

A Study to Determine the Higher Education Enrollments of West Bengal Prior to and Subsequent to Incorporation of e-Learning: A Discriminant Analysis Approach

Anirban Das, Member, IACSIT, Subhabrata Patra, Kajla Basu, and Dilip K. Banerjee

Abstract— In recent days, learning has radically changed and become beyond the conventional way of classroom teaching. With the advent of e-learning, the higher education aspiring students can leverage knowledge at their own time, own place and own pace. In West Bengal, e-learning has not been incorporated as a tool for higher education and as a result of which the GER is not at all impressive. This paper discusses the percentage of students' enrollment in higher education in traditional way of learning and after incorporation of e-learning in higher education up to what extent the percentage increases. This paper also classifies the students who are willing to go for higher education along with the students who are not willing to go for higher education in both traditional ways of education as well as when e-learning is incorporated in higher education using Fisher's Linear discriminant analysis.

Index Terms— Covariance matrix, e-learning, Fisher's linear discriminant analysis, GER, Kolmogorov-Smirnov Z normality test, pooled estimator

I. INTRODUCTION

In global scenario it has been already proved that e-learning has revolutionary impact in enhancing the gross enrollment ratio (GER). In West Bengal e-learning has not yet been applied in higher education which results to only 26.23% enrollments in higher education. Basically there are several reasons for which the students of West Bengal, who may have the opportunity to go for higher education, are least interested to it. According to the students' perspective, the reasons of not getting interested in higher education in West Bengal are as in Fig. 1.

In the Fig. 1, the reasons of students' less interests in higher education are represented with some values which are represented as:

A= [Improper guidance on higher education leads them least interested to it]

B= [Less number of higher education institutes]

C= [No secured job after doing higher education]

D= [Opportunity for average or below average students is not present in higher education]

Manuscript received September 29, 2011; revised October 25, 2011.

Anirban Das is with the Department of Computer Applications, Dr. B. C. Roy Engineering College, Durgapur, West Bengal, India (Phone: +919474623504, email: anirban-das@live.com)

Subhabrata Patra is with Department of Mathematics, Bengal College of Engineering & Technology, Durgapur, West Bengal, India (Phone: +919832157999, e-mail: suvi.b.patra@gmail.com).

Kajla Basu is with Department of Mathematics, National Institute of Technology, Durgapur, West Bengal, India (Phone: +919474532576, e-mail: kajla.basu@gmail.com).

Dilip K. Banerjee is with Globsyn Business School, Kolkata, West Bengal, India (Phone: +917278003690, e-mail: dkbanrg@gmail.com).

E= [Expensive higher education]

F= [Poverty]

G= [Limited seats for desired subjects in higher education institutes]

H= [Early marriage]

I= [Distance from home to higher education institutes and bad transportation]

J= [They become earning member of family]

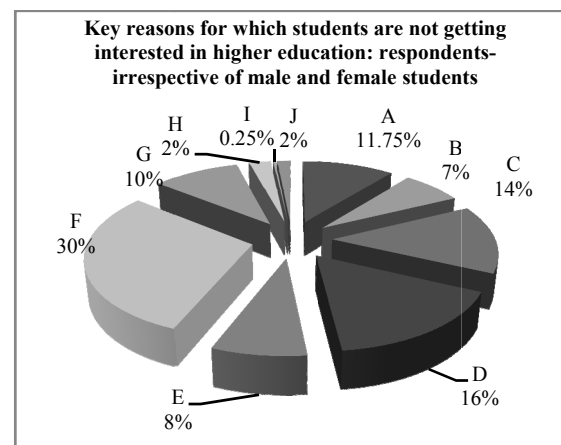


Fig. 1. Key reasons for which students of West Bengal are not getting interested to higher studies

This paper delineates that the responses, given by the students i.e. whether they opt for higher studies or not, are up to what extent in reality they will go for it or do not go for it. This classification is made here through Fisher's linear discriminant analysis concepts so that it will be clear that what percentage of eligible students will go for higher education through conventional system of higher education and higher education through e-learning.

II. PRELIMINARIES

A. Kolmogorov – Smirnov Normality

To use a parametric test, it's often assumed that the population under investigation is normal. Kolmogorov – Smirnov test is used for verifying that whether a sample comes from a population with some known distribution or not. The Kolmogorov – Smirnov test can be modified to serve as goodness of fit test.

