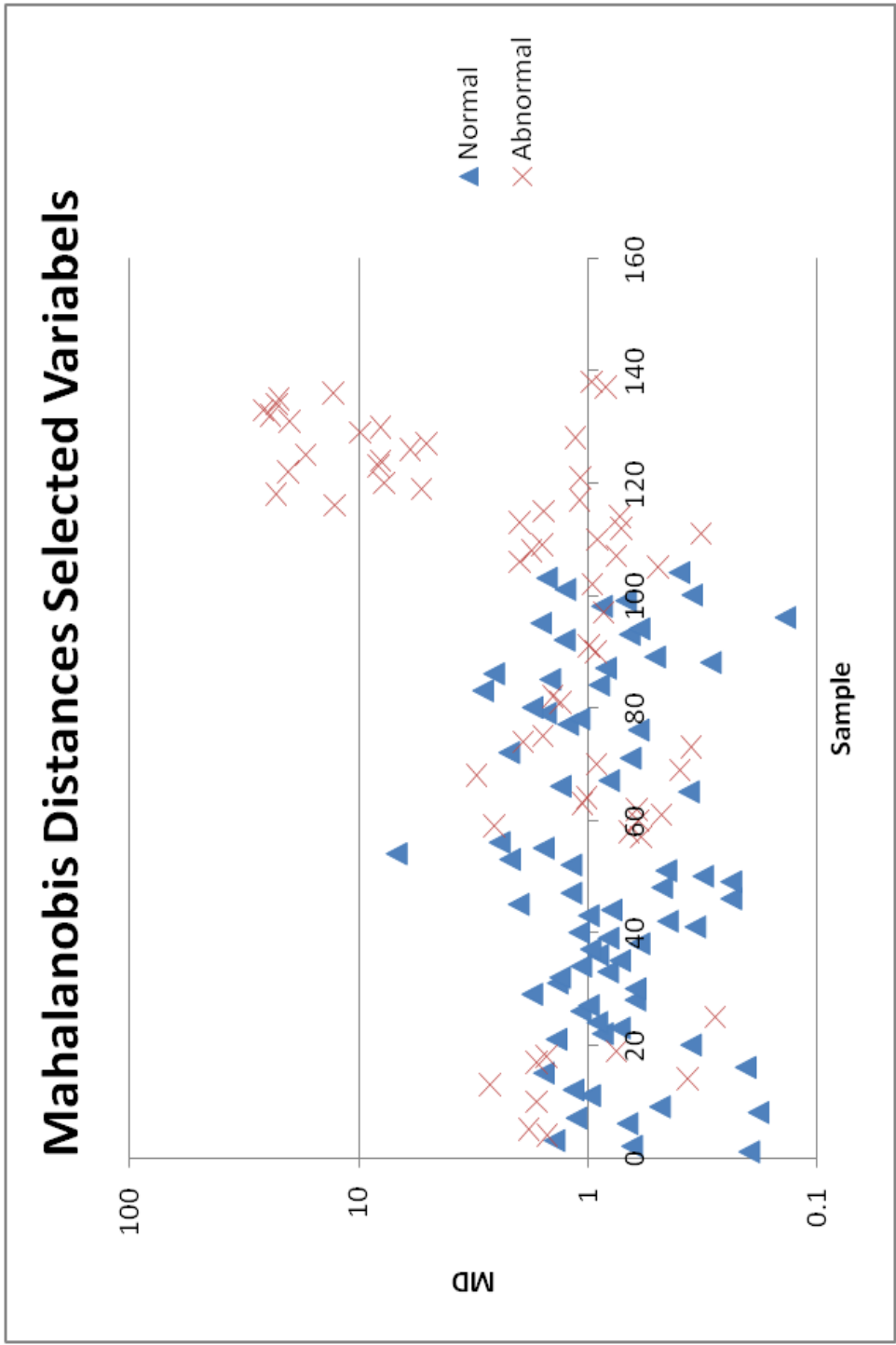


Figure 15 Mahalanobis Distances using Selected Variables

Comparison with the Logistic Regression Solution

Casting Defect Dataset have been previously studied in İşkol et al. (2005) and Erdural (2006) using Logistic Regression (LR). The optimal factor levels in these two studies compared to the results of this study are given in Table 24.

Table 24 Comparison of the proposed factor levels

	İşkol et al. (2005)	Erdural (2006)	Proposed Method
X1	-0.86	0.25	-0.860
X2	-1.19	-1.19	-0.874
X3	-1.68	Insignificant	-0.710
X4	Insignificant	Insignificant	Insignificant
X5	-1.6	-1.6	-0.847
X6	-1.68	-1.68	Insignificant
X7	1.97	1.97	1.530

It is seen that there are differences in the solutions. In the solution of İşkol et al. (2005), variable X4 is found insignificant. In Erdural (2006) work, the variables X3 and X4 are declared as insignificant. Using MTS, X4 and X6 are found insignificant and are not included into the model.

It is also seen that the variables in İşkol et al. (2005) and Erdural (2006) reached their limits. In the proposed method, 2 variables, X1 and X7 do take boundary values.

As the parameter levels are different in 3 of the methods and as there is not a verified optimal solution, it is best to compare the methods by looking at the differences in the classification of the samples.

The classification rate (PCC) of the LR model is calculated as %81.5 using a cutoff point of $c=0.023$. Cutoff point is the probability value that the samples with $P(Y=1)>c$ are considered to be in the 1st class (Erdural 2006) Although the classification rate is high, it is measured differently in Logistic Regression Model than MTS. In the LR model, the classification rate is

calculated using each of the 12574 production runs, not each of the 138 repeated production results. For example, the first row in the dataset is a parameter combination that has been experimented 180 times with 4 defectives. In Erdura (2006), defining this combination as non-defective adds 176 (180-4) samples to correctly classified samples, and 4 samples to incorrectly classified samples. However, it was also explained before that considering each item produced as a separate sample, the dataset achieves defective/total ratio of 220/12574, making the dataset highly imbalanced. Non-defectives produced are 0.983 of all production. Thus, even if all of the samples are classified as non-defective a correct classification rate of %98.3 would be obtained.

In MTS calculations, the samples had to be classified into defective and non-defective classes before the model is built. Therefore, the correct classification rate is measured according to the discretized response of the dataset.

Erdural (2006) formulated probability of having a defective product, $P(Y=1)$ by the formula:

$$\hat{P}(Y = 1) = \exp(-4.32801 + 0.16207(X1) + 0.63291(X2) - 0.05438(X5) - 0.50325(X6) - 0.43786(X7) + 0.36527(X1) * (X6) - 0.33881(X5) * (X6))$$

The probability of obtaining a non-defective product is

$$\hat{P}(Y = 0) = 1 - \hat{P}(Y = 1)$$

These two models can be compared by observing the classification results of the samples. The Figure 16 shows the MDs of the samples versus the probability of being defective ($P(Y=1)$).

In order to observe the results more clearly, the results of Logistic Regression and MTS were both log transformed. The lines on the graph represent the thresholds for both methods. The threshold for MTS was determined to be 1.4. For the Logistic Regression model, the cutoff point was selected as 0.023 by Erdural (2006). Using these thresholds, the samples that have both MD and $P(Y=1)$ smaller than these thresholds are the samples that are agreed to be non-defective by both methods. The following confusion matrix shows the comparison of both methods.

Table 25 Confusion Table of the MTS and LR models

		Logistic Regression	
		Defective	Non-defective
MTS	Defective	14	35
	Non-defective	17	72

The results given in Table 25 are the predictions of both methods. To measure the agreement level of both methods, Cohen's kappa measure can be used. From these results, Kappa=0.6075. Kappa measure ranges from 0 to 1, 1 being perfect agreement. This result shows that there is a moderate agreement between both methods in this case. From the classification table, it is seen that the disagreement mainly arises from the points that are labeled defective by MTS, but are labeled non-defective by the LR model.

If MTS is formed using the variables used in LR, the performance measures are as given in Table 26.

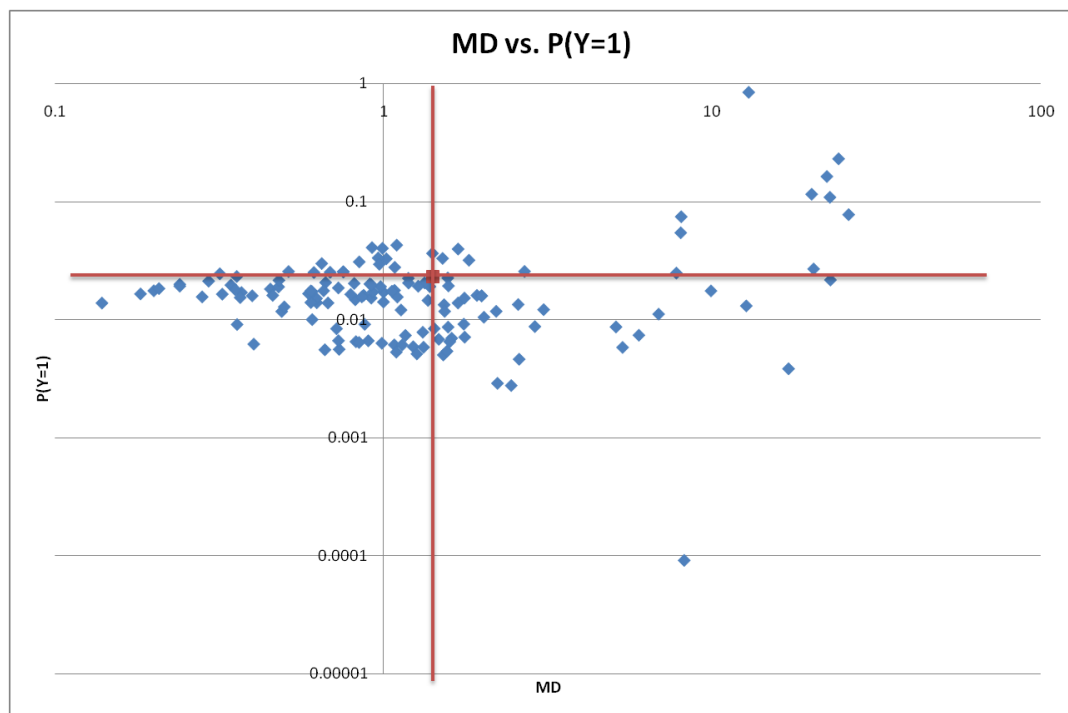


Figure 16 MD vs. P(Y=1) graph of the samples

Table 26 Comparison of the Original MTS and the MTS formed using LR variables

	Original MTS	LR Variables
Threshold	1.41	1.310
PCC	0.681	0.652
KAPPA	0.662	0.631
G-mean	0.694	0.647
Precision	0.657	0.631
Recall	0.54	0.524
Specificity	0.8	0.760
F05	0.656	0.618
F1	0.607	0.579
F2	0.565	0.545
AUC	0.697	0.671

Using X1, X2, X5, X6 and X7, it is seen that there is a slight decrease in all of the performance measures. The classification table in Table 27 shows that there are actually two samples that are classified as defective by MTS, but were originally classified as non-defective. These two samples have thus decreased the classification performance of the model.

Table 27 Confusion table for the MTS (using LR variables) and LR Models

		Logistic Regression	
		Defective	Non-defective
MTS	Defective	15	36
	Non-defective	16	71

After the models are formed, the next step is the determination of the optimal parameter values.

Using the probability of having a defective product as the objective function, and using the variable limits as constraints, a nonlinear programming model was formed by Erdural (2006) as

$$\text{Minimize } f(x) = P(Y = 1)$$

subject to

$$-0.86 \leq x_1 \leq 0.25$$

$$-1.19 \leq x_2 \leq -0.03$$

$$-1.6 \leq x_5 \leq 0.37$$

$$-1.68 \leq x_6 \leq -0.97$$

$$1.53 \leq x_7 \leq 1.97$$

The optimal solution to this Logistic Regression model is given in Table 28.

Table 28 Optimal parameter levels of the LR model

x1	x2	x5	x6	x7
0.25	-1.19	-1.6	-1.68	1.97

These parameter values result in $P(Y=0) = 1 - P(Y=1) = 0.997612$.

It is seen that the optimal values are different in the MTS optimization and LR optimization. As the used variables are different in these methods, the optimal values of a model cannot be inserted into the other to observe if there is any agreement between the methods. Different from LR, MTS has X3 instead of X6.

To compare the two methods, the parameter values of the two cases can be substituted into the each other. MTS and LR have X1, X2, X5 and X7 in common.

In the first case, the optimal values of X1, X2, X5 and X7 are taken from LR and inserted into the MTS model. The optimal levels are found by keeping these variables constant and changing X3.

Optimizing MTS model using LR variables result in the following solution in Table 30.

