

# Selection of Input Variables in DEA using 2 - Level Fractional Factorial Design

C.V. Pradeepa, V. Prakash,

**Abstract:** Data Envelopment Analysis (DEA) can be a statistics oriented, non – parametric method to gauge relative efficiency supported pre – selected Inputs and Outputs. It is a implemented mathematics based technique for measuring the relative overall performance of organisational units in which the presence of Multiple Inputs and Outputs makes evaluation difficult. In a few cases, the performance model isn't well defined, so it's important to pick the proper Inputs and Outputs by way of other means. We used, Morita and Avkiran proposed technique after it has been developed an Input – Output Selection Method that uses Fractional Factorial design, which is an Statistical method to locate a best and optimal combination. In this study  $2^{k-p}$  Fractional Factorial design is applied to demonstrate the proposed method relates to the Manufacture of Pharmaceuticals, Medicinal Chemical and Botanical Products from the Manual of Annual Survey of Industries (ASI) 2016 – 2017.

**Keywords:** Data Envelopment Analysis (DEA), Decision Making Units (DMUs), CCR, BCC, Super Efficiency, Mahalanobis Distance.

## I. INTRODUCTION

Companies at some stage in this Industry segment supply the lively elements employed with the aid of pharmaceutical companies to manufacture completed products, called the pharmaceutical preparations. Companies in this Industry segment furnish the active ingredients utilized by pharmaceutical corporations to manufacture completed merchandise, known as the pharmaceutical preparations. Active aspect constitute the part of a finished drug that creates the preferred effect – therapeutic or preventive for human and animals assets are vital examples of the additives produced with the aid of this industry sector. By the 1960's synthesized chemicals both a manufactured reproduction of an organic or in natural substance or a New Chemical Entity (NCE) had become commonplace active ingredients in pharmaceuticals from vitamin drugs to hormones. Meanwhile, the bio generation revolution beginning the earnest in the 1980's, resulted in methods of putting genetic cloth into small micro organisms. This made them miniature factories for the meeting of active drugs components like Insulin and in the process, created New Molecular Entities (NME's) that might be patented.

By the overdue 1990's, the primary market of this Industry the pharmaceutical industry risked high studies and development prices for the prized billions of dollars it may generate with new merchandise. Although many drug agencies have vertically manufacturing lines, a fashion grew to favour outsourcing chemical intermediates and active ingredients to smaller first-rate chemical organizations.

**Revised Manuscript Received on December 5, 2019.**

C.V. Pradeepa, Ph.D Research Scholar, Department of Statistics, Presidency College (Autonomous), University of Madras, Chennai, India.

V. Prakash, Associate Professor and Head, Department of Statistics, Presidency College (Autonomous), University of Madras, Chennai, India.

This category covers institutions generally engaged in manufacturing bulk natural and inorganic medicinal chemical compounds and their derivatives additionally as processing grading, grinding, Milling bulk botanical capsules and herbs.

In this paper, we are interested to show the green performance of the organizations in manufacturing pharmaceutical, Medicinal Chemicals and Botanical products and also makes use of the diagonal layout test to select output variables which is a statistical approach to find an optimal mixture to analyse the performance of the above point out Industry through the usage of the approach Data Envelopment Analysis.

Data Envelopment Analysis (DEA) is a linear programming based method for measuring the performance performance of similar kind of organisational gadgets termed as Decision Making Units (DMU's), with Multiple Inputs and Outputs. It identifies a subset of green "Best Practice" DMUs and for closing DMUs, the importance of their non – performance is measured by comparing to a frontier comprised of the green DMUs. Data Envelopment Analysis (DEA) can be a statistics oriented, non – parametric method to evaluate relative efficiency base on pre – selected Inputs and Outputs.

Decision Making Units (DMUs) refers to the similar form of businesses that consumes a variety of same Inputs to produce a variety of equal outputs. DMU can encompass production units, Department of massive Organisations consisting of Universities, Schools, Bank Branches, Hospitals, Power Plants, Police Stations, Tax Offices etc.

The rest of the paper is organized as followings: Section II, review of literature. In Section III, a procedure of Selecting Input variables using 2 – level Fractional Factorial Design and Mahalanobis distance is presented. An empirical example is described in Section IV to illustrate the methodology presented herein. The main results of the paper are summarized in Section V.

