

P A R T I I

Note on Part II

In Part I, (we) have dealt with the prediction of the degree of success, separately for each subject-language, Mathematics and Science and lastly the overall degree of success as measured by PSc Grand Total Percent. Though this is very useful for validation research, yet accurate measurement and evaluation of complex mental processes is so difficult that the prediction, by any instrument or method, remains far from being perfect. For instance, even if the prediction in terms of correlation reaches the upper level of accuracy of .70 (i.e. $r = .70$) the variance accounted for is only nearly 50 percent ($r^2 \times 100 = 49.00$) and a large residual variance (about an equal amount) remains unaccounted for. This makes the interpretation of the implications of the results all the more difficult from the practical view-point. Moreover, the success in the PSc Examination is not determined on the basis of the Grand Total Percent. It then becomes practically more meaningful to obtain prediction in terms of the two categories of Pass-Fail dichotomy and efficiency of such classification. Part II therefore deals with the discriminant analysis of this type.

CHAPTER V

PREDICTION OF COLLEGE SUCCESS BY LINEAR DISCRIMINANT

Introduction

High failure rate in our universities and subsequent wastage in higher education has been a concern for educationalists since long. Crores of rupees are wasted after those who do not make successful careers at college. Besides this, there occurs a lot of waste of energy, time and money resulting in failure and frustration for the individual and huge waste of human resources and manpower for the nation.

If wastage at university level is to be reduced, the problem of admissions to universities has to be tackled. This means that only those who have capacity to go successfully through the course in question, should be admitted to the university. The selection of such suitable persons presents a classificatory problem i.e. the problem of classifying whether individuals are fit or unfit for the given course.

Problem

The college admission officer then is confronted with the problem of predicting success and failure of the individuals who seek admissions and prior to this, it is necessary for him to know how far his method of selection and the measuring instrument at his disposal would be reliable. This study attempts to find answer to this.

At present, in most universities in India, admissions are given on the basis of Secondary School Certificate, Examination marks. It is the purpose of this study to examine in the first instance, how far the Secondary School Certificate Examination marks in English and Mathematics would be reliable for selecting right persons for admission to the Preparatory Science Course which leads to other important professional course such as Engineering and Medicine. The study is in relation to Secondary School Certificate Examination (1957) of former Bombay State and Preparatory Science Examination (1958) of the M.S. University of Baroda. The solution to the above problem can be obtained by the following Discriminant Analysis Method.

Method

The discriminant problem starts with the data on two populations, say P_1 and P_2 , one representing the individuals who are fit, and the other population of the individuals who are unfit for the given course. This data on two groups is further analysed to obtain discriminant function to be used for predicting the individuals fit or unfit for the course in question i.e. for predicting whether individuals would pass or fail in the course.

The aforementioned data on 278 students along with their SSCE marks, was split up into two groups-Pass Group and Fail Group-according as the students passed or failed in the Preparatory science course of the M.S. University of Baroda. The variables selected in this analysis are marks in : (1) SSCE English, (2) SSCE Higher Mathematics.

As both methods- (1) due to R.A. Fisher and (2) Abraham Wald are closely related to each other and yield almost similar results under the assumptions, the analysis was first done by Discriminant Function and then Wald's U-statistic was used for classification.

Analysis

The statistical computations necessary for the analysis are shown below step by step.

Notations

Let N_1 and N_2 be the number in the Pass Group and Fail Group respectively.

x_1 and y_1 denote English variable in the Pass Group and Fail Group respectively.

x_2 and y_2 denote Mathematics variables in the Pass Group and Fail Group respectively.

