# CHAPTER VII

## PREDICTION OF COLLEGE SUCCESS BY MORE GENERAL QUADRATIC DISCRIMINANT

Hitherto we have assumed equality of variance - covariance matrices for the two groups of pass-fail dichotomy, and the results of the previous two chapters are based on this assumption.

Smith, Cedric A.(1947) has shown that when differences in variances and covariances of the two groups to be discriminated, cannot be assumed equal, the discriminant is a quadratic function of the normally distributed predictor-variables, which may be termed as Quadratic Discriminant.

### Problem

In our data under study, we observe considerable differences in the variance-covariance matrices for pass group and fail group as determined by the results of the

PScE. It is therefore the purpose of this study to investigate whether Quadratic Discriminant can be used to improve the prediction of pass-fail dictotomy, over the linear discriminant.

Method

Let-

$X_1$ denote English variable

$X_2$ denote Mathematics variable

$V_{x_1x_1}$ denote variance of the English variable in Pass Group

$V_{x_2x_2}$ denote variance of the Mathematics variable in Pass Group

$V_{x_1x_2}$ denote covariance of English-Mathematics variables in Pass Group

and

$V_{y_1y_1}$ denote variance of the English variable in Fail-Group

$V_{y_2y_2}$ denote variance of the Mathematics variable in Fail Group

$V_{y_1y_2}$ denote covariance of English-Mathematics variable in Fail Group

$V_{xx}$ denote variance-covariance matrix for Pass Group

$V_{yy}$ denote variance-covariance matrix for Fail Group

Then

$$V_{xx} = \begin{bmatrix} V_{x_1x_1} & V_{x_1x_2} \\ V_{x_1x_2} & V_{x_2x_2} \end{bmatrix}$$

and

$$V_{yy} = \begin{bmatrix} V_{y_1y_1} & V_{y_1y_2} \\ V_{y_1y_2} & V_{y_2y_2} \end{bmatrix}$$

Let $I_{xx}$ be the matrix inverse to $V_{xx}$

and $I_{yy}$ be the matrix inverse to $V_{yy}$

If-

$$I_{xx} = \begin{bmatrix} I_{x_1x_1} & I_{x_1x_2} \\ I_{x_1x_2} & I_{x_2x_2} \end{bmatrix}$$

where

$$I_{x_1x_1} = \frac{V_{x_2x_2}}{V_{x_1x_1} V_{x_2x_2} - V^2_{x_1x_2}}$$

$$I_{x_2x_2} = \frac{V_{x_1x_1}}{V_{x_1x_1} V_{x_2x_2} - V^2_{x_1x_2}}$$

$$I_{x_1x_2} = \frac{-V_{x_1x_2}}{V_{x_1x_1} V_{x_2x_2} - V^2_{x_1x_2}}$$

and

$$I_{yy} = \begin{bmatrix} I_{y_1y_1} & I_{y_1y_2} \\ I_{y_1y_2} & I_{y_2y_2} \end{bmatrix}$$

where $$I_{y_1 y_1} = \frac{V_{y_2 y_2}}{V_{y_1 y_1} V_{y_2 y_2} - V^2_{y_1 y_2}}$$

$$I_{y_2 y_2} = \frac{V_{y_1 y_1}}{V_{y_1 y_1} V_{y_2 y_2} - V^2_{y_1 y_2}}$$

$$I_{y_1 y_2} = \frac{- V_{y_1 y_2}}{V_{y_1 y_1} V_{y_2 y_2} - V^2_{y_1 y_2}}$$

then the Quadratic Discriminant (Dq) is given as-

$$Dq = \alpha_{x_1 x_1} X_1^2 + 2\alpha_{x_1 x_2} X_1 X_2 + \alpha_{x_2 x_2} X_2^2 + 2\alpha_{x_1} X_1 + 2\alpha_{x_2} X_2 + \alpha$$

where $$\alpha_{x_1 x_1} = I_{y_1 y_1} - I_{x_1 x_1}$$

$$\alpha_{x_1 x_2} = I_{y_1 y_2} - I_{x_1 x_2}$$

$$\alpha_{x_2 x_2} = I_{y_2 y_2} - I_{x_2 x_2}$$

$$\alpha_{x_1} = I_{x_1 x_1} M_{x_1} + I_{x_1 x_2} M_{x_2} - I_{y_1 y_2} M_{y_2} - I_{y_1 y_1} M_{y_1}$$