## II. REVIEW OF LITERATURE

Farrel (1957) is taken into consideration to be the most influential paper on DEA. The similarly pioneering contributions were made by using CCR (1978, 1979) and CCR (1981). Banker, Charnes and Cooper (1984) and Charnes et al (1985). Banker and Morey (1986) have evaluated the relative technical and scale efficiencies of DMUs by way of mathematical programming formulations when some of the inputs and outputs are exogenously fixed and past the discretionary control of DMU personnel. A massive quantity of papers have prolonged and implemented the DEA Methodology (Coelli, 1996).Zhu (1996b) and Seiford and Zhu (1998d)

develop some of new remarkable Efficiency fashions to determine the efficiency stability regions. Anderson and Peterson (1993) advocate the usage of the CRS Super Efficiency version in rating the efficient DMUs. Also, the Super Efficiency can be used in detecting influential observations (Wilson 1995) and in identifying the intense efficient DMUs (Thrall, 1966). Seiford and Zhu (1999c) examine the infeasibility of diverse Super Efficiency fashions developed from the envelopment fashions offers other first-rate performance models which can be utilized in Sentivity Analysis. Hiroshi Morita and Necmi. K. Avkiran (2008) advanced the method for selecting Inputs and Outputs in information envelopment analysis with the aid of designing statistical experiments. HulyaBayrak, OdayJayiesand Kubra Durukan demonstrated the proposed approach of Morita and Avkiran using of Fractional

Factorial layout rk – p in information Envelopment Analysis to selection of Outputs and Inputs.

III. METHODOLOGY

DEA converts multiple inputs and outputs into scalar measure of efficiency. There are essentially two varieties of DEA Models CCR Model (Constant Returns to Scale) and BCC Model (Variable Return to Scale).The two envelopment versions, one involving  $\theta$  and the other involving  $\phi$ . The version involving  $\theta$  aims to produce the observed Outputs with minimum Inputs and it is referred to as Input Oriented Envelopment DEA Program. The other version involving  $\phi$  refers to as an Output Oriented Envelopment DEA Program as it aims to maximize output production, subject to the given resource level.

Models	Input Oriented	Output Oriented
<b>I.CCR</b>	(i) Min $\theta$ Subject to, $Y\lambda \geq Y_0$ $X\lambda \leq \theta_0 X_0$ $\lambda \geq 0; \theta$ free	(ii) Max $\phi$ Subject to, $Y\lambda \geq \phi Y_0$ $X\lambda \leq X_0$ $\lambda \geq 0; \phi$ free
<b>BCC</b>	(iii) Min $\theta$ Subject to, $Y\lambda \geq Y_0$ $X\lambda \leq \theta_0 X_0$ $\sum_{n=1}^N \lambda_n = 1$ $\lambda \geq 0; \theta$ free	(iv) Max $\phi$ Subject to, $Y\lambda \geq \phi Y_0$ $X\lambda \leq X_0 \sum_{n=1}^N \lambda_n = 1$ $\lambda \geq 0; \phi$ free

Where  $\theta$  and  $\phi$  = Efficiency Measure  
 $X = [X_1, X_2, X_3, \dots, X_N]$  = Vector of Inputs  
 $Y = [Y_1, Y_2, Y_3, \dots, Y_N]$  = Vector of Outputs  
 $\lambda = [\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_N]$  = Vector of Weights

$Y_0$  = Output of the Observed DMU  
 $X_0$  = Input of the Observed DMU  
 $N$  = Number of DMUs.

The Model (1) is an Input and Output Oriented CCR Model. The Input and Output Oriented Envelopment Model is solved for each DMU in the set, it gives an efficiency score  $\theta$  or  $\phi$  and DMU weights  $\lambda_N$ . If  $\theta = 1$  and  $\phi = 1$ , then the DMU is said to be Efficient and there is no slacks. If  $\theta < 1$  and  $\phi > 1$ , then the DMU is Inefficient related to best practice DMU in the sample and efficient DMUs comprise the efficient frontier.

The Model (2) is an Input and Output Oriented BCC Model. The Input and Output Oriented Envelopment Model is solved for each DMU in the set, it gives an efficiency score  $\theta$  or  $\phi$  and DMU weights  $\lambda_N$ . If  $\theta = 1$  and  $\phi = 1$ , then the DMU is said to be Efficient and there is no slacks. If  $\theta < 1$  and  $\phi > 1$ , then the DMU is Inefficient related to best practice DMU in the sample and efficient DMUs comprise the efficient frontier.

3.1 Super Efficiency Model:

In this paper, with the intention to rank DMUs, we use the assessment contexts which can be acquired by means of partitioning the set of DMUs into several degrees of Efficiency, and rank all DMUs with two criteria: the High and Low performers. They have an effect on of all DMUs, each Efficient and Inefficient in Ranking, Author have proposed strategies for rating the first-class performers, as an instance using Super – Efficiency DEA Model. When a DMU underneath evaluation isn't always included within the

reference set of the envelopment Models, the ensuing DEA Models are called Super – Efficiency DEA Models.