Table 29 Optimal Parameter Levels of MTS using LR variables

x1	x2	x3	x5	x7
0.250	-1.190	-1.585	-1.600	1.970

The resulting MD is 7.513, which is classified as abnormal by the MTS method.

In the second case, the optimal values of X1, X2, X5 and X7 are taken from MTS and inserted into the LR model. The optimal levels are found by changing X6. The optimal parameter levels are as follows, as given in Table 30.

Table 30 Optimal Parameter Levels of LR using MTS variables

x1	x2	x5	x6	x7
-0.860	-0.874	-0.847	-0.970	1.530

These parameter levels result in $P(Y=0) = 0.994118588$, which is %0.35 worse than the original optimal LR solution.

When the parameter values found using MTS model are inserted into the LR model, the result is only %0.35 less than the optimal parameters of the LR Model. Thus the difference is small. On the contrary, if the parameter values of the LR model are inserted into the MTS

model, the MD found is 7.513, which is a large MD that would be classified as abnormal by the MTS model.

The comparison of the optimal parameter levels of both models are shown in Table 31.

Table 31 Comparison of the Optimal Solutions of MTS and LR

	MTS Model	LR Model
MTS Parameters	0.309	0.994118588
LR Parameters	7.513	0.997611995

It can be seen from the optimal solutions that Logistic Regression is less sensitive to the changes in the variables. Although the MTS parameters result in satisfactory non-defective probabilities in LR, the use of LR variables in the MTS model results in a highly defective case.

Unless the suggested MTS solution is tested in a real setting, it is not possible to reach a strong conclusion about which method is better for this case.

4.2.2. Customer Satisfaction from a Driver's Seat Case

As a second case, the customer satisfaction data from the thesis work of Çabuk (2008) have been used. This dataset includes questions from a questionnaire used to study the factors affecting the satisfaction from the driver's seat. In the questionnaire, there are 75 questions asked to customers. In addition, there are 14 anthropometrical measurements of the customers. In her thesis, Çabuk analyzed the results of the questionnaire using Logistic Regression. This problem does not exactly fit into the area of robust design. Therefore, the function between the control factors and the response is solved inversely, trying to find the factor levels required to reach the desired response class.

Out of a total of 89 variables, Question 63 – Overall Satisfaction with the Vehicle- was used as the response of the questionnaire. The response is binary, where 1 denotes "Somewhat Satisfied" customers and 2 denotes "Highly Satisfied" customers.

17 questions have ordered categorical, and 2 questions have nominal categorical answers. 14 anthropometrical questions have continuous measurements as answers. The remaining 57 variables have binary answers.

In her work, Çabuk removed some of the questions after preprocessing. As a result, 12 categorical variables, 4 continuous variables and 28 binary variables are now present in the data set. The variables are given in Table 32.

Upon further investigation, it is observed that Question 32 and Question 33 are equal to each other for the highly satisfied customers. This means one of them is redundant, and would result in singularity in taking the inverse of the covariance matrix. As Question 33 is used in the Logistic Regression model, Question 32 is removed.

The dataset contains 21 "Somewhat satisfied" customers and 51 "Highly Satisfied" customers. Here, the reference class that would be used to form the normal group is selected to be "Highly Satisfied" customers.

2 nominal categorical answered questions, question 4 and 12, are also removed prior to forming MTS model. In order to calculate Mahalanobis distances using a nominal variable with k levels, k dummy binary variables have to be used, each representing a level. However, in datasets where the number of variables is large compared to the number of samples, addition of new variables may result in the number of variables exceeding the number of samples, preventing the calculation of the Mahalanobis Distances. Even if not, it was shown that a low ratio of sample size/variable causes the sample correlation matrix to be falsely estimated. (Young 1978)

Table 32 Variables of the Customer Satisfaction Dataset

Variables	
Question1	Question31
Question2	Question32
Question3	Question33
Question4	Question34
Question5	Question37
Question6	Question38
Question7	Question39
Question8	Question41
Question10	Question42
Question11	Question43
Question12	Question44
Question13	Question45
Question14	Question46
Question15	Question47
Question16	Question48
Question18	Question53
Question22	Question56
Question24	Question58
Question26	Question76
Question27	Question78
Question28	Question85
Question29	Question87

In the Customer Satisfaction case, question 4 has 3 levels and question 12 has 4 levels. Although dummy variables can be included without exceeding the number of samples, these questions have been removed from the dataset as they are not used in the Logistics Regression model formed by Çabuk.

The reference group was selected as the highly satisfied customers. The aim is to determine if the customers will be highly satisfied or not. There are 51 highly satisfied and 21 somewhat satisfied customers in the dataset.

The properties that form the reference group are the correlation matrix, the mean and the standard deviations of the variables. Thus, these can be examined prior to modeling the dataset with MTS.

The mean values show that in X2, the normal group has a mean significantly larger than the abnormal group. Similarly, in variables X36, X37 and X38, abnormal group has higher means.

Standard deviations in Figure 18 show that high standard deviations make it difficult to conclude that any mean of the variables are different.

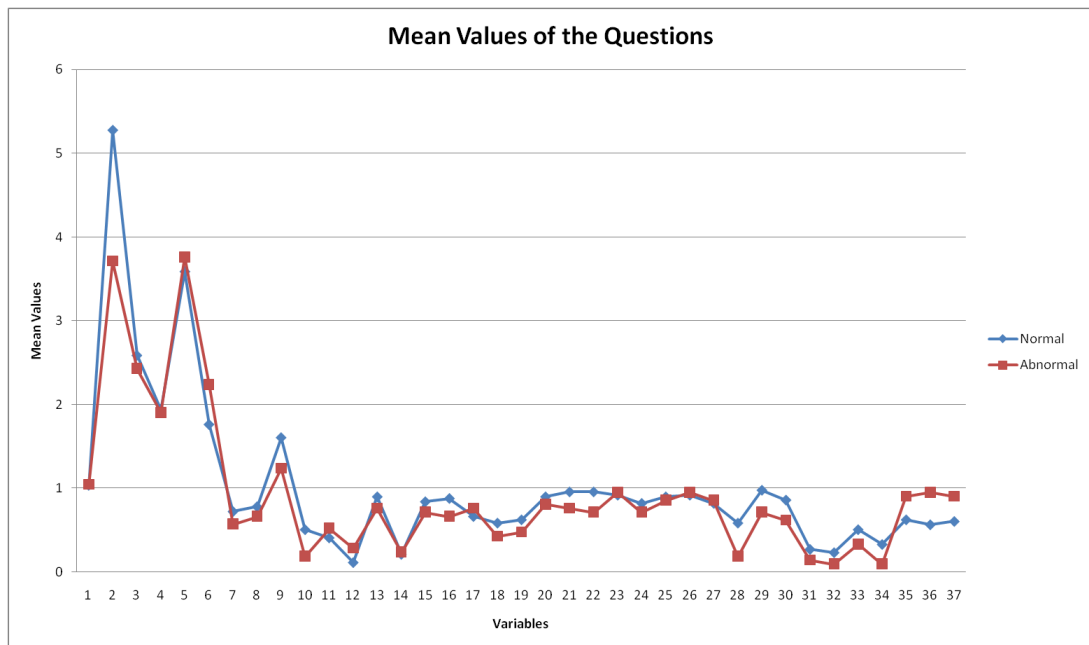


Figure 17 Mean Values of the Questions

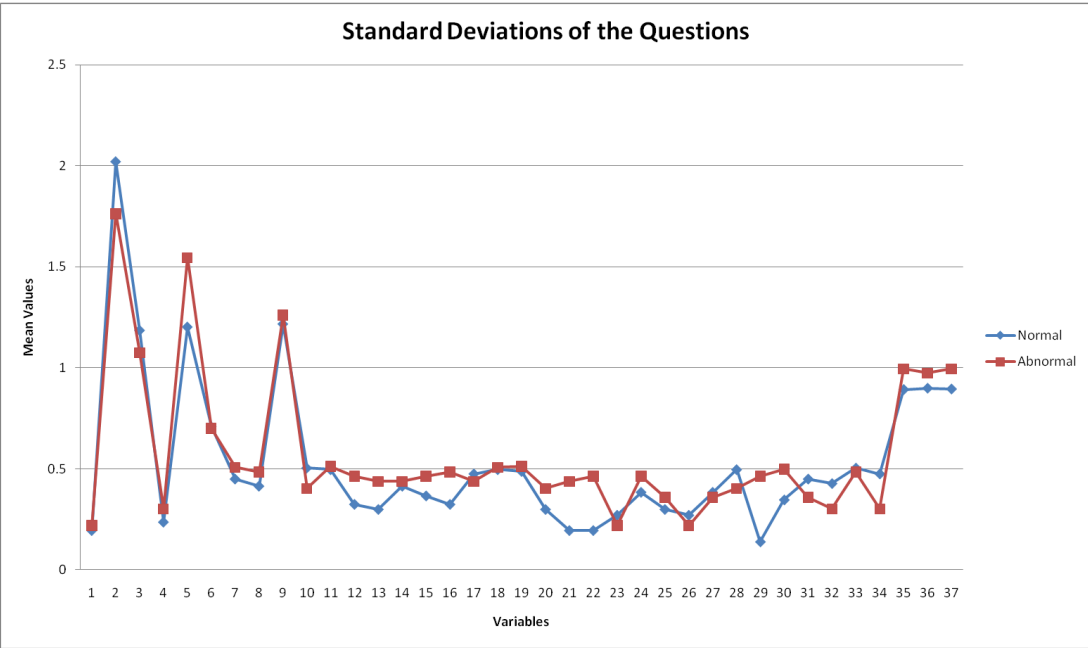


Figure 18 Standard Deviations of the Questions

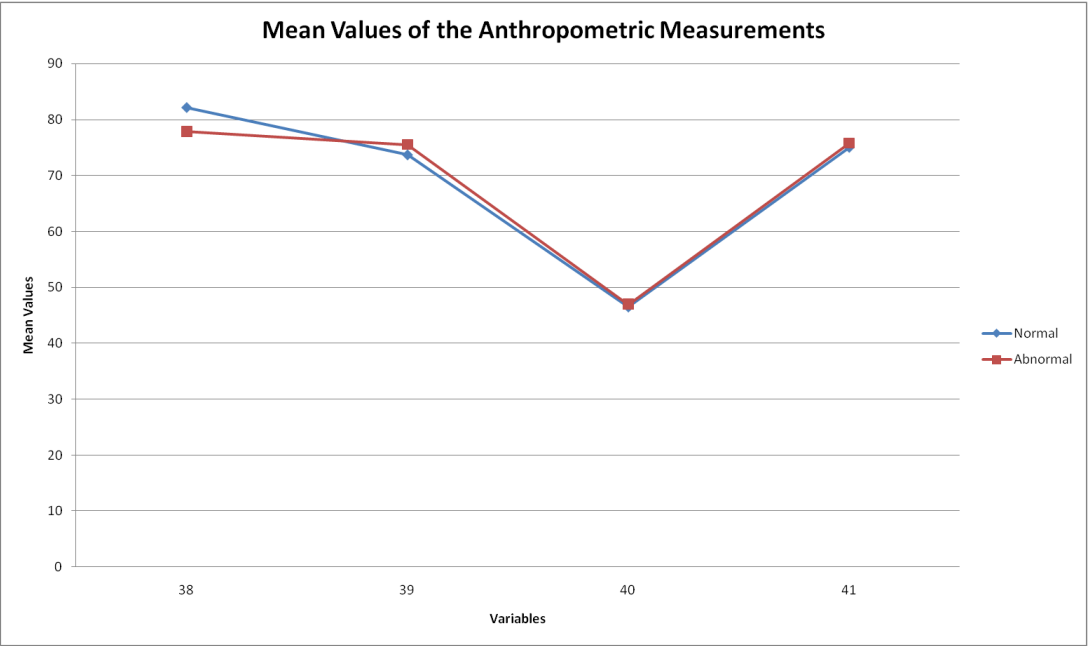


Figure 19 Mean Values of the Anthropometric Measurements

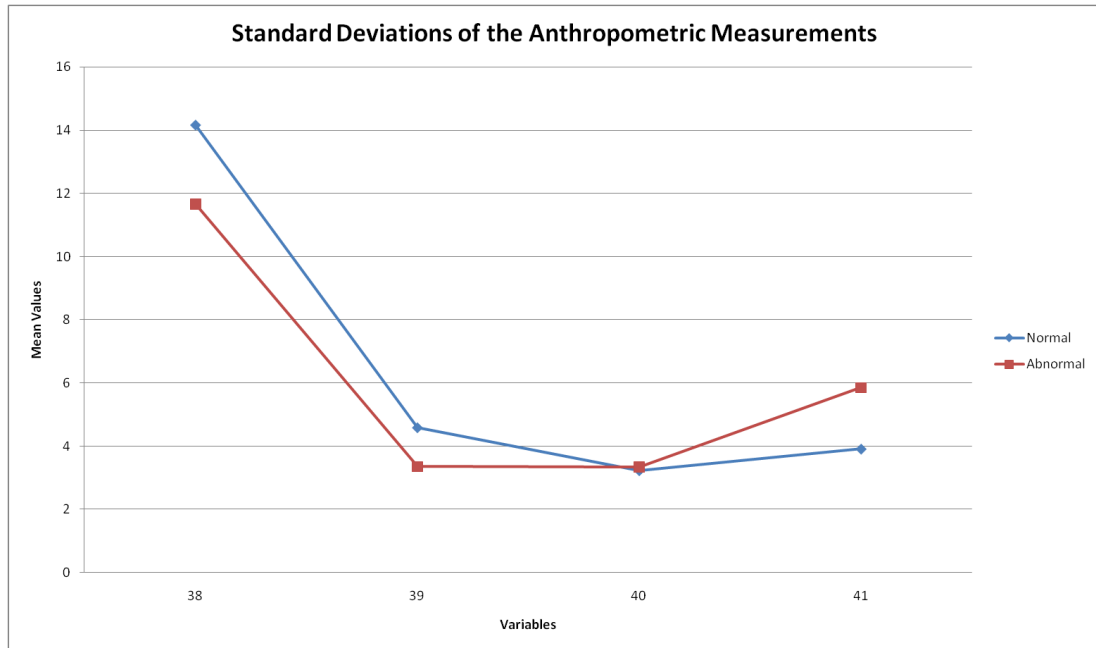


Figure 20 Standard Deviations of the Anthropometric Measurements

Using the remaining 41 variables, the MTS model is formed. As stated before, there are 51 samples in the normal group.