Step I. Computation of Summations:

N_1	= 175	N_2	= 103
$\sum x_1$	= 10511	$\sum y_1$	= 5388
$\sum x_2$	= 12636	$\sum y_2$	= 6122
$\sum x_1^2$	= 643429	$\sum y_1^2$	= 289782
$\sum x_2^2$	= 942308	$\sum y_2^2$	= 378284
$\sum x_1 x_2$	= 763577	$\sum y_1 y_2$	= 320181

Step 2. Computation of Means and Mean-differences:

$$\begin{aligned}\bar{x}_1 &= 60.0629 & \bar{y}_1 &= 52.3107 & d_1 &= 7.7522 \\ \bar{x}_2 &= 72.2057 & \bar{y}_2 &= 59.4369 & d_2 &= 12.7688\end{aligned}$$

Step 3. Computation of Matrix of within-Group sums of squares and products.

$$S_{ij} = \sum (x_i - \bar{x}_i)(x_j - \bar{x}_j) + \sum (y_i - \bar{y}_i)(y_j - \bar{y}_j)$$

$$(S_{ij}) = \begin{bmatrix} 20040.37 & 4557.76 \\ 4557.76 & 44327.93 \end{bmatrix}$$

Step 4. Computation of Inverse Matrix (S^{ij}):

$$(S^{ij}) = \begin{bmatrix} .000051094066 & -.000005253448 \\ -0.000005253448 & .000023099290 \end{bmatrix}$$

Step 5. Computation of D, relative weights and F-ratio:

$$D = .0003290120X_1 + .0002542250X_2$$

Relative	44	56
Weights	Per cent	Per cent

$$F_{2, 275} = 51.68^{***}$$

Step 6. Computation of the classification equation and U-Statistic:

$$U = .09080731X_1 + .07016610X_2$$

Step 7. Computation of the Critical Region:

$$A_1 = .09080731\bar{x}_1 + .07016610\bar{x}_2 = 10.52054215$$

$$A_2 = .09080731\bar{y}_1 + .07016610\bar{y}_2 = 8.92064890$$

$$\frac{1}{2}(A_1 + A_2) = 9.720595$$

Therefore

For $U > 9.72$ the individual is classified as coming from P_1 population-Pass.

$U \leq 9.72$ the individual is classified as coming from P_2 population-Fail

Step 8. Computation of the Error of classification and thereby assessing the Efficiency of classification:

$$\begin{aligned} \bar{\sigma}^2 = & s^{11} (\bar{x}_1 - \bar{y}_1) (\bar{x}_1 - \bar{y}_1) + s^{12} (\bar{x}_1 - \bar{y}_1) (\bar{x}_2 - \bar{y}_2) \\ & + s^{21} (\bar{x}_2 - \bar{y}_2) (\bar{x}_1 - \bar{y}_1) + s^{22} (\bar{x}_2 - \bar{y}_2) (\bar{x}_2 - \bar{y}_2) \end{aligned}$$

$$\bar{\sigma}^2 = 1.59989334$$

$$\therefore \bar{\sigma} = \sqrt{1.59989334} = 1.2649$$

$$\frac{A_2 - A_1}{2\bar{\sigma}} = .6324$$

$$p_1 = 1 - p_2 = \frac{1}{\sqrt{2\pi}} \int_{.6324}^{\infty} e^{-t^2/2} dt = .2635 \text{ or } 26.35\%$$

where p_1 is the probability of making an error of Type I, that is, of classifying a student as one who will go successfully through the course when he actually does not; and $1-p_2$ is the probability of making an error of Type II, that is of classifying a student as one who will fail in the course in question while he actually passes.

In using the above classification equation to classify 278 students used in this study, 23 errors of Type I were made or 22.3 per cent while 46 errors of Type II were made or 26.3 per cent. These percentages seem reasonably close to the expected 26.35 per cent. The following table will clarify the results obtained (Table 5.1):