As in Charnes, Cooper and Thrall (1991), the DMUs can be partitioned into four classes E, E', F and N described as follows and the extreme efficient DMUs can be identified by the Super .efficiency Models.

E – The set of Efficient DMUs

E'–The set of Efficient DMUs that aren't extreme points and expressed as linear combinations of the DMUs in set E.

F – The set of Frontier factors (DMUs) with non zero slack(s) are known as Weakly Efficient.

N – The set of Inefficient DMUs.

3.2 Selection of Input Variables Using a 2 – Level Fractional Factorial Design:

Based on the Efficiency score, the whole DMU's are categorized into two groups namely High performers and low performers with the objective of selecting the appropriate Input variables. The distance among the two groups is measured by the Mahalanobis distance.



### 3.3 Mahalanobis Distance:

In Statistics, Mahalanobis distance (MD) can be a distance measure brought by means of P.C. Mahalanobis in 1936. It's supported correlations between variables by way of which exclusive styles are often identified and analysed. It gauges similarity of an unknown sample set to a acknowledged one. It differs from Euclidean distance therein it takes below consideration the correlations of the data set and is scale – invariant. In other words, Mahalanobis distance is a Multivariate effect size.

Formally, the MD of a Multivariate vector  $x = (x_1, x_2, \dots, x_N)^T$  from a set of values with mean vector  $\mu = (\mu_1, \mu_2, \dots, \mu_N)^T$  and the Covariance Matrix  $\Sigma$ .

$$D_M(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)} \text{ -----(1)}$$

The above equation is rewritten for the pattern following as

$$\hat{D}_M(x) = \sqrt{(x - \bar{x})^T S^{-1} (x - \bar{x})} \text{ ---- (2)}$$

Where, the Mean Vector and Covariance Matrix of the pattern are given as  $\bar{x}$  and S respectively

The MD threshold is another important detail of prognostics analysis. An MD threshold value that's either too massive or too small results in false negatives or false positives, respectively. In this study, the writer considers the space of one – dimensional variables, wherein MD coincides with Welch Statistics. The Welch Statistics is given as

$$d = \frac{\bar{\theta}_h - \bar{\theta}_l}{\sqrt{\frac{V_h + V_l}{n_h + n_l}}} \text{ ----- (3)}$$

Where the average and variance of each sets are given as  $\bar{\theta}_h$ ,  $V_h$ , for  $n_h$  high performance and  $\bar{\theta}_l$ ,  $V_l$ , for  $n_l$  low performance.

### 3.4 2 – Level Fractional Factorial Design:

The intention is to locate that combo of Input variables which maximizes the distance d. In our evaluation, we carried out the 2 – level Factorial Design. When there are k candidates of Input variables, the overall total number of combination is  $2^k$ . Full Factorial designs perform all of  $2^k$  combos for k candidates. On the other hand, we will outline a  $2^{k-p}$  design to be a Fractional Factorial Design with k candidates, each at 2 levels, consisting together with  $2^{k-p}$  runs. The first (k – p) candidates are a part of  $2^{k-p}$  combinations as a Full Factorial Design, and the final p candidates may be generated as interactions with the first (k – p) columns. Table 1 shows an example of a Fractional Factorial Design where k = 7, p = 3, and  $x_i$  is a candidate variable. ‘+’ shows that the variable is selected as an Input, and ‘-’ shows that the variable is not selected as an Input. For example, the variables  $x_2, x_5, x_6$  are decided as a selected Input variables in Run No.3.

Based on the Fractional Factorial design in Table 1, we calculate the performance efficiency scores by means of Model 1 (i) and Mahalanobis distance by using (3) between two performers using the selected Input variables. The analysis of Variance (ANOVA) for the Fractional Factorial Design is shown in Table 2.