The first step is the removal of outliers from the normal group. Outliers were defined as the samples with MD larger than the mean+3*standard deviation. In this case, none of the samples were found to be outliers.

In the next step, the Mahalanobis Distances of the samples are calculated using all of the variables. Basic statistics of the resulting MDs are given in Table 33.

Table 33 Basic Statistics of the Normal/Abnormal Groups

	Normal	Abnormal
Mean	0.980	21.85548
Standard Deviation	0.124	16.32707
Maximum	1.196	68.73277
Minimum	0.714	1.590844

It can be seen that the mean of the abnormal group is 21.855, a much higher value than the mean of the normal group. The minimum MD of the abnormal group is higher than the maximum of the normal group, indicating that the two classes do not overlap.

The MDs of the normal and abnormal groups are plotted in Figure 21.

From Figure 21, it can also be seen that the abnormal group is separated from the normal group.

To measure how well the classification is done, a threshold value must be selected. The threshold was selected by determining the value that maximizes g-mean. The threshold maximizing g-mean was found to be 1.2 by searching through the possible thresholds.

The performance measure results of the MTS model formed with all the variables is as given in Table 34.

Table 34 Performance Measures of the MTS Model Using All Variables

MTS	
Threshold	1.200
PCC	1
KAPPA	1
Precision	1
G-Mean	1
Recall	1
Specificity	1
F05	1
F1	1
F2	1
AUC	1

Using 41 variables, the MTS model classified all the samples correctly.

In the next step, S/N ratios of different combinations of variables are examined to check if the discrimination power of the model can be improved. Since all the samples were correctly classified using all variables, the expectation is that the classification is at least as

good as the prior one while removing some of the variables. Decreasing the number of variables would decrease the required minimum number of samples for the normal group.

From the ANOVA table in Appendix A, it is seen that there are 23 variables with p value less than 0.1. These 23 variables are determined to be changing the S/N ratio when included in the model. These variables are X3, X6, X10, X11, X12, X15, X16, X17, X18, X20, X21, X22, X25, X27, X28, X29, X30, X32, X33, X34, and X36.

Still, the change may be in the positive or in the negative direction. To determine the direction, the main effects of the variables are plotted. The main effects plots of the variables are provided in Appendix B.

The main effects plot shows that although there are other variables that have positive main effects, they were not found significant by ANOVA. In the end, remaining 23 variables will be used in model forming.

Using the selected 23 variables, the MTS model is constructed. Basic statistics of the calculated MDs in comparison with the 41 variable case are shown in Table 35.

Table 35 Basic Statistics of the two MTS models

	41 Variables		23 Variables	
	Normal	Abnormal	Normal	Abnormal
Mean	0.980	21.85548	0.980	5.152
Standard Deviation	0.124	16.32707	0.344	2.389
Maximum	1.196	68.73277	2.131	10.483
Minimum	0.714	1.590844	0.514	0.962

The results show that the discrimination did not improve in this case. The maximum value of the normal group increased while the minimum value of the abnormal group decreased below the mean of the normal group. It is clear that the abnormal group got closer to the normal group. Figure 22 shows the MDs of the normal and abnormal groups using the 23 variables.

Table 36 shows the performance measures of the two cases.

Table 36 Comparison of two MTS models

	41 Variables	23 Variables
Threshold	1.200	2.140
PCC	1	0.972
KAPPA	1	0.971
Precision	1	1.000
G-Mean	1	0.951
Recall	1	0.905
Specificity	1	1.000
F05	1	0.979
F1	1	0.950
F2	1	0.922
AUC	1	0.975

It is seen that all performance measures decreased when the variables are reduced. This reduction is caused by the false classification of 2 abnormal samples.

In the literature, it was discussed that the use of OAs did not always result in the optimal selection of variables. (Abraham and Variyath 2003). The allocation of the variables to different columns of an OA did result in different S/N ratios. Jugulum also pointed out this situation as a future research topic. This issue can be a reason of the decrease in the performance measures.

The second possible reason may be the use of ANOVA. It was seen in the main effects plot that some variables that had a positive gain were not included into the model as their effect were not found significant by ANOVA. Whether this is a reason for the performance decrease or not can be tested by using only the main effect plots for the determination of the variables to include.

Using only the main effect plot given, 34 variables are selected. The basic statistics are shown in Table 37.

Table 37 Basic Statistics of the MTS model with 34 Variables

	Normal	Abnormal
Mean	0.980	7.759
Standard Deviation	0.207	4.034
Minimum	1.442	15.095
Maximum	0.612	0.839

It is seen that the abnormal class got closer to the normal class. Figure 23 shows the MDs of all samples. It is observed that although the abnormal class got closer to the normal class, there is still a separation between the classes.

The performance of the MTS model using these 34 variables is shown in Table 38.

Table 38 Performance Measures of the MTS model with 34 variables

34 Variables	
Threshold	1.450
PCC	0.986
KAPPA	0.986
Precision	1.000
G-Mean	0.976
Recall	0.952
Specificity	1.000
F05	0.990
F1	0.976
F2	0.962
AUC	0.966

It is seen that 11 variables that had small positive gain but were not found significant by ANOVA did in fact increase the performance of the classifier. The increase in recall is the result of correctly classifying an additional abnormal sample.

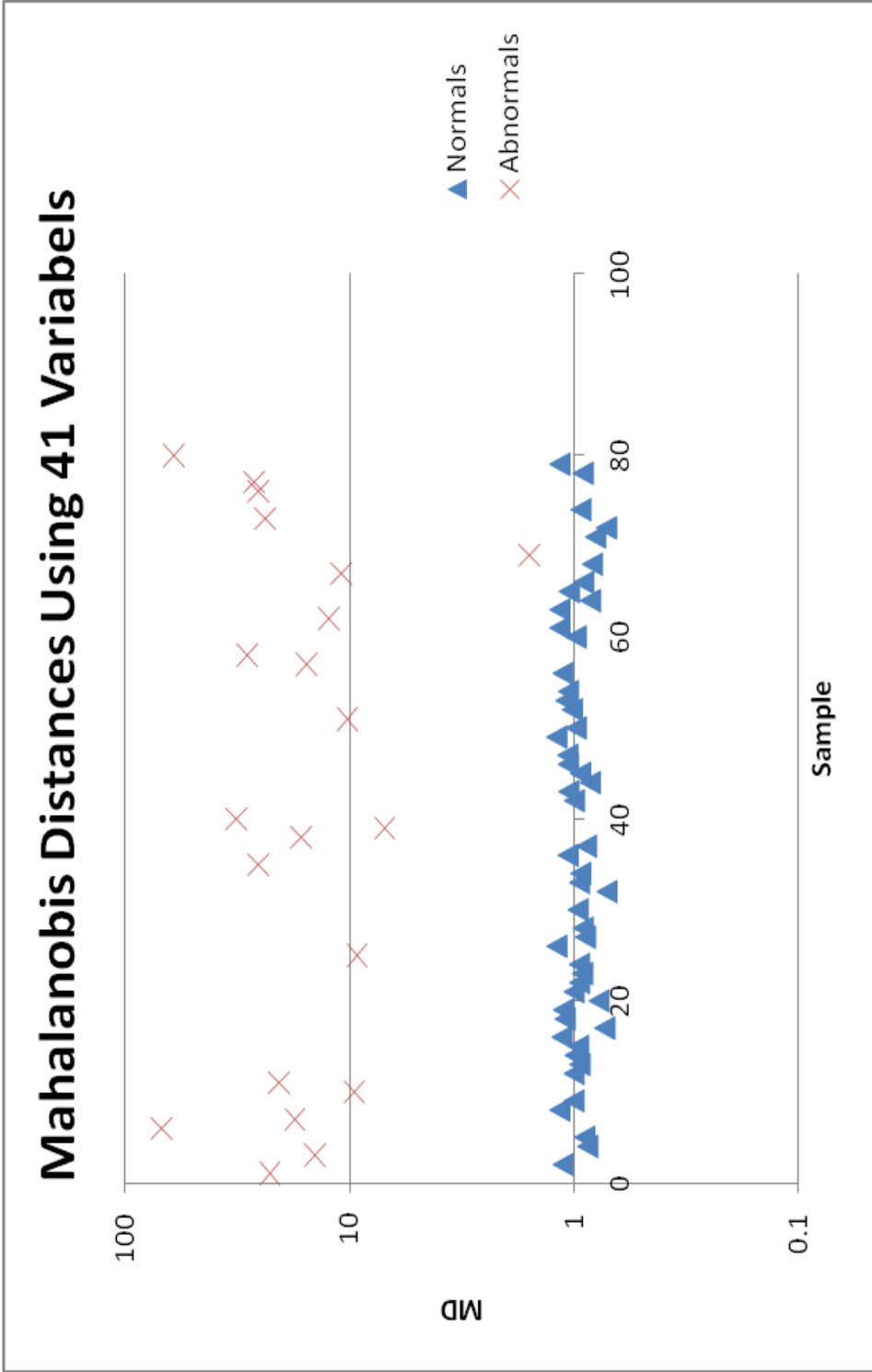


Figure 21 Mahalanobis Distances Using 41 Variables

Comparing all 3 cases, it is clear that it was not possible to reduce the number of variables used using S/N ratios or Orthogonal Arrays. The best performance is obtained when all variables are included into the model.

The MTS model will then be used for optimization. For this purpose, a quadratic programming model has been set up. The model is

$$\text{Minimize } f(\mathbf{x}) = \frac{1}{p} (\mathbf{x} - \mathbf{m})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mathbf{m})$$

subject to

$x_1, x_2, x_3, x_5, x_6, x_7, x_8, x_{10}, x_{11}, x_{13}, x_{14}, x_{15}, x_{16}, x_{18}, x_{22}, x_{24}, x_{26}, x_{27}, x_{28}, x_{29}, x_{31},$

$x_{33}, x_{34}, x_{37}, x_{38}, x_{39}, x_{41}, x_{42}, x_{43}, x_{44}, x_{45}, x_{46}, x_{47}, x_{48}, x_{53}, x_{56}, x_{58}, x_{76}, x_{78}, x_{85}, x_{87}$

$x_8, x_{10}, x_{13}, x_{14}, x_{15}, x_{16}, x_{18}, x_{22}, x_{24}, x_{26}, x_{27}, x_{28}, x_{29}, x_{31}, x_{33}, x_{34}, x_{37}, x_{38}, x_{39}, x_{41},$

$x_{42}, x_{43}, x_{44}, x_{45}, x_{46}, x_{47}, x_{48} : \text{binary}$

$x_1, x_2, x_3, x_6, x_7, x_{11}, x_{53}, x_{56}, x_{58} : \text{integer}$

$x_{76}, x_{78}, x_{85}, x_{87} : \text{continuous}$

$$1 \leq x_1 \leq 2$$

$$1 \leq x_2 \leq 9$$

$$1 \leq x_3 \leq 4$$

$$1 \leq x_6 \leq 7$$

$$1 \leq x_7 \leq 3$$

$$0 \leq x_{11} \leq 3$$

$$0 \leq x_{53} \leq 2$$

$$0 \leq x_{56} \leq 2$$

$$0 \leq x_{58} \leq 2$$

$$50 \leq x_{76} \leq 125$$

$$62.1 \leq x_{78} \leq 82.4$$

$$37.5 \leq x_{85} \leq 56$$

$$64 \leq x_{87} \leq 87.5$$

M = means of the variables

Σ = covariance matrix

Means of the variables are calculated as given in Table 39.

