

# Applying Mahalanobis Taguchi System for Analyzing the Effect between University Admission Requirements and Student's Academic Accomplishment

Jung-Eui Hong\*

홍 정 의\*

## Abstract

Mahalanobis Taguchi System (MTS) is a pattern information technology, which has been used in different diagnostic applications to make quantitative decisions by constructing a multivariate measurement scale using data analytic methods. In MTS approach, Mahalanobis distance (MD) is used to measure the degree of abnormality of patterns and principles of Taguchi methods are used to evaluate accuracy of predictions based on the scale constructed. The advantage of MD is that it takes into consideration the correlations between the variables and this consideration is very important in pattern analysis. The purpose of this study is constructing admission diagnosis system and define the effect of admission requirements for student's academic accomplishment.

**Keywords:** Taguchi System, Diagnostic, Mahalanobis

---

\* Department of Industrial and Management Engineering, Chungju National University

## 1. Introduction

In this research, the use of Mahalanobis Distance methodology is preferred over other classical statistical approaches for the following reasons. First, MD considers the variance and covariance of the measured variables rather than just the average value. It weights the differences by the variability range in the sample point direction. This accounts for natural variation within a set. Second, it also accounts for the ranges of acceptability between the variables. It compensates for the interactions between variables. This is useful since most systems are not comprised of independent variables rather systems of dependent variables impacting the system performance. Also, it calculates distances in units of standard deviation from the group mean. By using standard deviation, the attribute under consideration is given in the original units, which relates the results to the original data.

At this study, the requirements for raduate admission consists of three variables including GPA (X1), SAT (X2), and SAT - Math (X3). These characteristics and their designated variable notations are shown in <Table 1>

<Table 1> University Admission Variables

Variable	Characteristic
X1	Grade Point Average (GPA)
X2	SAT
X3	SAT - Math

The Normal Group consists of University Admission scores for successful graduates. We define the normal group with fifteen samples ( $n = 15$ ) and three variables ( $k = 3$ ) as shown in <Table 2>

&lt;Table 2&gt; Normal Group Data

	GPA(X1)	SAT (X2)	SAT-MATH(X3)
1	3.0	1010	670
2	2.9	990	428
3	3.2	1035	712
4	2.8	980	546
5	3.9	1310	677
6	3.2	990	650
7	2.8	965	646
8	3.7	1380	715
9	3.4	1300	645
10	3.2	1205	645
11	2.6	895	490
12	3.1	950	555
13	3.6	1110	520
14	2.7	1045	625
15	3.7	1235	690

## 2. Generalization of Data

### 2.1 Calculate the mean

The mean for each characteristic in the normal group of this data is calculate as:

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} \quad (1)$$

### 2.2 Calculate the standard deviation

The standard deviation for each characteristic in the normal group is calculated. In the literature for this example, the standard deviation calculation for a population is used although the number of samples is relatively small. This example will also use the standard deviation equation for a population to remain consistent with the literature.

$$s = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}} \quad (2)$$

### 2.3 Normalize the Data

The next step in the Mahalanobis-Taguchi System is to normalize the data from the normal group.

$$Z_i = \frac{x_i - \bar{x}}{s} \quad (3)$$

### 2.4 Calculate the Mean of the Normalized data

By normalizing the data, the mean for each characteristic in the normal group is now the zero point. This is illustrated in Table 3.3 where the mean for all normal group characteristics is equal to zero.

$$\bar{Z} = \frac{\sum_{i=1}^N Z_i}{N} \quad (4)$$

### 2.5 Calculate the Standard Deviation of the Normalized Data

The standard deviation of the normalized data for each characteristic in the normal group is one. This is shown in Table 3.3 in which the standard deviation for all characteristics is equal to 1.0.

$$s = \sqrt{\frac{\sum_{i=1}^N (Z_i - \bar{Z})^2}{N}} \quad (5)$$

### 2.6 Calculate the Correlation matrix for the Normalized Values

The correlation matrix for the normalized values, which gives the correlation coefficient between the variables, is calculated below.

$$r = \frac{\sum_{i=1}^N Z_1 Z_2}{N} \quad (6)$$

## 2.7 Calculate the inverse of the Correlation matrix

The next step in MTS is to calculate the inverse of the correlation matrix.

## 2.8 Calculate W Values

The W values are calculated by multiplying the matrix of the normalized data by the inverse of the correlation matrix (A-1).

$$W_i = ZA^{-1} \quad (7)$$

## 3. Mahalanobis Space for Normal Group

### 3.1 Calculate MD for the Normal Group

Mahalanobis Distance is calculated for each sample in the normal group. The D<sup>2</sup> calculation treats this as a vector multiplication.

$$D^2 = \frac{1}{k} \sum_{ij=1}^k a_{ij} \left( \frac{x_i - m_i}{\sigma_i} \right) \left( \frac{x_j - m_j}{\sigma_j} \right) \quad (8)$$

$$D^2 = \frac{1}{k} \sum_{ij=1}^k W_i \left( \frac{x_j - m_j}{\sigma_j} \right) \quad (9)$$

$$D^2 = \frac{1}{k} \sum_{ij=1}^k W_i Z_i \quad (10)$$

The MD for the normal group is given in Table 3.8. The test group is any group of data that is unknown if it is normal or abnormal. It is compared to the normal group MD to determine the degree of abnormality.