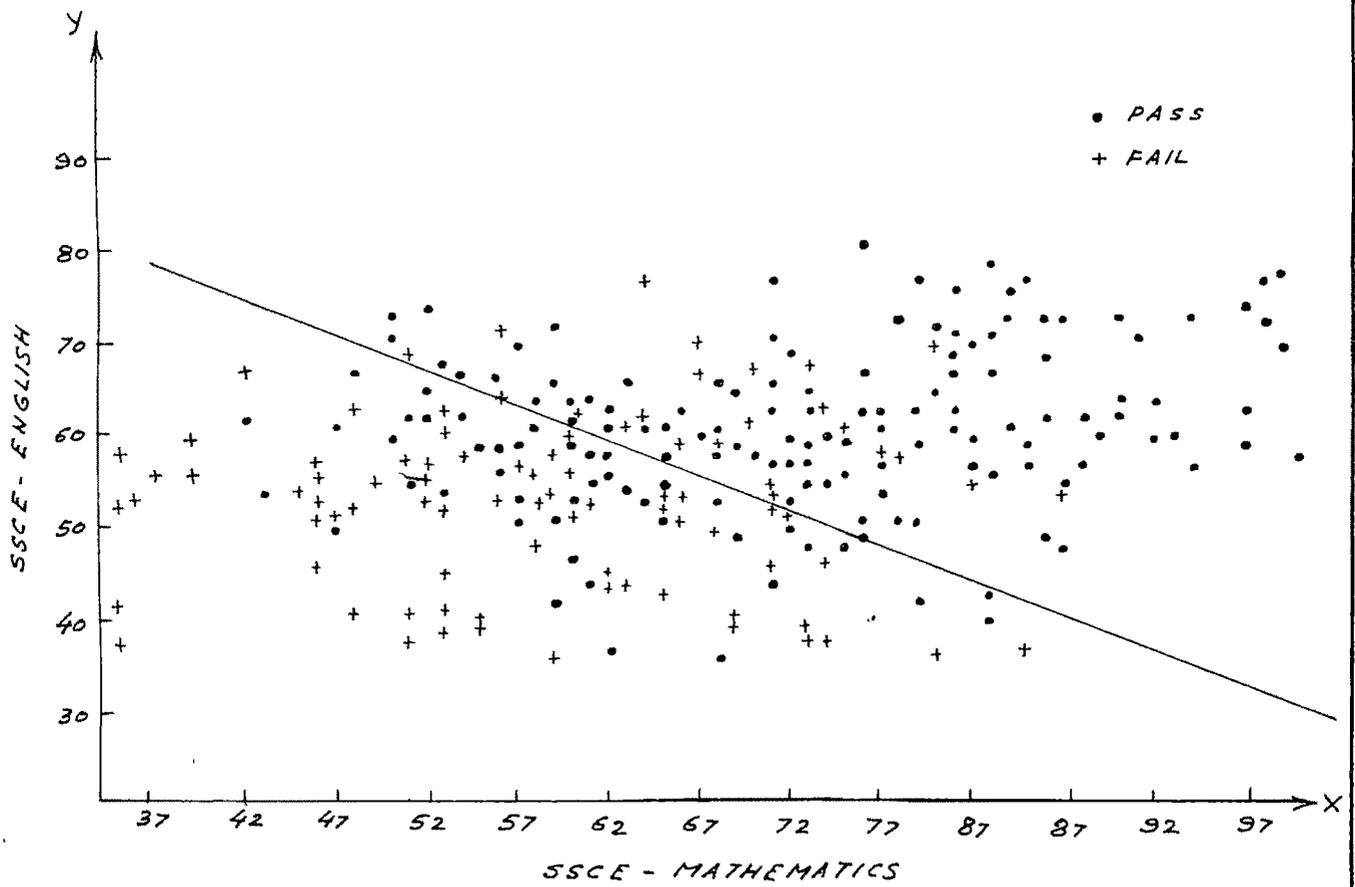
Table 5.1 Classification of Predicted Pass-Fail Dichotomy Versus Actually Observed with Two-Variable Discriminant

	Actually Pass	Actually Fail	Total
Predicted Pass (by Discriminant Function)	129	23	152
Predicted Fail (by Discriminant Function)	46	80	126
Total	175	103	278

Figure 3 represents the scatter-diagram of pass-fail points and the graph of the discriminant line.

Conclusion:

It can be concluded from the above study that the error of classification with regard to the measuring instrument-SSCE English and Mathematics is nearly 26 percent.



SCATTER-DIAGRAM OF PASS-FAIL POINTS
AND DISCRIMINANT LINE

FIG. 3

From Table 5.1, we observe that the number of correctly classified cases is $129 + 80 = 209$ and so, the observed efficiency of correct classification is $209 \times 100 / 278$, that is 75.2 percent. This percentage compares well with the theoretically derived one.