Table 39 Mean Values of the Variables

X1	X2	X3	X5	X6	X7	X8	X10	X11
1.039	5.275	2.588	1.941	3.588	1.765	0.725	0.784	1.608
X13	X14	X15	X16	X18	X22	X24	X26	X27
0.510	0.412	0.118	0.902	0.216	0.843	0.882	0.667	0.588
X28	X29	X31	X33	X34	X37	X38	X39	X41
0.627	0.902	0.961	0.961	0.922	0.824	0.902	0.922	0.824
X42	X43	X44	X45	X46	X47	X48	X53	X56
0.588	0.980	0.863	0.275	0.235	0.510	0.333	0.627	0.569
X58	X76	X78	X85	X87				
0.6078	82.1157	73.6569	46.4745	74.9976				

Solving this QP, the optimal parameter values are found are given in Table 40.

The optimal MD of 0.424 is obtained with these parameters. With the threshold of 1.45, this point is within the normal group and is close to the centroid of the normal group.

Mahalanobis Distances Using 23 Variabels

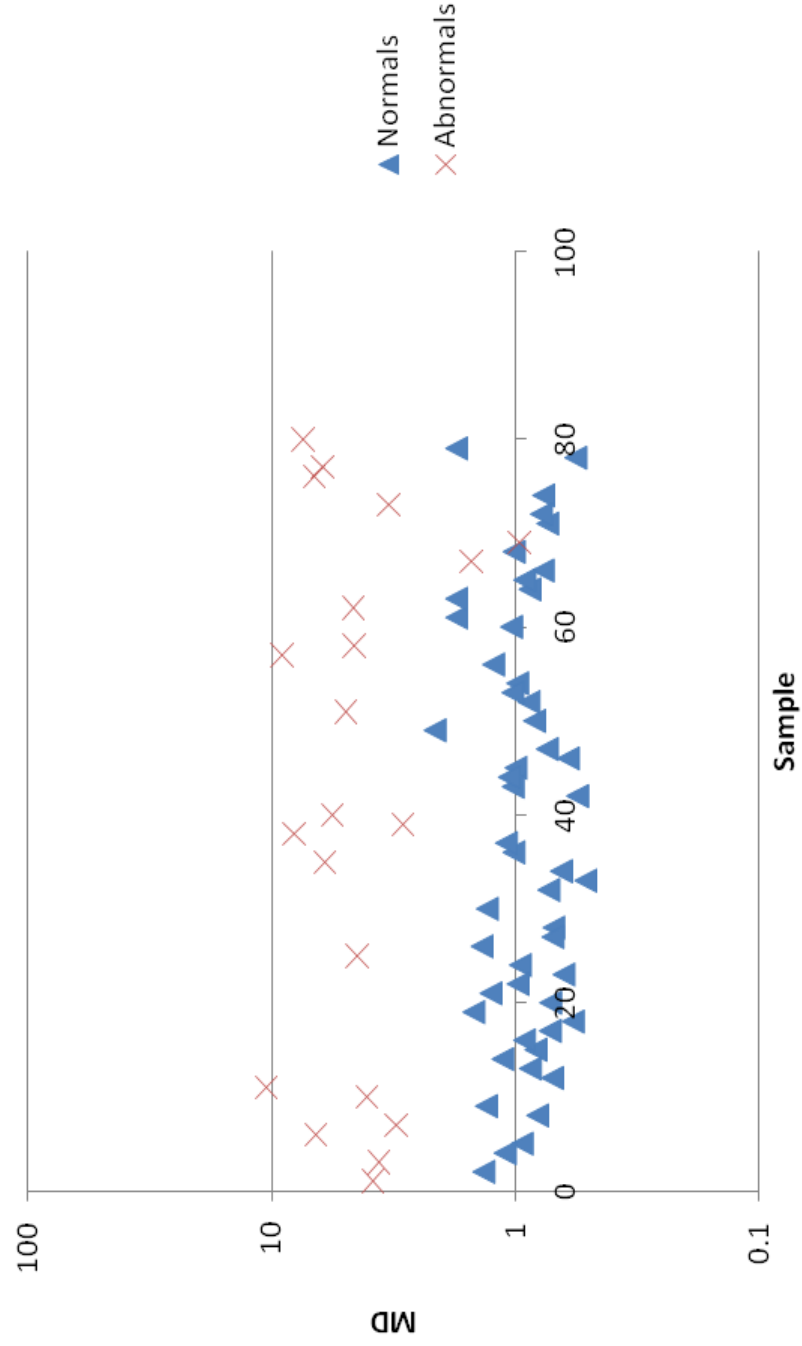


Figure 22 Mahalanobis Distances Using 23 Variables

Mahalanobis Distances Using 34 Variabels

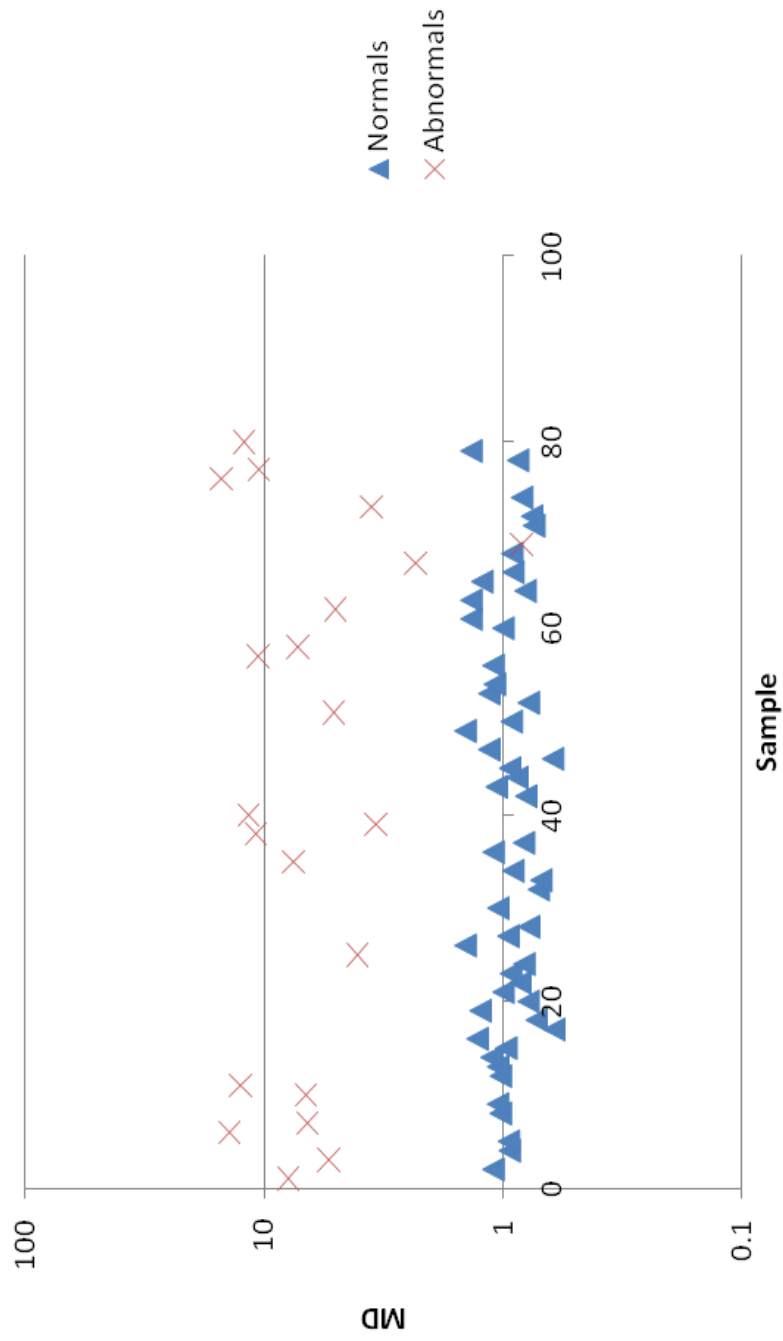


Figure 23 Mahalanobis Distances Using 34 Variables

Table 40 Optimal Parameter Levels of the MTS model

X1	X2	X3	X5	X6	X7	X8	X10	X11
1	5	1	2	4	1	1	1	2
X13	X14	X15	X16	X18	X22	X24	X26	X27
1	1	0	1	0	1	1	1	0
X28	X29	X31	X33	X34	X37	X38	X39	X41
1	1	1	1	1	1	1	1	1
X42	X43	X44	X45	X46	X47	X48	X53	X56
0	1	1	0	0	1	1	0	0
X58	X76	X78	X85	X87				
0	86.53	70.74	46.87	76.76				

The MTS model formed can also be used to find the point that is furthest from the centroid of the normal group. In order to do this, the QP model is maximized. The results are provided in Table 41.

Table 41 Parameter Levels for Maximum Distance From the Normal Group

X1	X2	X3	X5	X6	X7	X8	X10	X11
1	9	1	2	1	1	1	0	0
X13	X14	X15	X16	X18	X22	X24	X26	X27
1	1	0	0	0	0	1	0	1
X28	X29	X31	X33	X34	X37	X38	X39	X41
1	1	0	1	1	0	1	1	0
X42	X43	X44	X45	X46	X47	X48	X53	X56
1	1	1	1	1	0	0	0	2
X58	X76	X78	X85	X87				
2	125	83	37.5	64				

These parameters result in a MD of 727 to the normal group. Although this is a very distant point to the normal group, it is not exactly possible to state that these parameters would lead to the least satisfaction level. As explained in Appendix D, it is not possible to assess the direction of the point relative to the normal group. Mahalanobis Distance is a dimensionless scale, and the point may be abnormal in either "good" or "bad" way.

Therefore, if it is desired to find the parameter levels with least satisfaction, it is suggested that the dataset is remodeled using the less satisfied customers as the reference group.

Comparison to Logistic Regression Solution

Originally, this dataset was modeled by Çabuk (2008) using Logistic Regression in her thesis. The model formed is

$$\hat{P}(Y = 2) = 1 - \frac{e^{g(x_1, \dots, x_9)}}{1 + e^{g(x_1, \dots, x_9)}}$$

where

$$g(x) = \{3.517 + 4.147(1 - x_{13}) - 3.157x_{24} + 1.882(1 - x_{27}) - 5.079x_{33} - 4.055x_{42} - 6.015x_{43} + 2.393x_7 - 1.046x_8 + 1.101x_9\}$$

$\hat{P}(Y = 2)$ is the probability that the customer will be "Highly Satisfied". Out of 44 variables, 9 variables entered the model.

Using the $\hat{P}(Y = 2)$ equation, the samples with probability greater than the specified cutoff level are labeled as "Highly Satisfied" and the samples with probability less than the cutoff point are labeled as "Somewhat Satisfied". This cutoff point can be determined using the dataset, or it can be set to a dataset independent value. In this case, cutoff point is selected as 0.5.

According to this cutoff point, the performance measures of Logistic Regression model compared to the MTS results are shown in Table 42.

Using a cutoff point of 0.5, correct classification rate of the model is %93.1. Area Under ROC Curve (AUC) is 0.896, which shows the model is very good in the sense of prediction. Highly satisfied customers were predicted %92.3 percent correctly while less satisfied customers were predicted %90.5 correctly. From the obtained performance measures, it is seen that MTS performed better in classifying the customers.

Table 42 Performance Measure of the LR and the MTS models

	Logistic Regression (Cutoff=0.5)	MTS
PCC	0.931	1.200
KAPPA	0.927	1
Precision	0.864	1
Recall	0.923	1
Specificity	0.905	1
F05	0.941	1
F1	0.872	1
F2	0.884	1
AUC	0.896	1

As MTS correctly classified all the samples, the performance of the LR model is actually tested on itself. The confusion matrix of the LR model is given in Table 43.

Table 43 Confusion matrix for the LR model

		Logistic Regression	
		Somewhat Satisfied	Highly Satisfied
Actual	Somewhat Satisfied	19	3
	Highly Satisfied	2	48

It is seen that 3 abnormal have been classified as normal, and 2 normals have been falsely classified as abnormal. The graph below shows this agreement between the methods. Out of 55 highly satisfied customers, 48 were correctly classified by both methods. The results are also shown in Figure 24. The red lines indicate the thresholds used in both methods.