In the special case of testing for Normality of the distribution, samples are standardized and compared with the standard normal distribution. This is equivalent to setting the mean and variance of the reference distribution equal to the sample estimates.

B. Linear Discriminant Analysis

There are many possible techniques for classification of data. Linear Discriminant Analysis (LDA) is the commonly used technique for data classification. This technique divides the whole data set into the groups, classifying the membership of a particular case or data value or scale value (given by a respondent). Here the two group classification has been used with the help of Fisher's Linear Discriminant analysis, taking the data samples which were originally from two group population.

III. METHODOLOGIES AND RESULTS

From 8 districts of West Bengal 300 students of rural schools are taken as the respondent sample. They were asked several questions about the reasons for which the students are unable to enroll themselves in higher education in West Bengal. Eight reasons were already collected from the pilot survey and these eight questions were placed to them. In accordance with them the importance of the questions regarding less enrollments in higher education are plotted graphically in Fig.2.

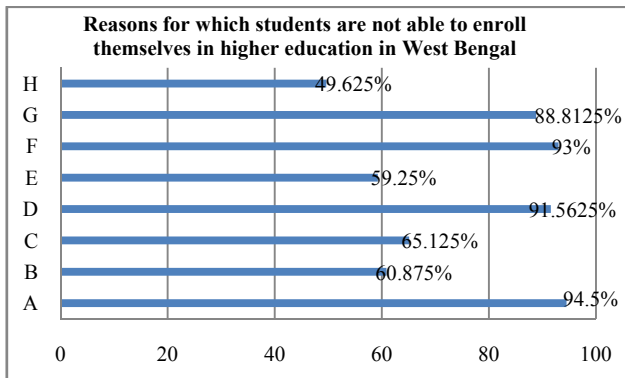


Fig. 2. Reasons of inability of students to enroll themselves in higher education in West Bengal

In Fig.2 the ordinates represent:

A= [After fulfilling basic needs of family, it becomes difficult to bear expenses for higher education]

B= [Due to engagement with family occupation; students, even those who are aspiring to higher education, get no time to study]

C= [Distance along with transportation system between residence and higher education institutes is not that much good, which are key reasons for not going for it]

D= [A very few number of seats and less opportunities are there in higher education institutes in West Bengal for mediocre students. So they are not getting chances for desired subjects due to competition]

E= [The least interests of students in higher education make them reluctant to it]

F= [Students do not get proper guidance which leads them in not going for higher education]

G= [Involvements of politics in higher educational institutes in West Bengal affect the environment of education that leads to fewer enrollments in higher education]

H= [Job opportunities of fresh secondary or below secondary students in retail based industries demotivate student's higher education]

In the data sample the cases arose according to the scaling values [1 - 4] given by the respondents i.e. 300 students of rural schools of 8 districts of West Bengal. Now we are to determine that the scaling values in terms of the reasons of inability of students' enrollment in higher education, they provided, are at all matching with their willingness decisions [YES/NO]. Now we need to classify them into two categories [YES and NO] with the help of the scaling values they provided. Now the data sample has been channelized through Kolmogorov-Smirnov Z normality test to check whether the data sample is applicable for discriminant analysis or not. The normality test satisfied well. The correlations among variables i.e. the scaling of the reasons are very low. And hence we can take the data sample suitable for discriminant analysis.

From the data sample of 300 respondents in traditional learning we got, 163 candidates responded 'YES' and 137 candidates responded 'NO' in terms of going for higher education. Here taking $n_1=163$ and $n_2=137$, we can have a pooled estimator for the covariance of the original population as

$$S_{pooled} = \frac{n_1-1}{(n_1-1)+(n_2-1)} S_1 + \frac{n_2-1}{(n_1-1)+(n_2-1)} S_2 = 0.544S_1 + 0.456 S_2$$

where S_1 and S_2 are the covariance matrices for the 'YES' and 'NO' groups respectively. And the matrices S_1 and S_2 obtained are as follows:

$$S_{1(YES)} = \begin{bmatrix} .281 & -.024 & .034 & -.026 & .020 & .059 & .007 & -.045 \\ & .854 & -.101 & .080 & -.189 & -.079 & .080 & -.023 \\ & & 1.025 & -.022 & .155 & .104 & -.209 & .126 \\ & & & .473 & .015 & .082 & .043 & -.101 \\ & & & & 0.764 & .096 & -.140 & .125 \\ & & & & & .462 & -.047 & -.089 \\ & & & & & & .523 & -.071 \\ & & & & & & & .784 \end{bmatrix}$$

$$S_{2(NO)} = \begin{bmatrix} .250 & -.120 & .018 & -.036 & .067 & .041 & .006 & .028 \\ & .717 & -.082 & .068 & -.047 & -.012 & .008 & -.096 \\ & & 1.098 & -.011 & .277 & .117 & -.129 & .063 \\ & & & .531 & -.068 & .068 & .059 & -.027 \\ & & & & 0.751 & .109 & -.046 & .079 \\ & & & & & .358 & -.094 & .017 \\ & & & & & & .535 & -.105 \\ & & & & & & & .721 \end{bmatrix}$$

$$S_{pooled} = \begin{bmatrix} .2669 & -.0678 & .0267 & -.0306 & .0414 & .0508 & .0065 & -.0117 \\ & .7915 & -.0923 & .0745 & -.1242 & -.0484 & .0472 & -.0563 \\ & & 1.0583 & -.0170 & .2106 & .1099 & -.1725 & .0973 \\ & & & .4994 & -.0228 & .0756 & .0503 & -.0673 \\ & & & & 0.7581 & .1019 & -.0971 & .1040 \\ & & & & & -.4146 & -.0684 & -.0407 \\ & & & & & & .5285 & -.0865 \\ & & & & & & & .7553 \end{bmatrix}$$

Here the data sample we have got according to the scaling values of the respondents between 1 and 4. Now we need to verify the responses (YES/NO) for authenticity. We need to classify the respondents according to the two groups for YES and NO, whether a respondent who opines YES (NO) for higher studies is predicted for (not) aspiring to higher studies or not. To classify the respondents into the said groups we are to make use of Fisher's linear discriminant analysis approach. Let's suppose there exist k no. of categories denoted by π_k , in the actual population consisting of the covariance matrix Σ , where k denotes the categories.

According to R. A. Fisher the Minimum Total Probability Misclassification rule for normal populations with unequal covariance matrices is given by

Allocate X to π_k if the linear discriminant score

$$\hat{d}_k(X) = \max[\hat{d}_1(X), \hat{d}_2(X), \hat{d}_3(X), \dots, \hat{d}_g(X)]$$

Here $g=2$, the two groups for YES and NO. And the number of categories in the original population is also $k = 2$. The estimate of the linear discrimination score

$$\hat{d}_i(X) = \bar{X}_i S_{pooled}^{-1} X - \frac{1}{2} \bar{X}_i S_{pooled}^{-1} \bar{X}_i + \ln p_i$$

where $i=1 \sim \text{YES}$, $2 \sim \text{NO}$; and p_i are the respective prior probabilities. An estimate $\hat{d}_k(X)$ of the linear discriminant score $d_i(X)$ is based on the pooled estimate of the population covariance Σ i.e.

$$S_{pooled} = \frac{1}{n_1 + n_2 + \dots + n_{g-1}} [(n_1 - 1)S_1 + (n_2 - 1)S_2 + \dots + (n_{g-1} - 1)S_{g-1}]$$

This pooled estimate is used in the position of covariance matrix in the discrimination score. We have already got the pooled estimator earlier. This way the two discrimination scores we can have:

$$\hat{d}_1(X) = (13.8591, 4.3244, 2.2361, 6.3099, 1.8934, 7.0923, 8.1818, 4.5092)X - 79.075$$

$$\hat{d}_2(X) = (13.9384, 4.3331, 2.3579, 6.1030, 1.6799, 7.3594, 8.0706, 4.6225)X - 79.498$$

Here X = random vector consisting of the random variables which represent the scaling values of the reasons in an individual case.