**Table 1: Fractional Factorial Design for  $2^{7-3}$  and selected Inputs**

Runs	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	Selected Inputs	Mahalanobis Distance
1	-	-	-	-	-	-	-	None	$d_1$
2	+	-	-	-	+	-	+	$x_1, x_5, x_7$	$d_2$
3	-	+	-	-	+	+	-	$x_2, x_5, x_6$	$d_3$
4	+	+	-	-	-	+	+	$x_1, x_2, x_6, x_7$	$d_4$
5	-	-	+	-	+	+	+	$x_3, x_5, x_6, x_7$	$d_5$
6	+	-	+	-	-	+	-	$x_1, x_3, x_6$	$d_6$
7	-	+	+	-	-	-	+	$x_2, x_3, x_7$	$d_7$
8	+	+	+	-	+	-	-	$x_1, x_2, x_3, x_5$	$d_8$
9	-	-	-	+	-	+	+	$x_4, x_6, x_7$	$d_9$
10	+	-	-	+	+	+	-	$x_1, x_4, x_5, x_6$	$d_{10}$
11	-	+	-	+	+	-	+	$x_2, x_4, x_5, x_7$	$d_{11}$
12	+	+	-	+	-	-	-	$x_1, x_2, x_4$	$d_{12}$
13	-	-	+	+	+	-	-	$x_3, x_4, x_5$	$d_{13}$
14	+	-	+	+	-	-	+	$x_1, x_3, x_4, x_7$	$d_{14}$
15	-	+	+	+	-	+	-	$x_2, x_3, x_4, x_6$	$d_{15}$
16	+	+	+	+	+	+	+	$x_1, x_2, x_3, x_4, x_5, x_6, x_7$	$d_{16}$

The total sum of squares  $S_T$  is given as

$$S_T = \sum_{i=1}^n (d_i - \bar{d})^2 \text{ ----- (8)}$$

The sum of squares  $S_i$  for candidate i reflects the main effect of the variable, which is the difference between ‘+’ and ‘-’ as,

$$S_i = 2 \{ \bar{d}(x_i +) - \bar{d}(x_i -) \} \text{ ----- (9)}$$

Where  $\bar{d}(x_i +)$  is the Mean of the Mahalanobis distances observed when  $x_i = +$ . The residual sum of squares  $S_E$  is given by subtracting the sum of  $S_i$  from  $S_T$ .



## Selection of Input Variables in DEA using 2 - Level Fractional Factorial Design

$$S_E = S_T - (S_1 + S_2 + S_3 + S_4 + S_5 + S_6 + S_7 + S_8 + S_9 + S_{10} + S_{11} + S_{12} + S_{13} + S_{14} + S_{15} + S_{16}) \quad \text{----- (10)}$$

The total degree of freedom is  $\phi_T = 7$ , which is the number of runs minus 1, and the degree of freedom for each sum of squares is  $\phi_i = 1$ . Therefore the degree of freedom for the residual is given as,

$$\phi_E = \phi_T - (\phi_1 + \phi_2 + \phi_3 + \phi_4 + \phi_5 + \phi_6 + \phi_7 + \phi_8 + \phi_9 + \phi_{10} + \phi_{11} + \phi_{12} + \phi_{13} + \phi_{14} + \phi_{15} + \phi_{16}) \quad \text{--- (11)}$$

The Null Hypothesis that the candidate has no effect as an Input is tested by using F Statistics,  $F = \frac{S_i/\phi_i}{S_E/\phi_E}$  ----- (12)

The test rejects the Null Hypothesis at level  $\alpha$  if F – value exceeds  $\alpha$  percentile of F distribution with degrees of freedom  $(\phi_i, \phi_E)$ , and the hypothesis tests is as following:

H<sub>0</sub>: The variable candidate has no impact on Input.

H<sub>1</sub>: The variable candidate has impact on Input.

This results within the optimal aggregate of Input variables.

**Table 2: ANOVA table for Fractional Factorial Design of 27 – 3**

Variables	Sum of Squares	Degrees of Freedom	Mean Sum of Squares	F Statistics
$x_1$	$S_1$	$\phi_1 = 1$	$V_1 = \frac{S_1}{\phi_1}$	$\frac{V_1}{V_E}$
$x_2$	$S_2$	$\phi_2 = 1$	$V_2 = \frac{S_2}{\phi_2}$	$\frac{V_2}{V_E}$
$x_3$	$S_3$	$\phi_3 = 1$	$V_3 = \frac{S_3}{\phi_3}$	$\frac{V_3}{V_E}$
$x_4$	$S_4$	$\phi_4 = 1$	$V_4 = \frac{S_4}{\phi_4}$	$\frac{V_4}{V_E}$
$x_5$	$S_5$	$\phi_5 = 1$	$V_5 = \frac{S_5}{\phi_5}$	$\frac{V_5}{V_E}$
$x_6$	$S_6$	$\phi_6 = 1$	$V_6 = \frac{S_6}{\phi_6}$	$\frac{V_6}{V_E}$
$x_7$	$S_7$	$\phi_7 = 1$	$V_7 = \frac{S_7}{\phi_7}$	$\frac{V_7}{V_E}$
Error	$S_E$	$\phi_E = 8$	$V_E = \frac{S_E}{\phi_E}$	
Total	$S_T$	$\phi_T = 15$		

The following is the summary procedure for the selection of variables.