Çabuk (2008) used this Logistic Regression model to optimize the parameters so that the probability of high satisfaction of maximized. The model is

$$\text{Maximize } \hat{P}(Y = 2) = 1 - \frac{e^{g(x_1, \dots, x_9)}}{1 + e^{g(x_1, \dots, x_9)}}$$

Subject to

$$g(x) = \{3.517 + 4.147x_{13} - 3.157x_{24} + 1.882x_{27} - 5.079x_{33} - 4.055x_{42} - 6.015x_{43} + 2.393x_7 - 1.046x_8 + 1.101x_9\}$$

$$x_8 = \text{Std76} = \frac{x_{76} - 81.14}{13.96}$$

$$x_9 = \text{Std85} = \frac{x_{85} - 46.59}{3}$$

$$x_{13}, x_{24}, x_{27}, x_{33}, x_{42}, x_{43} = 0,1$$

$$x_7 = 1,2,3$$

$$50 \leq x_{76} \leq 125$$

$$37.5 \leq x_{85} \leq 56.5$$

$$0 \leq \hat{P}(Y = 2) \leq 1$$

The optimal parameter levels for this model are given in Table 44.

Table 44 Optimal Parameter Levels of the LR model

X13	X 24	X27	X 33	X 42	X 43	X 7	X 76	X 85
1	1	1	1	1	1	1	125	37.5

The resulting optimal solution obtained from these parameters is:

$$P(Y=2) = 0.99994677111614200$$

In MTS and LR, the variables used are different. In MTS all the variables are used, but only 9 variables are present in the LR model. As the variables were different in MTS and LR, the optimal parameter levels can be compared in different ways.

If the levels of 9 variables used in LR model are taken from the optimal MTS solution, the resulting probability of high satisfaction is

$$P(Y=2) = 0.98152235292466600$$

This result is %1.84 less than the optimal solution found in the LR solution.

Next, MTS model is optimized, using the optimal parameter levels of the optimal LR solution. The optimal values are shown in Table 45.

The optimal MD, fixing the common variables in this case is 1.37. As the defined threshold is 1.2, this MD value signals a parameter combination that is slightly outside the high satisfaction region.

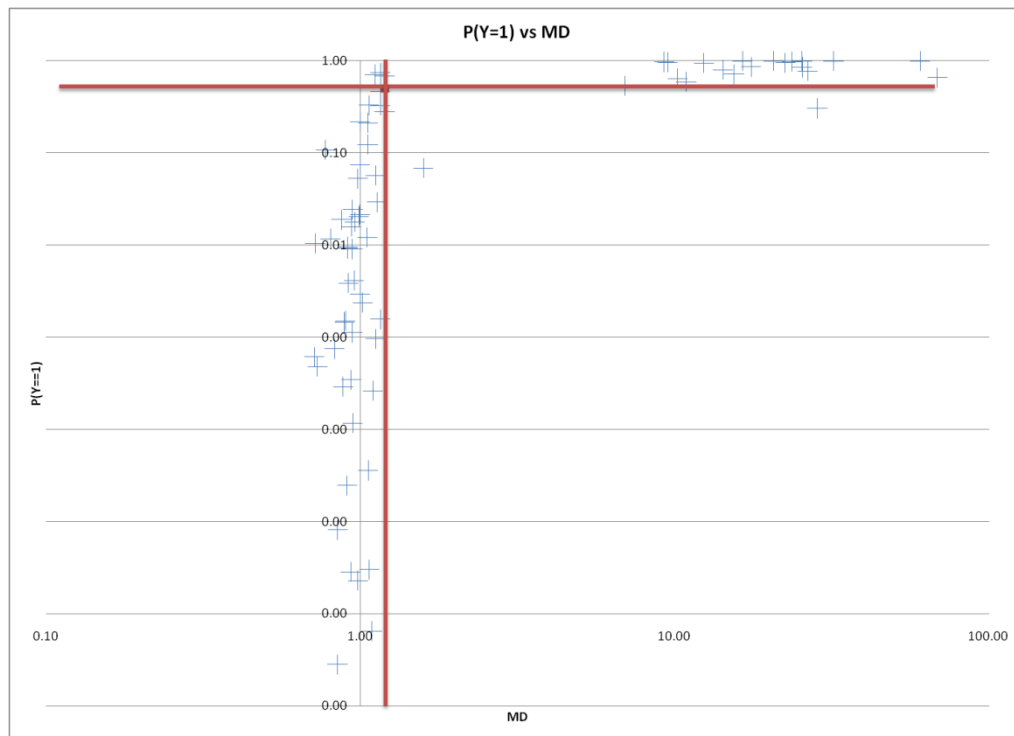


Figure 24 P(Y=1) vs. MD graph for the observations

Table 45 Optimal parameter levels of the MTS model using optimal LR solution

X1	X2	X3	X5	X6	X7	X8	X10	X11
1	6	3	2	4	1	1	0	3
X13	X14	X15	X16	X18	X22	X24	X26	X27
1	0	0	1	0	1	1	0	1
X28	X29	X31	X33	X34	X37	X38	X39	X41
0	1	1	1	1	1	1	1	1
X42	X43	X44	X45	X46	X47	X48	X53	X56
1	1	1	0	0	1	1	1	0
X58	X76	X78	X85	X87				
0	125	67.53	37.5	76.35				

Next, the MTS model is formed using the variables of LR. The performance measures of the resulting model are given in Table 46.

Table 46 Performance measures of the MTS model formed with LR variables

MTS with LR Variables	
Threshold	1.530
PCC	0.875
KAPPA	0.870
Precision	0.800
Recall	0.838
Specificity	0.762
F05	0.922
F1	0.792
F2	0.780
AUC	0.769

As expected, correct classification rate has decreased from 100% to 87.5%. 4 normals have been misclassified as abnormal and 5 abnormal have been misclassified as normal.

Optimizing this MTS model results in the optimal variables provided in Table 47.

Table 47 Optimal parameter levels for the MTS model formed using LR variables

X7	X13	X24	X 27	X 33	X42	X 43	X 76	X 85
2	0	1	1	1	1	1	81.23	45.04

The optimal Mahalanobis Distance is found to be 0.296. With the threshold of 1.53, this solution is also within the normal group. When these parameter values are inserted into the LR model, $P(Y=2) = 0.99970768763702600$. This probability is only %0.024 smaller than the probability obtained by optimizing the LR model.

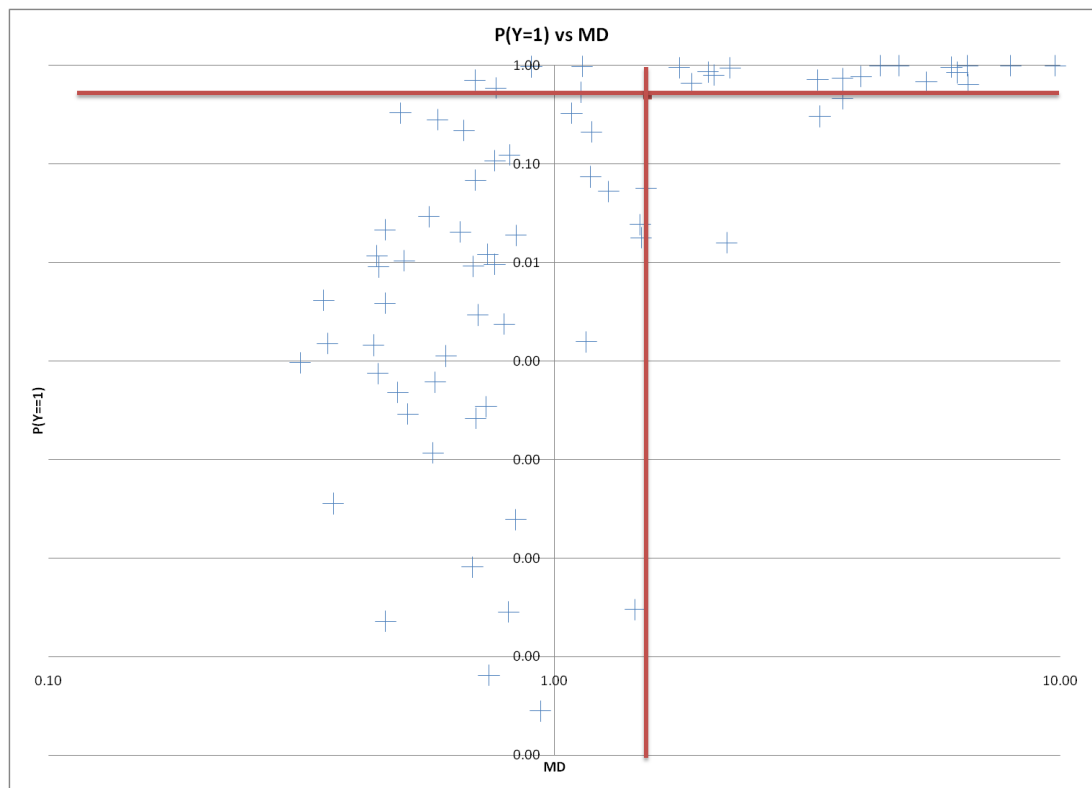


Figure 25 P(Y=1) vs. MD graph for the observations

From these results, it is seen that in classification of the customers, MTS performed well. All the samples in the dataset were correctly classified using the MTS model. In the optimization phase, it was possible to obtain parameter combinations that are close to the centroid of the group. Inserting the solution found in MTS into the LR model also resulted in high probability of satisfaction. However, it was seen that MTS model was more sensitive to changes in the variables. This sensitivity mainly results from the anthropometric variables. LR is less sensitive to changes in the anthropometric variables. The most significant difference between the optimal solutions of the MTS and the LR model are observed in X76. This variable takes the upper bound, 125 in the optimal LR solution and 86.5 in the optimal MTS solution. In the LR model, the difference in high satisfaction probability $P(Y=2)$ when assigning the lower bound of 50 and the higher bound of 125 to X76 while keeping the other variables constant is -0.0003%. However, changing the value of X76 from 86.5 to 50 and 125 in the MTS model changes the solution 1118.3% and 1240.4% respectively.

4.3. Comparison of the Proposed Method and Other Robust Design Methods

Different methods have been developed for modeling and optimizing based on data with qualitative responses. Some of these methods have been summarized in Section 2.7. In addition, the performance of the proposed method was compared to that of the logistic regression in the two case studies provided.

From the studies, it was seen that MTS is particularly suitable for robust design with a binary categorical response. However, as MTS uses the properties of the reference group to form the model, selection of the reference group is very important. It is required that the reference, or the normal, group is originated from the same distribution and it is as uniform as possible. This requirement is not present for the observations outside the normal group, namely the abnormal. If the wrong group is chosen as the reference, then the classification performance may be limited. Therefore, it is more appropriate that the normal class is

formed using the non-defective samples in a defective vs. non-defective case. Although there can be different types of defects, the non-defective samples are more likely to be uniform.

This problem was observed in the Casting Defect Case. The data consisted of measurements from different production cases. As the aim of the original experiment was to reduce the nitrogen gas defects in the casting process, the outcome of the experiments were selected as the presence of nitrogen defects. However, there was no information on whether the group that has no nitrogen defects is actually free of other defects or not. From the classification results, it is seen that the classification performance of MTS is not well. One of the main results of this is that the two responses are taken as the number of defectives produced and the total produced. This is not from a designed experiment, and the parameter combinations have been tested different number of times. For the parameter combinations with small number of runs, the actual defect rates cannot be estimated accurately from this number of runs. Even though that combination might be actually good, the effect of total produced was taken into consideration using hypothesis testing. However, it is seen from the classification performance at the end that either the variables were not properly representing the response, or the separation of the samples into two groups was not very successful.

At the optimization phase, it was seen that the optimal solution obtained with the MTS model did also result in low defect probability when inserted into the LR model. The LR model was less sensitive to changes in the variables, and the MTS solution did also result in a good solution. This shows that although the classification performance of the MTS model is not as good as the LR model, MTS could result in good optimal solutions.

In the Customer Satisfaction case, the response is binary. It was seen that the classification performance of the MTS model is better than the LR model. A clear definition of the normal class (highly satisfied customers) did result in good classification of classes.

Although the performance measure of MTS is better in the Customer Satisfaction case, the number of variables present in the model is much higher than the number of variables in the LR model. There are 41 variables in the MTS model, but there are only 7 in the LR model. The increase in the number of variables increases the computational effort in the

optimization phase. From the experiments, it was observed that when the number of variables exceeds 20, the quadratic model cannot be solved using the standard Excel solver add-in. Other mathematical modeling programs specialized in QP problems have to be used. Optimization of the LR model requires much less computational effort.

As the core of MTS method is the MD, which is a dimensionless metric, it is not possible to examine the direction of an observation relative to the reference group. Therefore, it is possible to say that location and dispersion effects cannot be analyzed separately using MTS. Although Jugulum (2000) aimed to use Mahalanobis Taguchi Gram-Schmidt (MTGS) method for this purpose, this work has been highly criticized.