<Table 3> Abnormal Group data

	GPA (X1)	SAT (X2)	SAT-Math (X3)
1	2.40	1210	540
2	1.80	765	280
3	0.90	540	280
4	3.60	990	230
5	2.10	930	480
6	2.60	1140	530

<Table 4> MD of Normal

	MD
1	0.576
2	1.825
3	1.102
4	0.401
5	1.137
6	0.865
7	0.712
8	1.355
9	1.123
10	0.572
11	1.109
12	0.703
13	1.799
14	1.150
15	0.661

### 3.2 Calculate Z value for the Test Group

The Z values for the test group are calculated using the mean and standard deviation from the normal group.

$$Z_i = \frac{x_i - \bar{x}}{s} \quad (11)$$

### 3.3 Calculate W value for the test Group

The next step in MTS is to calculate the W values for the test group by multiplying the matrix of normalized data from the test group by the inverse of the correlation matrix from the normal group.

$$W_i = ZA^{-1} \quad (12)$$

### 3.4 Calculate MD for the test group

D2 is calculated for the test group by multiplying the W values by the normalized values.

$$D^2 = \frac{1}{k} \sum_{ij=1}^k a_{ij} \left( \frac{x_i - m_i}{\sigma_i} \right) \left( \frac{x_j - m_j}{\sigma_j} \right)$$

$$D^2 = \frac{1}{k} \sum_{ij=1}^k W \left( \frac{x_j - m_j}{\sigma_j} \right) \quad (13)$$

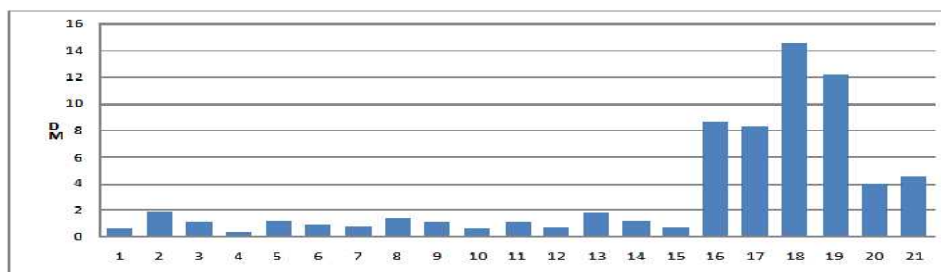
$$D^2 = \frac{1}{k} \sum_{ij=1}^k WZ$$

$$D^2 = \frac{W_1Z_1 + W_2Z_2 + W_3Z_3 + \dots + W_iZ_i}{i}$$

<Table 5> MD of Abnormal

	MD
1	8.640
2	8.292
3	14.498
4	12.232
5	3.991
6	4.598

A comparison of normal group MD and test group MD is given in Figure 3.1. The bar chart illustrates the significant difference in MD values between the normal and test group.



<Figure 1> MD of Normal and Abnormal Case

## 4. Optimization of the System

### 4.1 Apply Orthogonal Array

The orthogonal array is selected based on the number of factors and number of levels. In this case, a 23 (2levels - 3factors)orthogonal arrayw as used to determine the number of runs. The general guidelines published [8] based on the number of factors and levels are used to determine the appropriate OA. A three factor - two level experiment is used to optimize the system following an L8 orthogonal array.

The three factors include GPA (X1), SAT (X2), and SAT-Math (X3). The two levels are include and exclude. Include means to use the factor to construct the space and exclude means to ignore and not use the factor to construct the space.

This shows the effect of each factor on MD. For example, in the run (1, 1, 2) only the factors GPA and SAT are used. The third factor SAT-Math is not included in the MD calculation and only two variables are considered.

### 4.2 Evaluation Matrix

Since dynamic signal-to-noise ratios are utilized, an evaluation matrix is constructed to assign levels to the input signals. These inputs values or severity levels,  $M_i$ ,are generally assigned arbitrarily. Larger values of  $M_i$  indicate a greater degree of abnormality. The  $M$  values are levels of the signal factor that span the range of expected values. The relationship between the input signals ( $M$ ) and the output ( $y$ ) is given by equation (16).

$$y_i = \beta M_i \quad (14)$$

$$y_i = \sqrt{MD} \quad , \quad (15)$$

$\beta$  is the slope, and  $M_i$  is the working mean of the  $i$ th class.