- List potential Input Output Variables
- Used Super Efficiency Model as external criteria to distinguish the performance of two groups, e.g. high and low performance. ( Here the author used BCC Model and Super Efficiency Model for grouping)
- Assign the variables to the selected 2 – level Orthogonal layout and determine the combination of selected variables used in the experiments. (To Select Input or Output Variables alone 2 – level Fractional Factorial Orthogonal layout is used and to select both Input and Output Variables then choose 3 – level Fractional Factorial design)
- Calculate the DEA Efficiency Score and Mahalanobis distance among the two groups by using the chosen variables.
- Determine the most effective combination of Input and Output variables based totally on consequences of Analysis of Variance.
- Identify the most efficient designation of statistically great variables as either an Input or an Output the usage of Mahalanobis distance.

## IV. DATA DESCRIPTION

The data analysed in this paper is a secondary data of Manufacturing Pharmaceuticals, Medicinal Chemical and Botanical Products taken from the published Manual of Annual Survey of Industries (ASI) 2016 – 2017. The data provides the information about the Number of Employees, Number of Factories in Operation, Fixed Capital, Gross Value of Plant and Machinery, Materials Consumed, Fuel Consumed, Wages and Salaries, Net Value Added, Total Output, Income, Profit and Gross Capital Information for each States (ie., 35 – Including Union Territories) during the year 2016 - 2017. Here, the author considered only the 24 States (Including Union Territories) as DMUs by considering only the positive values of the variables and then each DMU is characterized by the following Inputs and Outputs as shown in Table 3. (Appendix 1)

### 4.1 EMPIRICAL ANALYSIS:

Based on the data structure, the results of the BCC Model in respect of data structure are given below

**Table 4: Efficiency Score, Peers And Rank By BCC Model**

S. No	State/UT/Division	Efficiency Score	Status	Peer Weights	Peer Count	Rank
1	Andhra Pradesh	1	Efficient			
2	Assam	1	Efficient		2	4

3	Bihar	1.0088	Inefficient	$\lambda_2 = 0.02;$ $\lambda_4 = 0.78;$ $\lambda_6 = 0.20;$		
4	Chhattisgarh	1	Efficient		5	2
5	Dadra and NagerHaweli	1	Efficient		1	5
6	Delhi	1	Efficient		1	5
7	Goa	1	Efficient		7	1
8	Gujarat	1	Efficient		2	4
9	Haryana	1	Efficient			
10	Himachal Pradesh	1	Efficient			
11	Jammu and Kashmir	1	Efficient		1	5
12	Karnataka	1.2437	Inefficient	$\lambda_7 = 0.59;$ $\lambda_8 = 0.13;$ $\lambda_{19} = 0.29;$		
13	Kerala	1	Efficient		1	5
14	Madhya Pradesh	1.2	Inefficient	$\lambda_4 = 0.33;$ $\lambda_7 = 0.52;$ $\lambda_{18} = 0.15;$		
15	Maharashtra	1	Efficient			
16	Odisha	1.6097	Inefficient	$\lambda_2 = 0.00;$ $\lambda_4 = 0.98;$ $\lambda_7 = 0.00;$ $\lambda_{11} = 0.01;$ $\lambda_{19} = 0.00;$		
17	Puducherry	1.2931	Inefficient	$\lambda_4 = 0.83;$ $\lambda_7 = 0.11;$ $\lambda_{18} = 0.07;$		
18	Rajasthan	1	Efficient		2	4
19	Sikkim	1	Efficient		3	3
20	Tamil Nadu	1.4197	Inefficient	$\lambda_7 = 0.08;$ $\lambda_8 = 0.10;$ $\lambda_{19} = 0.83;$		
21	Telegana	1	Efficient			
22	Uttar Pradesh	1	Efficient		1	5
23	Uttarakhand	1.0436	Inefficient	$\lambda_7 = 0.90;$ $\lambda_{13} = 0.10;$		
24	West Bengal	1.0436	Inefficient	$\lambda_4 = 0.75;$ $\lambda_5 = 0.03;$ $\lambda_7 = 0.12;$ $\lambda_{22} = 0.10;$		

From the above table, the author could not categorize the DMUs in groups because of the tie ranks. Assam, Gujarat and Rajasthan with Rank 4, Similarly, Dadar& Haveli, Delhi, Jammu and Kashmir, Madhya Pradesh and Uttar Pradesh with Rank 5. So, for instance using Super – Efficiency DEA Model the tie ranks were solved and the

DMUs were categorized in two groups namely High Performance and Low Performance by considering the set of extreme efficient DMUs in High Performance and the set of inefficient DMUs in Low Performance as shown in the below Table 5.