Accumulation Analysis (AA) is also a method introduced by Taguchi (1991) for the analysis and optimization of ordered categorical data. In AA, the cumulative response categories are determined and the effects of factor levels on the probability distributions by the categories are calculated using an ANOVA-like procedure. Using this information, the optimal factor levels can be calculated.

Accumulation Analysis has been widely criticized by Nair (1986) and Hamada and Wu (1986) for being inefficient and complicated. The use of this method in categorical data is not suggested (See Section 2.7.2).

Scoring Schemes method assigns two scores to the categories and tries to identify the location and dispersion effects separately. Nair (1986) As these effects are considered separately, two different optimal factor settings for location and dispersion effects are calculated. This limits the use of scoring schemes, for it can be difficult to choose factor settings using two different optimal factor calculations obtained for location and dispersion effects.

Weighted Signal-to-Noise Ratio (WSNR) is another method developed by Taguchi (1991). In this method, a weight proportional to the quality loss is assigned to each response category. The categorical response is used as it was continuous and smaller-the-better SNR is used. The optimal parameter levels are the ones that results in the maximum SNR. This method is preferred to AA and Scoring Schemes method as one single optimization solution can be obtained.

In their work, Wu and Yeh (2006) apply Taguchi methods on a continuous production case study data for parameter optimization. Then they categorize the variables and use AA, WSNR and two different Scoring Schemes methods to obtain optimal parameter levels. These results from the categorized data are compared to those of the continuous data. It is concluded that WSNR method is both simpler and more efficient than the other methods.

Accumulation analysis, Scoring Schemes and WSNR methods are all methods developed solely to be used with data with categorical independent variables and categorical response. It is not possible to use a continuous variable in these methods before categorizing. This is a loss of information that is present. In addition, the optimal parameter levels obtained are also limited to the categorical data; it is not possible to obtain a continuous variable level.

MTS can be preferable to these methods as different types of independent variables can be used in forming the model. One clear disadvantage of MTS to these methods is the sample size limit. Due to the formula of Mahalanobis Distance, the sample size must be at least as many as the number of variables.

The use of OAs in MTS is also criticized. It was also shown in the case studies that the use of OAs together with SNR did not always give the best model.

Taguchi and Jugulum (2000) state that there are no assumptions present for the MTS method. However, Mahalanobis distance metric used in MTS is originally developed with the assumption of normality of the groups. Although there is a normality assumption, this assumption is not taken into consideration in the MTS literature without any reported problems.

Most important limitation of MTS comes from the use of correlation matrix in the MD formula. In order to calculate MD, there must be satisfactory number of samples to calculate the correlations between the variables, and the variables must not be highly correlated. The variables with high correlation must be removed prior to model forming. If the sample size is not enough, either additional samples must be collected or some of the variables must be removed.

CHAPTER 5

CONCLUSION & FUTURE WORK

In this study, an optimization method for robust design with a binary response is proposed. The proposed method uses Mahalanobis Taguchi System to form a classification model, and optimizes the model using quadratic programming. Mahalanobis Taguchi System is a multivariate classification method that is specifically useful for experiments with binary responses (or two classes)

The proposed method was applied into two different datasets. These datasets have been previously studied using Logistic Regression, and thus the performance of MTS and LR in classification and optimization could be compared.

In the proposed method, ANOVA was used in variable selection process in combination with main effects calculations. A threshold selection method was also proposed, in order to select a threshold that would be suitable for the performance measures found important.

Quadratic Programming was used in the parameter optimization process. Excel Premium Solver Add-In in combination with the XPRESS solver engine was used. Different programs designed for quadratic programming models can also be used.

The first dataset studied was the Casting Defect Dataset. The number of variables was 7, allowing the trial of all variable combinations. Thus a full factorial design was used. Due to the fact that the response could not be accurately categorized into two classes, the classification performance of MTS was not satisfactory. The variables entering the model and the optimum levels were different from two previous studies. However, it was also observed that the optimal levels obtained using the MTS model did result in low defect probabilities when entered into the LR model.

The second dataset studied was the Customer Satisfaction Dataset. This data includes survey questions asked to customers and the anthropometric measurements of the customers. It was observed in this dataset that the solution obtained using OAs did not always result in the best model. Using all the variables, MTS classified all the customers correctly. The optimal solutions obtained using MTS did also result in high satisfaction probabilities when inserted into the LR model. However, LR solution results in a MD slightly outside the normal region.

In the two cases studied, it was observed that there are differences in optimal solutions of the compared methods. A better comparison with other methods can be made by doing verification experiments.

From these cases studied, it was also seen that the variable selection phase of MTS does not always work as intended. As it was pointed out by various authors, the use of OA is problematic as the way the variables are assigned to the OA columns changes the calculated factor effects. This variable selection phase could be a further study area, where variable selection is made using heuristic algorithms or heuristic methods.

Finally, after studying the use of MTS method in multi-class problems, this study can be extended to robust design problems with more than 2 qualitative responses.

REFERENCES

- Abraham, Bovas, and Asokan Mulayath Variyath. "Discussion." *Technometrics* 45, no. 1 (February 2003): 22-25.
- Ayhan, Dilber. "Multi-Class Classification Methods Utilizing Mahalanobis Taguchi System and a Re-Sampling Approach For Imbalanced Data Sets." Masters Thesis, Industrial Engineering Department, Middle East Technical University, Ankara, 2009.
- Batista, Gustavo E. A. P. A., Ronaldo C. Prati, and Maria Carolina Monard. "A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data." *Sigkdd Explorations* 6, no. 1 (2004): 20-28.
- Box, G, and S Jones. "Discussion of "Testing in Industrial Experiments with Ordered Categorical Data"." *Technometrics* 28, no. 4 (1986): 295-301.
- Bradley, Andrew P. "The Use of Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms." *Pattern Recognition* 30, no. 7 (1997): 1145-1159.
- Çabuk, Vuşlat. "Modeling and Analysis of Customer Requirements From a Driver's Seat." Ms Thesis, Graduate School of Natural and Applied Sciences, Middle East Technical University, Ankara, 2008.
- Center for Machine Learning and Intelligent Systems, University of California, Irvine. *UCI Machine Learning Repository*. February 2, 2009. <http://archive.ics.uci.edu/ml/>.
- Chawla, Nitesh V. "Data Mining for Imbalanced Datasets: An Overview." In *Data Mining And Knowledge Discovery Handbook*, edited by Oded Maimon and Lior Rokach, 853-867. Springer, 2005.
- Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and Philip W. Kegelmeyer. "SMOTE: Synthetic Minority Over-sampling TEchnique." *Journal of Artificial Intelligence Research*, 2002: 341-378.
- D., Tracy N., Young J. C., and Mason R. L. "Multivariate Control Charts for Individual Observations." *Journal of Quality Technology* 24 (1992): 88-95.
- Dasgupta, Somesh. "The Evolution of the D2-Statistic of Mahalanobis." *Sankhya: The Indian Journal of Statistics* 55 (1993): 442-459.

De Maesschalck, R, D Jouan-Rimbaud, and D L Massart. "Tutorial : Mahalanobis Distance." *Chemometrics and Intelligent Laboratory Systems* 50 (2000): 1-18.

Erdural, Serkan. "A Method for Robust Design of Products or Processes with Categorical Response." Msc Thesis, The Graduate School of Natural and Applied Sciences, Middle East Technicla University, Ankara, 2006.

Foley, Donald H. "Considerations of Sample and Feature Size." *IEEE TRANSACTIONS ON INFORMATION THEORY* IT-18, no. 5 (September 1972): 618-626.

Hamada, M, and C. F. J, Wu. "Should Accumulation Analysis and Related Methods be Used for Industrial Experiments?" *Technometrics* 28, no. 4 (1986): 302-306.

Hawkins, Douglas M. "Discussion." *Technometrics* 45, no. 1 (February 2003): 25-30.

Hirohisa, A, N Mochiduki, and H Yamada. "A Model for Detecting Cost-Prone Classes Based on Mahalanobis-Taguchi Method." *IEICE Transactions on Information and Systems* 89 - D, no. 4 (April 2006): 1347-1358.

İşkol, M E, Erkul İ, Karagüzel B., Gürsoy K, Aral B, and Gülbay B. "Six Sigma Applications at Casting Process." *Altı Sigma Forum* 5 (2005): 24-32.

Itagaki, Masayuki, Eri Takamiya, Watanabe Kunuhiro, Takanori Nukaga, and Umemura Fumio. "Diagnosis of quality of fresh water for carbon steel corrosion by Mahalanobis distance." *Corrosion Science* 49 (2007): 3408–3420.

Jugulum, Rajesh. "New Dimensions in Multivariate Diagnosis to Facilitate Decision Making Progress." Industrial and Manufacturing Engineering, Wayne State University, Detroit, Michigan, 2000.

Jugulum, Rajesh, Genichi Taguchi, Shin Taguchi, and James O Wilkins. "Discussion." *Technometrics* 45, no. 1 (February 2003): 16-21.

Kubat, Miroslav, and Stan Matvin. "Adressing the Curse of Imbalanced Training Sets: One Sided Selection." *In Proceedings of the Fourteenth International Conference on Machine Learning*. Nashville, Tennessee: Morgan Kaufmann, 1997. 179-186.

Leese, M N, and P L Main. "The Efficient Computation Of Unbiased Mahalanobis Distances and Their Interpretation in Archaeometry." *Archaeometry* 36 (1994): 307-316.

Mahalanobis, Chandra Prasanta. "On the Generalized Distance in Statistics." *Proceedings of the National Institute of Science of India*, 1936: 49-55.

Montgomery, Douglas C. *Introduction to Statistical Quality Control*. John Wiley & Sons, 2005.

- Nair, V. N. "Testing in Industrial Experiments with Ordered Categorical Data." *Technometrics* 28, no. 4 (1986): 283-291.
- Olver, Peter J., and Chehzad Shakiban. *Applied Linear Algebra*. Upper Saddle River, NJ: Prentice Hall, 2006.
- Phadke, Madhav S. *Quality Engineering Using Robust Design*. New Jersey: Prentice Hall PTR, 1989.
- Piziak, Robert. *Matrix theory : from generalized inverses to Jordan form*. Boca Raton: Chapman & Hall/CRC, 2007.
- Riho, Taketoshi, Akihito Suzuki, Junji Oro, Kenji Ohmi, and Hajime Tanaka. "The Yield Enhancement Methodology for Invisible Defects Using the MTS+ Method." *IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING*, 18, no. 4 (November 2005): 561-568.
- Rousseeuw, Peter J, and Bert C van Zomeren. "Unmasking Multivariate Outliers and Leverage Points." *Journal of American Statistical Association* 85, no. 411 (September 1990): 633-651.
- Srinivasaraghavan, Jayanth, and Venkat Allada. "Application of mahalnobis distance as a lean assessment metric." *The International Journal of Advanced Manufacturing Technology* 29 (August 2006): 1159-1168.
- Su, Chao-Ton, and Yu-Hsiang Hsiao. "An Evaluation of the Robustness of MTS for Imbalanced Data." *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, 19, no. 10 (October 2007): 1321-1332.
- Tabachnick, Barbara G., and Linda S Fidell. *Using Multivariate Statistics*. 4th Edition. MA: Allyn & Bacon, 2001.
- Taguchi, Genichi. *Taguchi Methods: Case Studies from the U.S. and Europe*. Michigan: American Supplier Institute, 1991.
- . *Taguchi Methods: Research and Development*. Michigan: American Supplier Institute, 1991.
- . *Taguchi Methods: Signal-to-noise ratio for quality evaluation*. Michigan: American Supplier Institute, 1991.
- Taguchi, Genichi, and R Jugulum. "New Trends in Multivariate Diagnosis." *Sankhya: The Indian Journal of Statistics* 62 (2000): 233-248.
- Taguchi, Genichi, Rajesh Jugulum, and Shin Taguchi. Multivariate data analysis method and uses thereof. United States of America Patent US 7,043,401 B2. May 9, 2006.

Taguchi, Genichi, Subir Chowdhury, and Wu Yun. *The Mahalanobis-Taguchi System*. New York: McGraw-Hill, 2001.

Tsui, Kwok-Leung. "A critical look at Taguchi's modelling approach for robust design." *Journal of Applied Statistics* 23, no. 1 (1996): 81-95.

Wang, Huei-Chun, Chih-Chou Chiu, and Chao-Ton Su. "Data Classification Using The Mahalanobis - Taguchi System." *Journal of the Chinese Institute of Industrial Engineers*, 21, no. 6 (2004): 606-618.

Watabe, A., K. Komiya, J. Usuki, K. Suzuki, and H. Ikeda. "Effective designation of specific shots on video service system utilizing mahalanobis distance." *IEEE Transactions on Consumer Electronics* 51, no. 1 (2005): 152-159.

Watkins, David S. *Fundamentals of Matrix Computations*. New York: Wiley, 1991.