There need to be at least three levels of the signal factors to show curvature or nonlinearity. The signal levels are chosen to span reasonable range for the signal factors. Typically one is near (but not on) the lower limit, one is at the target, and one is near (but not on) the upper limit. The  $M$  values selected in this case study are shown in <Table 6>



<Table 6> Evaluation Matrix

	Point	Evaluation	Sample
M1	10	Not so Bad	5,6
M2	20	Poor	1,2
M3	30	Very Poor	3,4

### 4.3 Calculate the Dynamic Signal to Noise Ratio

The signal-to-noise ratio is calculated using the dynamic signal-to-noise equation.

$$S/N = 10 \log_{10} \frac{\frac{1}{r}(S_{\beta} - V_e)}{V_e} \quad (16)$$

Where,

ST= total sum of square

$$S_T = \sum_{i=1}^t y_i^2 \quad (17)$$

r=sum of square due to input signal

$$r = \sum_{i=1}^t M_i^2 \quad (18)$$

Sβ=sum of square due to slope

$$S_{\beta} = \frac{1}{r} \sum_{i=1}^t (M_i y_i)^2 \quad (19)$$

Se=error sum of square

$$S_e = S_T - S_{\beta} \quad (20)$$

Ve=error variance

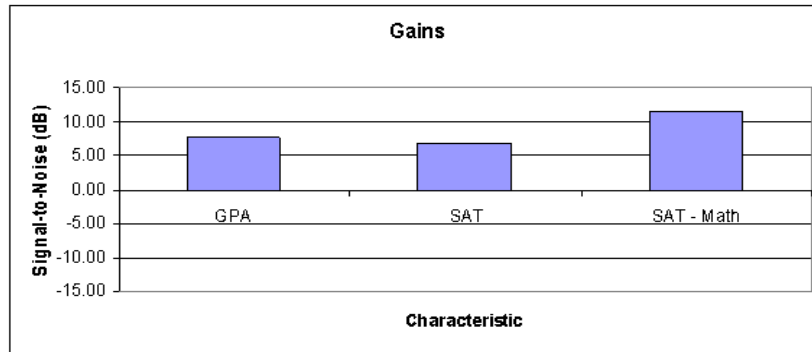
$$V_e = \frac{S_e}{t-1} \quad (21)$$

### 4.4 Determine the Factor Effects Using the Dynamic S/N ratio

The next step in applying MTS is to determine the factor effects using the Dynamic S/N ratio. This calculation is performed for the two factor levels. For example, the effect for a factor at the high level is calculated by adding the S/N ratios for that factor at the high levels and dividing by the number of runs that

factor is at the high level.

$$A1 = \frac{\sum A1}{N1} \quad (22)$$



The next step is to express the functionality of the system in terms of S/N ratios using the gain calculation. The S/N ratio gain indicates the functionality improvement in the system. The S/N ratio gain is measured in decibels (dB). The gain is calculated by subtracting the S/N values for an attribute at the level 2 from the level 1 in the orthogonal array. Using the dynamic signal-to-noise ratio, all characteristics have positive gains as shown in Figure 3.6. Therefore, it is determined that it is better to include all three characteristics.

## 5. Conclusion

At this research, the relationship between university admission requirement and student's performance at the university compared using Mahalanobis Taguchi System. MTS can be used as a diagnosis tool such as who can successful at the university or not. With optimization process, all of three entrance requirement factors are effective for evaluation for student's academic performance.

## 6. Reference

- [1] Taguchi, G and R. Jugulum, "New Trends in Multivariate Diagnosis", Indian Journal of Statistics, 62, Series B, 2233-248(2000).
- [2] Taguchi, G and R. Jugulum, The Mahalanobis-Taguchi Strategy: A Pattern Te

- chnology System, John Wiley & Sons, Inc., 2002.
- [3] Jain, Anil K., Duin, Robert P.W., Mao, Jianchang, "Statistical Pattern Recognition: A Review", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.22, No. 1, Jan. 2000.
  - [4] Taguchi, S., "Mahalanobis Taguchi System", ASITaguchiSymposium,2000.
  - [5] nde, U., "Mahalanobis Distance: A Theoretical and Practical Approach", <http://biologi.uio.no/fellesavdelinger/finse/spatialstats/Mahalanobis%20distance.ppt>,2003.
  - [6] Hayashi, S., Y. Tanaka, and E. Kodama, "A New Manufacturing Control System using Mahalanobis Distance for Maximizing Productivity", IEEETransactions,59-62,2001.
  - [7] Asada, M., "Wafer Yield Prediction by the Mahalanobis-Taguchi System", IETransactions,25-28,2001.
  - [8] Wu, Y., "Pattern Recognition using Mahalanobis Distance", TPDSymposium, 1-14,1996.
  - [9] Taguchi G., S. Chowdury, and Wu Y., The Mahalanobis Taguchi System, McGraw Hill Press New York, 2001.
  - [10] Hassoun, M.H., "Fundamentals of Artificial Neural Networks", Cambridge,MA, MITPress,1995.
  - [11] Huhnen, J. H., Hollon, K. R., and Lai, D. C., "Comparison of Mahalanobis Distance, Polynomial and Neural Net Classifier", SPIE proceedings-Application of Artificial Neural Networks, 1294, 557-567, 1990.
  - [12] Woodall, W. H. , R. Koulelik,, K. L. Tsui, S. B. Kim, Z. G. Stoumbos, and C. P. Carvounis, "A review and analysis of the Mahalanobis Taguchi", Technometrics, Vol.45,No.1,February2003,1-30..