Assume some particular cases containing the scale values of 1 to 4 scale given by four respondents are

Case1: A=4, B=1, C=2, D=2, E=2, F=3, G=4, H=1 (as per Fig. 2) opined YES

Case2: A=4, B=3, C=2, D=2, E=1, F=4, G=4, H=2 (as per Fig.2) opined YES

Case3: A=3, B=4, C=3, D=2, E=1, F=2, G=4, H=1 (as per Fig.2) opined NO

Case4: A=4, B=1, C=1, D=4, E=1, F=4, G=2, H=3 (as per Fig.2) opined NO

Then the authenticity of their opinions [YES/NO] is to be checked. Taking these as individual cases as

$$\text{Case1: } X = [4 \ 1 \ 2 \ 2 \ 2 \ 3 \ 4 \ 1]^T$$

$$\text{Case2: } X = [4 \ 3 \ 2 \ 2 \ 1 \ 4 \ 4 \ 2]^T$$

$$\text{Case3: } X = [3 \ 4 \ 3 \ 2 \ 1 \ 2 \ 4 \ 1]^T$$

$$\text{Case4: } X = [4 \ 1 \ 1 \ 4 \ 1 \ 4 \ 2 \ 3]^T$$

The respective Fisher's linear discriminant scores for group 1 (~YES) and group 2 (~NO):

For Case1: $\hat{d}_1(X) = 60.0776$ and $\hat{d}_2(X) = 59.8534$, here $\hat{d}_1(X) > \hat{d}_2(X)$ and opined YES so, respondent is classified that is it is predicted that the respondent will go for higher education. [YES \rightarrow YES]

For Case2: $\hat{d}_1(X) = 78.4348$ and $\hat{d}_2(X) = 78.8216$, here $\hat{d}_1(X) < \hat{d}_2(X)$ and opined YES so, respondent is misclassified that is it is predicted that the respondent will not go for higher education. [YES \rightarrow NO]

For Case3: $\hat{d}_1(X) = 52.4424$ and $\hat{d}_2(X) = 52.2329$, here $\hat{d}_1(X) > \hat{d}_2(X)$ and opined NO so, respondent is misclassified

that is it is predicted that the respondent will go for higher education. [NO \rightarrow YES]

For Case4: $\hat{d}_1(X) = 68.3153$ and $\hat{d}_2(X) = 68.4848$, here $\hat{d}_1(X) < \hat{d}_2(X)$ and opined NO so, respondent is classified that is it is predicted that the respondent will not go for higher education. [NO \rightarrow NO]

Out of 300 respondents 163 number of respondents opined YES and 137 respondents opined NO initially. Proceeding the methods used in the afore explained cases the classification result as matrix form is:

		Predicted memberships	
		YES	NO
Actual Membership	YES	125	38
	NO	86	51

So here 58.67% cases are correctly classified among which 71.02% respondents are willing (YES \rightarrow YES) for higher studies.

After proposing e-learning to the respondents out of 300 respondents, 229 respondents opined YES and 71 respondents opined NO. Proceeding the methods used in the previous cases the classification result as matrix form is: cases the classification result as matrix form is:

		Predicted memberships	
		YES	NO
Actual Membership	YES	225	4
	NO	68	3

So here 76% cases are correctly classified among which 98.68% respondents are willing (YES \rightarrow YES) for higher studies.

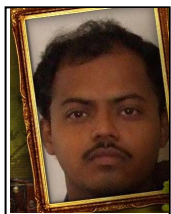
On the basis of uniform classification, the willing respondents for higher studies in traditional system of higher education are 41.66% whereas the willing respondents for higher education through e-learning are 75%. So the gain in terms of higher education enrollments through e-learning based on Fisher's linear discriminant analysis is 33.34% i.e. these 33.34% respondents, opined NO previously for higher studies through traditional system, opined YES for e-learning based higher education.

IV. CONCLUSION

E-learning has already shaped a revolutionary benchmark in higher education round the globe. Through e-learning, an approach can be undertaken to boost the enrollments in higher education in West Bengal. From the whole respondent sample, the willing or unwilling respondents are skillfully classified that these percentage of respondents will opt higher studies and these will not in both conventional way of higher education and also higher education through e-learning. After the classification it is seen that there is radical gain in the responses when they are suggested higher education through e-learning i.e. a major percentage of respondents voted their willingness to higher education through e-learning in West Bengal. It can be concluded that higher education system in West Bengal should be amalgamated with e-learning so that the said gain in enrollments in higher education can be visualized in reality.