**Table 5: Efficiency Score by Super Efficiency Model**

S. No	State/UT/Division	Efficiency Score	Status	Rank
1	Andhra Pradesh	1.6818	Extreme Efficient	High Performance
2	Assam	1.8113	Extreme Efficient	High Performance
3	Bihar	0.8821	Inefficient	Low Performance
4	Chhattisgarh	Big	Extreme Efficient	High Performance

## Selection of Input Variables in DEA using 2 - Level Fractional Factorial Design

5	Dadra and Nager Haweli	1.3029	Extreme Efficient	High Performance
6	Delhi	1.1993	Efficient	
7	Goa	1.8803	Extreme Efficient	High Performance
8	Gujarat	1.3534	Extreme Efficient	High Performance
9	Haryana	10.5115	Extreme Efficient	High Performance
10	Himachal Pradesh	0.8266	Inefficient	Low Performance
11	Jammu and Kashmir	1.5722	Extreme Efficient	High Performance
12	Karnataka	0.72	Inefficient	Low Performance
13	Kerala	1.0457	Efficient	
14	Madhya Pradesh	0.8307	Inefficient	Low Performance
15	Maharashtra	0.8135	Inefficient	Low Performance
16	Odisha	0.6101	Inefficient	Low Performance
17	Puducherry	0.7678	Inefficient	Low Performance
18	Rajasthan	1.1045	Efficient	
19	Sikkim	2.4227	Extreme Efficient	High Performance
20	Tamil Nadu	0.5373	Inefficient	Low Performance
21	Telegana	0.8074	Inefficient	Low Performance
22	Uttar Pradesh	1.9017	Extreme Efficient	High Performance
23	Uttarakhand	0.7991	Inefficient	Low Performance
24	West Bengal	0.9029	Inefficient	

The author constructed two performers, High Performance and Low Performances based at the value of Super Efficiency Scores and it is shown in Table 6 (Appendix 2).

Based on the Fractional Factorial layout in Table 1, the author calculated the Efficiency Scores and Mahalanobis distance between the two performers using selected Inputs and shown in Table 7 and Table 8 indicates the analysis of variance of the data, where the significance level is shown as the p – value. Thus, the candidate variable of  $X_1, X_2, X_3, X_4, X_5, X_6$  and  $X_7$  has no impact on Input.

The very last step in our process generates Table 9 which shows Sum of the Mean of the Mahalanobis

distance (MD) for each variable at every level. For example, when variable  $X_1$  is selected as an Input, the Sum of the Mean of MD is 6.9080 and if  $X_1$  is not selected as an Input then the Sum of the Mean of MD is 4.0935. Thus, given that the greatest value of MD for variable  $X_1$  is 6.9080 and it have to be decided on as an Input. Maximum values of the sum of the mean are indicated in bold font in Table 9. Thus we pick five variables as Inputs namely,  $X_1$  (Number of Employees),  $X_2$  (Number of Factories in Operation),  $X_4$  (Gross Values of Plants and Machinery),  $X_5$ (Materials Consumed) and  $X_6$  (Fuels Consumed). Then run the DEA Model 1(i) the usage of the chosen Inputs.

**Table 7: Selected Input Variables and Mahalanobis Distance**

Runs	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	Selected Inputs	Mahalanobis Distance
1	-	-	-	-	-	-	-	None	0.7564
2	+	-	-	-	+	-	+	$x_1, x_5, x_7$	7.5618
3	-	+	-	-	+	+	-	$x_2, x_5, x_6$	6.1733
4	+	+	-	-	-	+	+	$x_1, x_2, x_6, x_7$	6.0529
5	-	-	+	-	+	+	+	$x_3, x_5, x_6, x_7$	4.8482
6	+	-	+	-	-	+	-	$x_1, x_3, x_6$	5.0002
7	-	+	+	-	-	-	+	$x_2, x_3, x_7$	3.9544
8	+	+	+	-	+	-	-	$x_1, x_2, x_3, x_5$	6.4989
9	-	-	-	+	-	+	+	$x_4, x_6, x_7$	3.8209
10	+	-	-	+	+	+	-	$x_1, x_4, x_5, x_6$	12.671
11	-	+	-	+	+	-	+	$x_2, x_4, x_5, x_7$	4.4877
12	+	+	-	+	-	-	-	$x_1, x_2, x_4$	8.1002
13	-	-	+	+	+	-	-	$x_3, x_4, x_5$	4.3502
14	+	-	+	+	-	-	+	$x_1, x_3, x_4, x_7$	3.6715