Weiss, Gary M, and Provost Foster. "The Effect of Class Distribution on Classifier Learning." Technical Report ML-TR-43, Department of Computer Science, Rutgers University, New Jersey, 2001.

Woodall, William H, Rachelle Koudelik, Kwok-Leung Tsui, Seoung Bum Kim, and et al. "A review and analysis of the Mahalanobis-Taguchi System." *Technometrics* 45 (February 2003): 1-15.

Wu, Ful-Chiang, and Chi-hao Yeh. "A Comparative Study on Optimization Methods for Experiments with Ordered Categorical Data." *Computers & Industrial Engineering*, 2006: 220-232.

Young, Ian T. "Further Consideration of Sample and Feature Size." *Information Theory, IEEE Transactions on* 24, no. 6 (1978): 773-775.

APPENDIX A

ANOVA TABLE OF CUSTOMER SATISFACTION DATASET

Table 48 Anova Table of Customer Satisfaction Dataset

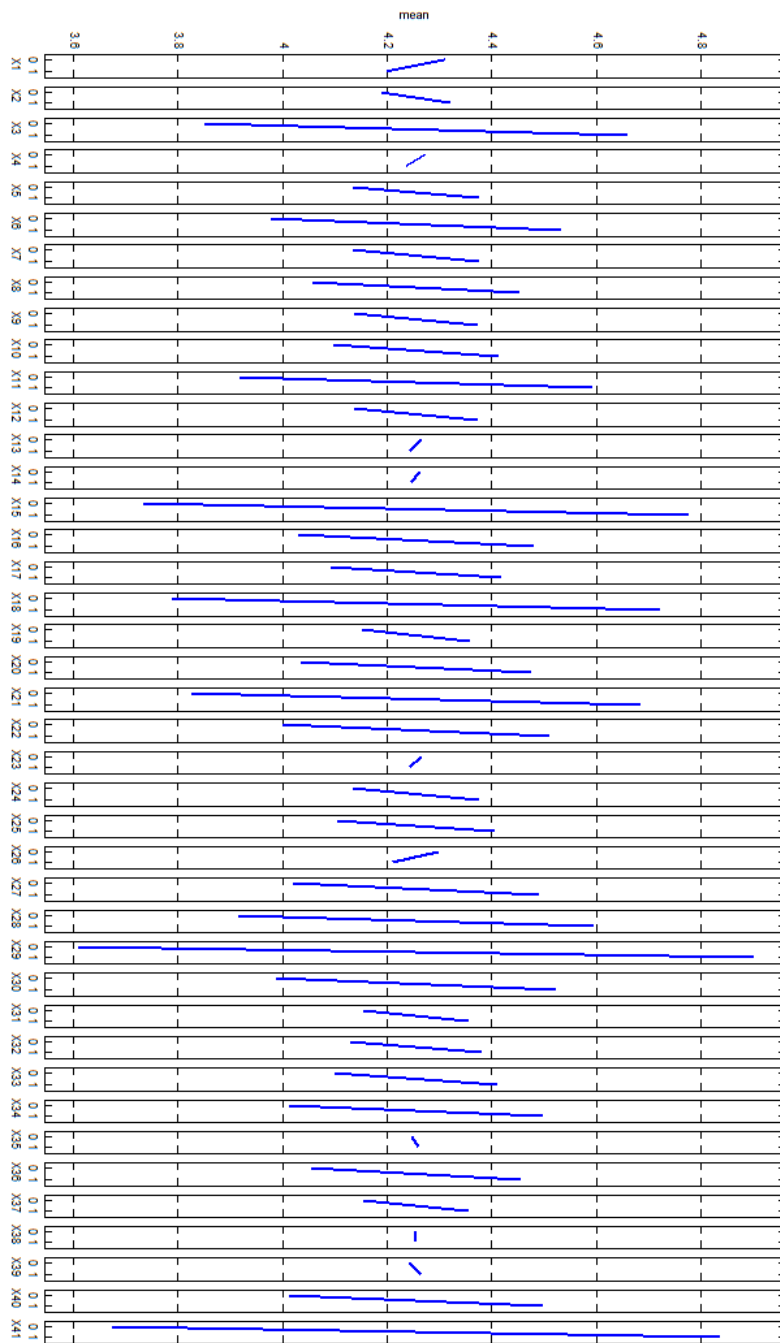
'Source'	'Sum Sq.'	'd.f.'	'Singular?'	'Mean Sq.'	'F'	'Prob>F'
'X1'	0.623233	1	0	0.623233	0.544119	0.461825
'X2'	0.871198	1	0	0.871198	0.760607	0.384462
'X3'	32.60171	1	0	32.60171	28.46319	3.26E-07
'X4'	0.068261	1	0	0.068261	0.059596	0.807452
'X5'	2.95823	1	0	2.95823	2.582707	0.110033
'X6'	15.46313	1	0	15.46313	13.50021	0.000326
'X7'	2.943721	1	0	2.943721	2.57004	0.110902
'X8'	7.856342	1	0	7.856342	6.859043	0.009678
'X9'	2.750062	1	0	2.750062	2.400964	0.123261
'X10'	5.002293	1	0	5.002293	4.367292	0.038237
'X11'	22.62536	1	0	22.62536	19.75326	1.65E-05
'X12'	2.752837	1	0	2.752837	2.403387	0.123074
'X13'	0.020525	1	0	0.020525	0.017919	0.893681
'X14'	0.012312	1	0	0.012312	0.010749	0.917556
'X15'	54.30139	1	0	54.30139	47.40827	1.28E-10
'X16'	10.2394	1	0	10.2394	8.939592	0.003238
'X17'	5.302125	1	0	5.302125	4.629064	0.032954
'X18'	43.67262	1	0	43.67262	38.12874	5.37E-09
'X19'	2.099153	1	0	2.099153	1.832683	0.177744
'X20'	9.78187	1	0	9.78187	8.540142	0.003984
'X21'	36.98121	1	0	36.98121	32.28675	6.24E-08
'X22'	12.92344	1	0	12.92344	11.28292	0.000981
'X23'	0.022199	1	0	0.022199	0.019381	0.889457
'X24'	2.915959	1	0	2.915959	2.545802	0.112586
'X25'	4.528481	1	0	4.528481	3.953627	0.048498
'X26'	0.368115	1	0	0.368115	0.321386	0.57158
'X27'	11.14724	1	0	11.14724	9.732193	0.002153
'X28'	23.24149	1	0	23.24149	20.29117	1.29E-05
'X29'	83.25166	1	0	83.25166	72.68354	1.15E-14
'X30'	14.38159	1	0	14.38159	12.55596	0.00052
'X31'	2.059471	1	0	2.059471	1.798038	0.181874

Table 49 Anova Table of Customer Satisfaction Dataset

'X32'	3.163927	1	0	3.163927	2.762292	0.098493
'X33'	4.904907	1	0	4.904907	4.282269	0.04014
'X34'	11.73744	1	0	11.73744	10.24747	0.001655
'X35'	0.006768	1	0	0.006768	0.005909	0.938824
'X36'	8.045854	1	0	8.045854	7.024498	0.008858
'X37'	2.013331	1	0	2.013331	1.757755	0.186818
'X38'	0.00012	1	0	0.00012	0.000105	0.991843
'X39'	0.03008	1	0	0.03008	0.026261	0.871471
'X40'	11.72834	1	0	11.72834	10.23952	0.001662
'X41'	67.89007	1	0	67.89007	59.27198	1.40E-12
'Error'	180.9731	158	0	1.145399	[]	[]
'Total'	704.2605	199	0	[]	[]	[]

APPENDIX B

MAIN EFFECTS PLOT FOR CUSTOMER SATISFACTION DATASET



APPENDIX C

MAHALANOBIS TAGUCHI GRAM-SCHMIDT SYSTEM

An alternative way of evaluating Mahalanobis Distance using the Gram-Schmidt orthogonalization process (GSP) is present. (Jugulum 2000) Two motives are stated by Jugulum (2000) for using MTGS instead of regular correlation matrix method. The first motive arises from situations when the normal method cannot function because the inverse of the correlation matrix cannot be calculated. MTGS can be used even when the sample size is smaller than the number of variables. The other reason is that although the Mahalanobis Distances can be calculated, whether the abnormality is in the "good" direction or in the "bad" direction cannot be calculated using the regular method. That is, although the sample is outside the Mahalanobis Space, it may be off the limits in the good direction. (Jugulum 2000)

The steps of MTGS are the same as those of MTS. The major difference is that orthogonal vectors are formed from the original variables and are used. Among different orthogonalization methods, Taguchi and Jugulum (2000) used Gram-Schmidt Method in their work.

Gram-Schmidt Method

Gram-Schmidt process forms an orthogonal basis from a set of vectors where an orthogonal basis

$$\mathbf{v} = \{v_1, \dots, v_n\} \in \mathbb{R}^n$$

is constructed from the set

$$\mathbf{w} = \{w_1, \dots, w_n\}$$

These orthogonal elements are constructed sequentially one at a time. The steps of Gram-Schmidt process is explained by Olver and Shakiban (2006) as follows:

(1) Any element of W can be used as the first orthogonal basis element v_1 as there is no limitation. Thus, the first vector in W is used

$$v_1 = w_1$$

(2) The second basis vector is constructed so that it is orthogonal to the first

$$\langle v_1, v_2 \rangle = 0$$

Let $v_2 = w_2 - cv_1$ where c is a scalar to be determined. The orthogonality condition is

$$0 = \langle v_2, v_1 \rangle = \langle w_2, v_1 \rangle - c \langle v_1, v_1 \rangle = \langle w_2, v_1 \rangle - c \|v_1\|^2$$

Rearranging this equation yields

$$c = \langle w_2, v_1 \rangle / \|v_1\|^2 \quad [2.39]$$

Substituting [2.39] into [2.38],

$$v_2 = w_2 - (\langle w_2, v_1 \rangle / \|v_1\|^2) v_1$$

(3) Next orthogonal vector v_3 is constructed by subtracting a linear combination of the first two orthogonal basis elements from w_3 .

$$v_3 = w_3 - c_1 v_1 - c_2 v_2$$

v_3 must be orthogonal to both v_1 and v_2 . Since v_1 and v_2 are already orthogonal,

$$0 = \langle v_3, v_1 \rangle = \langle w_3, v_1 \rangle - c_1 \langle v_1, v_1 \rangle$$

$$0 = \langle v_3, v_2 \rangle = \langle w_3, v_2 \rangle - c_2 \langle v_2, v_2 \rangle$$

Solving for the scalars c_1 and c_2 ,

$$c_1 = \langle w_3, v_1 \rangle / \|v_1\|^2$$

$$c_2 = \langle w_3, v_2 \rangle / \|v_2\|^2$$

Substituting c_1 and c_2 into their places in Eq. 5,

$$v_3 = w_3 - \frac{\langle w_3, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle w_3, v_2 \rangle}{\|v_2\|^2} v_2$$

The remaining orthogonal vectors v_4, \dots, v_n are constructed as linear combinations of W in the same manner.

The general Gram-Schmidt formula is therefore

$$v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\|v_j\|^2} v_j, k = 1, \dots, n$$

If the orthogonal vectors are normalized by dividing them to their norms orthonormal vectors with unit length are obtained:

$$u_k = \frac{v_k}{\|v_k\|} \quad k=1, \dots, n$$

In the general Gram-Schmidt formula, $\frac{\langle w_k, v_j \rangle}{\|v_j\|^2}$ are called the Gram-Schmidt coefficients.

Although the original Gram-Schmidt Process is easier in hand computations, in large scale computations it can be subject to numerical instabilities and the round-off errors can lead to vectors that are non-orthogonal. Thus, modifications were made on Gram-Schmidt Process to overcome this problem. The algorithm for obtaining orthonormal vectors is as follows:

$$u_1 = \frac{w_1}{\|w_1\|}$$

To ensure that the other vectors will be orthogonal to u_1 , appropriate multiples of u_1 is subtracted from all of the remaining basis vectors in W :

$$w_k^{(2)} = w_k - \langle w_k, u_1 \rangle u_1 \quad k=2, \dots, n$$

where the superscript (2) denotes the second stage computations.

The second orthonormal basis vector u_2 is then obtained by normalizing:

$$u_2 = \frac{w_2^{(2)}}{\|w_2^{(2)}\|}$$

In a similar way, it must be ensured that the remaining basis vectors are orthogonal to u_2 :

$$w_k^{(3)} = w_k^{(2)} - \langle w_k^{(2)}, u_2 \rangle u_2, k=3, \dots, n$$

As a result, they are orthogonal to both u_1 and u_2 . $u_3 = \frac{w_3^{(3)}}{\|w_3^{(3)}\|}$ is the next orthonormal basis element, and the process continues.