REFERENCES

- [1] A. Das, D. K. Banerjee, K. Basu, "Implementation of E-Learning in West Bengal to Enhance the Present GER in Higher Education," *International Journal of Innovation, Management and Technology*, vol. 2, no. 3, pp. 257-261, 2011.
- [2] McLachlan, "Discriminant Analysis and Statistical Pattern Recognition", Wiley Interscience 2004.
- [3] A. B. Martinez and Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2): 228-233, 2004.
- [4] S. Mika et al., "Fisher Discriminant Analysis with Kernels". *IEEE Conference on Neural Networks for Signal Processing IX*: 41-48. doi:10.1109/NNSP.1999.788121, 1999
- [5] C. J. Huberty, *Applied discriminant analysis*, Wiley and Sons New York, 1994
- [6] C. J. Huberty, R. M. Barton, *An introduction to discriminant analysis. Measurement and Evaluation in Counseling and Development*, 1989, pp. 22, 158-168
- [7] H. Abdi, *Discriminant correspondence analysis*, In: N.J. Salkind (Ed.) *Encyclopedia of Measurement and Statistics*. Thousand Oaks (CA): Sage, 2007, pp. 270-275.
- [8] W. R. Klecka, *Discriminant analysis*, Beverly Hills, CA: Sage, 1980.
- [9] R. A. Johnson, D. W. Wichern. "Applied Multivariate Statistical Analysis" PHI, 2009



Anirban Das is pursuing his research work in the Department of Mathematics from National Institute of Technology, Durgapur, India. He is a member of IAENG and IACSIT. He is having around 4 years of work experience including academics and industry. He started his career with HCL Technologies as a software Engineer. He is Microsoft Certified Technology Specialist (MCTS) in Microsoft Office Sharepoint Server 2007. Presently he is the Assistant Professor in Computer Applications department in Dr. B. C. Roy Engineering College, Durgapur, India. He has around 10 research papers in various eminent journal and conferences. Mr. Das has interests in Statistical Analysis, e-commerce, e-learning and computer applications as domain of research.



Subhabrata Patra is an Assistant Professor in Department of Mathematics in Bengal College of Engineering and Technology, Durgapur, India. Formerly he was Lecturer in National Power Training Institute, Durgapur, India. He passed M.Sc in Mathematics from Guru Ghasidas University, Bilaspur, India (2008) and M.Tech in Operations Research from National Institute of Technology, Durgapur, India (2011). Mr. Patra has interests in various fields such as Statistical analysis, Operations Research etc.



Kajla Basu is as Associate Professor in department of Mathematics in National Institute of Technology, Durgapur, India. She passed M. Sc. in Mathematics with national scholarship along with C.S.I.R, J.R.F and SRF (for 5 years) and Ph.D degree from IIT, Kharagpur. She is having around 15 years of work experience at National Institute of Technology, Durgapur. She has around 13 research papers in various eminent journals and conferences. Prof. Basu is a life member of ORSI and her domain of research interests are in Statistical Analysis, Operations Research and Fuzzy Mathematics.



Dilip Kumar Banerjee, Principal, Globsyn Business School, Kolkata since January 2011 has had his education in Mathematics and Statistics from Ranchi University and Bhagalpur University. He joined as a Research Fellow in Computer Science and Information Technology in Indian Statistical Institute, Kolkata and obtained his Ph. D. degree from the Jadavpur University. He attained the M.B.A degree from Sikkim Manipal University while working at the Sikkim Manipal Institute of Technology. He worked as a Scientist in the prestigious Knowledge Based Computer Systems project of ISI, Kolkata. Selected as a UNDP fellow from ISI, he went to the University of Washington, Seattle for advanced training and research in applications of Artificial Intelligence techniques. During 1991-1995 he visited France and Europe several times as a Fellow of the International Centre for Pure and Applied Mathematics (CIMPA) and worked with the leading researchers of his field. He taught at the undergraduate and postgraduate levels at ISI, Kolkata, Jadavpur University, Sikkim Manipal Institute of Technology (as the Head of the departments of Computer Science and Information Technology departments). He was the Director of the Durgapur Institute of Advanced Technology and Management during 2005-2007 and the Director of the Asia Pacific Institute of Management, New Delhi during 2008-2010. Besides academic experience for about 20 years he also worked in industry for about 15 years in Sarabhai Chemicals and Leica Imaging System Ltd., Cambridge, UK. Dr. Banerjee has over 35 research articles published in prominent national and international journals and conferences.