15	-	+	+	+	-	+	-	$x_2, x_3, x_4, x_6$	4.3571
16	+	+	+	+	+	+	+	$x_1, x_2, x_3, x_4, x_5, x_6, x_7$	5.7076

**Table 8: ANOVA Table for 27 – 3 Fractional Factorial Design**

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	67.338 <sup>a</sup>	7	9.620	2.384	.123
Intercept	498.736	1	498.736	123.609	.000
X1	25.608	1	25.608	6.347	.036
X2	.019	1	.019	.005	.947
X3	5.130	1	5.130	1.271	.292
X4	4.352	1	4.352	1.079	.329
X5	15.673	1	15.673	3.885	.084
X6	3.142	1	3.142	.779	.403
X7	6.020	1	6.020	1.492	.257
Error	32.278	8	4.035		
Total	583.752	16			
Corrected Total	99.616	15			

a. R Squared = .676 (Adjusted R Squared = .392)

**Table 9: The Sum of the Mean of the Mahanalobis distance**

Variables	Selected as an Input	Not Selected as an Input
X <sub>1</sub>	<b>6.9080</b>	4.0935
X <sub>2</sub>	<b>5.6665</b>	5.3350
X <sub>3</sub>	4.7985	<b>6.2030</b>
X <sub>4</sub>	<b>5.8958</b>	5.1058
X <sub>5</sub>	<b>6.5373</b>	4.4642
X <sub>6</sub>	<b>6.0789</b>	4.9226
X <sub>7</sub>	5.0131	<b>5.9884</b>

## V. CONCLUSION

The author considered an Input Selection approach that applies a 2 – level Fractional Factorial design, Mahanalobis Distance and ANOVA. The Concept of Super Efficiency Model is used to differentiate between the two set of performers namely high Performance and Low Performance. Variables are selected from the effects of ANOVA to Maximize the Mahanalobis Distance between the two performers. The writer unearths a powerful variable aggregate Method from a constrained quantity of experiments and proven the effectiveness of this new approach using a secondary information pertains to the Manufacture of Pharmaceutical Medicinal Chemical and Botanical products published in the Manual of Annual Survey of Industries (ASI) 2016 – 2017. The Selected Input variable measures the performance performance in a better way and gives a more potent optimality.

## REFERENCES

- Farrel, M.J. (1957): The Measurement Of Productive Efficiency, Journal Of The Royal Statistical Society Series A 120, (Part 3), 253-290.
- Charnes, A., Cooper, W.W and Rhodes, E (1978): Measuring the Efficiency of Decision Making Units, European Journal Of Operational Research, 2(6), 429-44.
- Charnes, A., Cooper, W.W and Rhodes, E (1979): Short Communication: Measuring the Efficiency Of Decision Making Units. European Journal Of Operational Research, 3(4): 339.
- Charnes, A., Cooper, W.W And Rhodes, E (1981): Evaluating Program and Managerial Efficiency: An Application Of Data Envelopment Analysis To Program Follow Through, Management Science, 29(6), 668-697.
- Banker, R.D., Charnes, A. And Cooper., W.W.(1984) : Some Models For Estimating Technical And Scale Inefficiencies In Data Envelopment Analysis, Management Science, 30(9): 1078-92.
- Charnes, A., Cooper, W.W., Golany, B., Seiford, L.M. And Stutz, J. (1985b): Foundation Of Data Envelopment Analysis For Pareto-Koopmans Efficient Empirical Production Function. Journals of Econometrics, 30(1/2), 91-107.
- Banker, R.D. And Morey, R.C. (1986a): Efficiency Analysis for Exogenously Fixed Inputs and Outputs, Operations Research (USA), 34(4): 513-521.
- Coelli T., (1996) A Guide To DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program, Centre For Efficiency And Productivity Analysis, CEPA Working Paper 96/08, University Of New England.
- Ramesh, B. and Patel, M.R. (1999), "Performance Evaluation Of Urban Co-Operative Banks In India", Asian Economic Review, 41(2): 323-330
- Asmild, M., Paradi, J.C., Reese, D.N. And Tam, F.(2007): Measuring Overall Efficiency and Effectiveness using DEA, European Journal Of Operational Research, 178,305-321.
- Bhattacharyya, Arunava, Lovell, C.A.K. And Sahay, Pankaj., (1997). The Impact Of Liberalization On The productive Efficiency Of Indian Commercial Banks. European Journal Of Operational Research, Vol. 98, Issue 2, 332-346.
- Subbarayan, A. And Prakash, V.(2009). "Ranking The Efficiency Performance Of Selected Set Of States In India By DEA." Bulletin Of Pure And Applied Sciences, Vol.28 E(No.1) Pp-133-145.