The algorithm starts with the initial basis vectors $w_j^{(1)}$ from $w_j = w_j^{(1)}$, $j=1, \dots, n$ and then recursively computes

$$u_j = w_j^{(j)} / \|w_j^{(j)}\|, w_k^{(j+1)} = w_k^{(j)} - \langle w_k^{(j)}, u_j \rangle u_j, j=1, \dots, n \quad k=j+1, \dots, n$$

$$u_1 = \frac{w_1}{\|w_1\|} \quad [2.49]$$

$$w_k^{(j+1)} = w_k^{(j)} - \langle w_k^{(j)}, u_j \rangle u_j, j = 1, \dots, n \quad k = j + 2, \dots, n$$

The modified Gram-Schmidt process forms the same orthonormal basis vectors in a more numerically stable way.

The Steps of Mahalanobis Taguchi Gram Schmidt System

1. Generation of the Normal Space

The first step in MTGS is to construct a basis for a normal group. This basis will be used as a reference to measure the normality of other observations. This step is the same as the first step in MTS. The p variables that are considered to be important in the definition of normality are determined.

2. Collection of Samples To Construct the Normal Group

In the second step, samples that will be used as the members of the normal group are collected. Different from MTS, the number of samples does not have to be greater than the number of variables if Gram-Schmidt Process will be used. However, if other matrix decomposition methods such as QR decomposition will be used to obtain orthogonal vectors, the number of samples has to be greater than or equal to the number of variables.

3. Calculation of the Mahalanobis Distance for Each Sample in the Normal Group

In a case where the number of samples in the normal group is n and the number of variables is p , each variable has n values from the samples. Considering each variable, there are p variable vectors of size $1 \times n$.

Similar to the original MTS method, variables are standardized to obtain

$$Z_i = (z_{i1}, z_{i2}, \dots, z_{in}) \quad i=1, \dots, p$$

Standardized vectors are then orthogonalized using Gram-Schmidt Orthogonalization Process, to obtain k orthonormal vectors:

$$U_i = (u_{i1}, u_{i2}, \dots, u_{in}) \quad i=1, \dots, p$$

After orthonormalization, the j^{th} sample becomes

$$u_j = (u_{1j}, u_{2j}, \dots, u_{pj})^T$$

Scaled Mahalanobis Distance is calculated using the orthonormal vectors of the vectors.

$$MD^2 = \frac{1}{p} u^T \Sigma^{-1} u$$

Σ is the covariance matrix of variable vectors U_1, \dots, U_p .

As the variables are orthonormalized, their vectors are orthogonal to each other and they are uncorrelated. Thus the covariance matrix is

$$\Sigma = \begin{pmatrix} \sigma_{U_1}^2 & 0 & \dots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \sigma_{U_p}^2 \end{pmatrix}$$

The inverse of the correlation matrix is

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma_{U_1}^2} & 0 & \dots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{\sigma_{U_{p1}}^2} \end{pmatrix}$$

Therefore, MD corresponding to the j th observation of the sample is computed using the equation

$$MD_j = \frac{1}{k} \left(\frac{u_{1j}^2}{\sigma_{U_1}^2} + \frac{u_{2j}^2}{\sigma_{U_2}^2} + \dots + \frac{u_{pj}^2}{\sigma_{U_p}^2} \right) \quad j = 1, \dots, n$$

The MD values calculated using the two methods are shown to be approximately equal to each other. (Taguchi and Jugulum 2000)

The following steps of MTGS are the same as the steps of MTS. In calculation of the MD of the abnormal samples, means, standard deviations and the Gram-Schmidt coefficients of the normal samples are used.

4. Removal of the Outliers from the Normal Group

After calculating the MD of the samples, the samples having higher MDs should be removed as outliers and MDs should be recalculated with the remaining samples. This process is repeated until a suitable MS with no outliers is generated. Although Taguchi and Jugulum (2000) suggest the removal of outliers from the normal group; they do skip this step in the examples provided. (Jugulum 2000), (Taguchi and Jugulum 2000) , (Taguchi, Chowdhury and Yuin 2001)

5. Gathering of r Samples from Outside the Normal Group

In order to check whether the model formed is able to determine whether a sample is from outside the normal group or not, samples that are known to be outside the normal group must be collected. This step is the same for MTS and MTGS.

6. Calculation of the MDs from outside the normal group.

In calculation of the MD of the abnormal samples, means, standard deviations and the Gram-Schmidt coefficients of the normal samples are used. Abnormal samples enter Gram-Schmidt Orthogonalization Process. Rather than orthogonalizing, the variables are processed using the orthogonalization information of the normal group, that is, the normal group coefficients.

The Gram-Schmidt orthogonalization formula was given as

$$v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\|v_j\|} v_j, k = 1, \dots, n$$

where $\frac{\langle w_k, v_j \rangle}{\|v_j\|}$ is the Gram-Schmidt coefficient.

7. Defining the Optimal System by Identifying Useful Variables

Similar to MTS, S/N ratios are used to find the variables that are contributing to the discrimination in a positive manner. Jugulum (2000) argue that "S/N ratios can be computed for all the variables directly from the orthonormal vectors, if the effects of partial correlations are not significant." If the effects of partial correlations are significant, then OA's will be used just like in MTS.

If partial correlations are not significant, the suggested larger-the-better S/N ratio for determining useful variables is calculated as follows:

$$S/N_i = -10 \log \left(\frac{1}{r} \right) \left[\sum_{j=1}^r \frac{1}{\left(\frac{u_{ij}}{\sigma_i} \right)^2} \right] i = 1, \dots, p$$

$U_i = (u_{i1}, u_{i2}, \dots, u_{ir})$: Orthonormalized vector for the i^{th} variable. $i=1, \dots, p$

r : The number of abnormal samples

σ_i : The standard deviation of the i^{th} variable.

Jugulum (2000) suggests that the orthonormal variables U_1, U_2, \dots, U_p correspond to the original variables X_1, X_2, \dots, X_p respectively. Thus after calculating the Signal to Noise ratio for each variable, the variables that have positive S/N values are included in the model.

However, Hawkins (2003) state that selection of important variables using orthogonal counterparts is not possible since a variable U_j is a function of all of X_1, X_2, \dots, X_j . U_j is computed using all of the first j^{th} variables.

It was shown that in Gram-Schmidt orthogonalization process orthogonalization is performed sequentially for each variable. For this reason the process depends on the ordering of the variables, especially the starting variable. Hawkins et al state that there are $p!$ different orderings of the variables, where each ordering will produce a different orthonormalization. Different combinations of important variables are shown to be found in each case. Because of these two major drawbacks, the determination of useful original variables from the orthonormal vectors is inappropriate. (Hawkins 2003) (Woodall, et al. 2003)

Jugulum (2000) also has pointed out that the orthogonalized vectors depend on the first vector. Therefore, the effects of variables depend on the order of variables. It was shown using examples that the useful set of variables changed when the order of orthogonalization changed. He adds that further research on this area is required. Actually, the process not only depends on the first variable, but on the entire ordering. (Woodall, et al. 2003)

Hawkins (2003) also state that the S/N ratios of the MTGS involve means of the inverse squares of the W (for the larger-the-better S/N ratio) In the idealized case, each W_j follows a distribution with mean 0 and standard deviation 1. Thus Hawkins (2003) points out that these squared inverses have a "hugely heavy tailed distribution". This makes S/N calculations highly sensitive to small changes in the data, such as "minute details of how the data were recorded and rounded". Also considering the fact that the order of the variables changes the selected variables, the use of S/N in MTGS is problematical.

Determination of the Direction of Abnormalities.

In some cases, the relative position of the sample to the normal group is important. Even if a sample is not inside the normality region and is found to be abnormal, it may be a "good" or a "bad" sample. One example for such situation may be exam grades. In this example, there can be very good and very bad grades, both being away from the general results of the students. Mahalanobis distance is a one dimensional distance metric that cannot separate these two cases. As a result, MTS method cannot be used to determine whether an abnormality is "good" or "bad". For this issue, it is suggested that MTGS method is used. Jugulum (2000) suggests the use of Gram-Schmidt coefficients to determine the direction of the vectors.

As the variable vectors were standardized prior to orthogonalization, the mean of the orthogonal vectors $U_i = (u_{i1}, u_{i2}, \dots, u_{in})$ $i=1, \dots, k$ are zero. For a variable j , u_{ij} is can be above or below zero. Based on the value u_{ij} takes, Jugulum proposes, the abnormality can be classified as "good" or "bad".

The approach is to determine if each u_i is better if large or small. Jugulum states that as the elliptical shape of the Mahalanobis space does not change even if the orthogonal vectors are used, the original variables can be used to find this information. After determining the directions each variable gets better, values of u_{ij} are used to determine whether the sample is a "good" abnormal or not.

Jugulum (2000) provides a two variable case to explain this process:

If both of the original variables W_1, W_2 are better if they are larger, then their orthogonalized vectors U_1 , and U_2 should also be larger. Since the mean of the standardized vectors is zero, it is sufficient to check if they are greater than zero or not. It is also known that the sample must have a MD larger than the threshold so that it is labeled as abnormal. Thus, for the j^{th} sample,

$$w_{1j} > 0 \text{ or } u_{1j} > 0 \text{ and } w_{2j} > 0 \text{ or } u_{2j} > 0$$

$$MD_j > T$$

where T is the threshold.

As a result, the sample satisfying these conditions is marked to be a "good" abnormality, if not; it is a "bad" abnormality.

This procedure can be applied to the cases where there is knowledge about the direction the variable gets better. If there is no clear distinction for the variables, then this procedure cannot be used.

It was previously explained that Hawkins had put forth some major shortcomings of this method. As a result of the way orthogonal vectors are constructed, Gram Schmidt variables (U_{ij}) cannot be used to determine the important variables as U is a function of all original variables. It was explained previously that the Gram Schmidt variables depends on the ordering of the variables and different orderings in the orthogonalization process may lead to different significance results of the variables. (Hawkins 2003)

It is usually not possible to determine whether a variable is better when it is larger or smaller in classification problems. Combined with the two major drawbacks Hawkins has explained, the procedure for the determination of the direction of abnormalities requires further research.

APPENDIX D

OTHER METHODS IN HANDLING SMALL DATA IN MTS

Adjoint Matrix Approach

Adjoint matrix of a square matrix is the transpose of its cofactor matrix. Cofactor matrix for an $n \times n$ matrix A is the matrix whose value in $A(i,j)$ is the cofactor C_{ij} of A . Cofactor $_{ij}$ of A can be defined as

$$C_{ij} = (-1)^{i+j} M_{ij}$$

where

M_{ij} =the minor of entry in $A(i,j)$. Minor is the determinant of the matrix A with where row i and column j is deleted.

Previously, it was stated that the inverse of a matrix A can be calculated using the formula

$$A^{-1} = \frac{1}{\det A} A_{\text{adj}}$$

In cases where the determinant is zero (singularity), or very close to zero (multicollinearity), Jugulum (2000) suggests the use of adjoint matrix instead of the inverse.

$$MDA^2 = \frac{1}{k} Z^T A_{\text{adj}} Z$$

Although MDA is similar to a MD value, because of the removal of the determinant in from the formula, the average MDA is not $n/(n-1)$.

Although the use of adjoint matrix of the covariance matrix instead of the inverse of the covariance matrix solves the calculation issues, the change in the mean of the MD of the normal group makes comparison between different MTS scales difficult. A distance measured in one scale cannot be compared to a distance in another scale in this method. Even different combinations of variables present in OA would yield different scales.

In this thesis, the research is aimed towards parameter optimization. Various MTS models are formed and their resulting Mahalanobis Distances are compared to each other to assess the performance of the models. The need of comparison of the different MTS models prohibits the use of Adjoint Matrix approach in our research.

Generalized Inverse

Another solution to the invertibility problem is the use of generalized inverse of the covariance matrix instead of its inverse.

Let A be any $m \times n$ matrix. A has a pseudoinverse if and only if there is a $n \times m$ matrix X satisfying

$$(1) AXA = A$$

$$(2) XAX = X$$

$$(3) (AX)^* = AX$$

$$(4) (XA)^* = XA$$

where $*$ denotes the conjugate transpose.

Generalized inverse X of a matrix A is a matrix satisfying one or more of the properties above. If it satisfies all of the properties, it is commonly called Moore-Penrose pseudoinverse. (Piziak 2007)

If A is a non-singular square matrix, then its pseudoinverse is equal to its ordinary inverse. If it is not possible to take the inverse of a matrix, pseudoinverse provides an estimate of the inverse that meets the four properties of an inverse given above. Singular Value Decomposition (SVD) method is commonly used to accurately calculate the pseudoinverse of a matrix.

In MTS, pseudo inverse X of a matrix A can also be used where inverse cannot be calculated. Due to the fact that pseudoinverse is actually an estimate solution for a problem with no solution, its use is not preferred in this study.