13. Taha, H.A (1997), Operational Research - An Introduction, Prentice-Hall Inc., New Jersey, USA.
14. Ramanathan, R. (1966). "An Introduction To Data Envelopment Analysis: A Tool For Performance Measurement, Sage Production Team.
15. Cooper, W.W., Seiford, L.M., and Tone, K (2000): Data Envelopment Analysis, Kluwer Academic Publishers.
16. Taguchi, G., Wu, Y., and Chowhury, S (2000): Mahalanobis – Taguchi System, McGraw – Hill Professional.
17. Morita, H and Habu, Y(2005): Variable Selection in data envelopment analysis based on external information. Proceedings of the eighth Czech – Japan Seminar 2005 on Data Analysis and Decision Making under Uncertainty, 181 – 187.
18. Edirisinghe, N.C.P., and Zang, X. (2007): Generalized DEA Model of fundamental analysis and its application to portfolio optimization. Journal of Banking and Finance, 31, 3311 – 3335.
19. Seiford, L.M., and Zhu, J. (2003): Context – dependent data envelopment analysis: attractiveness and progress, Omega, 31, 397 – 408.
20. Avkiran, N.K., and Fukuyama, H. (2008): An illustration of network SBM with simulated profit centre data for Japanese regional banks. Proceedings of DEA Symposium, 13 – 22.
21. Hirotsu, N., and Ueda, T., (2008): An evaluation of diversity in DEA. Proceedings of DEA Symposium 2008, 108 – 111.
22. Jeff Wu, C.F., and Hamada, M. (2000): Experiments – Planning, Analysis and Parameter Design Optimization, John Wiley & Sons, Inc.
23. Du, D.Z., and Pardalos, P.M. (1999): Handbook on Combinatorial Optimization, Springer.
24. Avkiran, N.K., and Morita, H.: Predicting bank stock performance with a model of fundamental relative analysis.
25. Jahanshahloo, G.R., and Hosseinzadeh, L.F. (2004): Introduction to Data Envelopment Analysis.
26. Morita, H., and Avrikran, N.K. (2009): Selecting Input and Output Variables in Data Envelopment analysis by designing statistical experiments. Journal of the Operations Research Society of Japan, 52(2), 163 – 173.
27. HulyaBayrak, OdayJarjies and Kubra Durukan, Using of Fractional Factorial Design ( $r^{k-p}$ ) in Data Envelopment Analysis to selection of Outputs and Inputs.

### AUTHORS PROFILE



**Mrs. C.V. Pradeepa**, received his M.Sc., M.Phil., in Statistics from University of Madras, Chennai, Tamilnadu, India. Currently, doing her part – time Doctorate in Research in Statistics under the guidance of Dr. V. Prakash, Associate Professor and Head, Department of Statistics, Presidency College (Autonomous), Chennai, Tamilnadu, India and also working as an Assistant Professor, Department of

Statistics, Justice Basheer Ahmed Sayeed College for Women (Autonomous), Chennai., Tamilnadu, India. She has totally 11 years of teaching experience in both UG and PG courses in Statistics. Her areas of specializations are Design of Experiments, Regression Analysis and Data Envelopment Analysis. She has published 2 research articles in National and International Journals of repute.



**Dr. V. Prakash** received his M.Sc., M.Phil., Ph.d., in Statistics from University of Madras, Chennai, Tamilnadu, India. Currently, he is working as an Associate Professor and holding the responsibility as Head, Department of Statistics, Presidency College (Autonomous), Chennai. He has totally 29 years of teaching experience in both UG and PG courses in Statistics. He has nearly 15 years of

research experience and assisted 25 M.Sc students to do their project work. He has guided and guiding M.Phil., and Ph.d., scholars. So far 4 of his candidates were awarded Ph.d degree by University of Madras and still half a dozen persons carrying out his research under his guidance. His areas of specializations are Optimization Techniques, Data Envelopment Analysis and Stochastic Frontier Analysis. He has published more than 20 research articles in National and International Journals of repute. He authored and acted as Chairperson for Business Mathematics and Statistics book for Higher Secondary students published by SCERT, Tamilnadu Text book Society.