$$\alpha_{x_2} = I_{x_1 x_2} M_{x_1} + I_{x_2 x_2} M_{x_2} - I_{y_1 y_2} M_{y_1} - I_{y_2 y_2} M_{y_2}$$

$$\alpha = I_{y_1 y_1} M^2_{y_1} + 2I_{y_1 y_2} M_{y_1} M_{y_2} + I_{y_2 y_2} M^2_{y_2} - I_{x_1 x_1} M^2_{x_1}$$

$$-2I_{x_1 x_2} M_{x_1} M_{x_2} - I_{x_2 x_2} M^2_{x_2} - l_n \, W_x / W_y$$

$M_{x_1}$ being the mean of English variable in Pass Group

$M_{x_2}$ being the mean of Mathematics variable in Pass Group

$M_{y_1}$ being the mean of English variable in Fail Group

$M_{y_2}$ being the mean of Mathematics variable in Fail Group

and $l_n \; W_{x/W_y}$ being the natural logarithm of $\dfrac{V_{x_1x_1} V_{x_2x_2} - V^2_{x_1x_2}}{V_{y_1y_1} V_{y_2y_2} - V^2_{y_1y_2}}$

Computation-

For the data in question, the variance co-variance matrices for the two groups are:

$$V_{xx} = \begin{bmatrix} V_{x_1x_1} & V_{x_1x_2} \\ V_{x_1x_2} & V_{x_2x_2} \end{bmatrix} = \begin{bmatrix} 69{,}5880 & 26.5675 \\ 26.5675 & 171.9344 \end{bmatrix}$$

$$V_{yy} = \begin{bmatrix} V_{y_1y_1} & V_{y_1y_2} \\ V_{y_1y_2} & V_{y_2y_2} \end{bmatrix} = \begin{bmatrix} 77.7653 & -0.6371 \\ -0.6371 & 141.2876 \end{bmatrix}$$

The corresponding inverse matrices are-

$$I_{xx} = \begin{bmatrix} I_{x_1x_2} & I_{x_1x_2} \\ I_{x_1x_2} & I_{x_2x_2} \end{bmatrix} = \begin{bmatrix} 0.01527119 & -.00235972 \\ -.00235972 & .00618080 \end{bmatrix}$$

and

$$I_{yy} = \begin{bmatrix} I_{y_1y_1} & I_{y_1y_2} \\ I_{y_1y_2} & I_{y_2y_2} \end{bmatrix} = \begin{bmatrix} .01285968 & .00005799 \\ .00005799 & .00707802 \end{bmatrix}$$

$M_{x_1} = 60.06, \quad M_{x_2} = 72.21 \quad M_{y_1} = 52.31; \quad M_{y_1} = 59.44$

$$\alpha_{x_1x_1} = I_{y_1y_1} - I_{x_1x_1} = \begin{matrix} .01285968 \\ -.01527119 \end{matrix} = -.00241151$$

$$\alpha_{x_1x_2} = I_{y_1y_2} - I_{x_1x_2} = .00005799 + .00235972 = .00241771$$

$$\alpha_{x_2x_2} = I_{y_2y_2} - I_{x_2x_2} = .00707802 - .00618080 = .00089722$$

$$\alpha_{x_1} = I_{x_1x_1}M_{x_1} + I_{x_1x_2}M_{x_2} - I_{y_1y_1}M_{y_1} - I_{y_1y_2}M_{y_2}$$

$$= (.01527119)(60.06)+(-.00235972)(72.21)-(.01285968)\times (52.31) - (.00005799)(59.44)$$

$$= 0.07065550$$

$$\alpha_{x_2} = I_{x_1x_2}M_{x_1} + I_{x_2x_2}M_{x_2} - I_{y_1y_2}M_{y_1} - I_{y_2y_2}M_{y_2}$$

$$= (-.00235972)(60.06)+(.00618080)(72.21)-(.00005799)(52.31)-(.00707802)(59.44)$$

$$= -.11916018$$

$$\alpha = I_{y_1y_1}M_{y_1}^2 + 2I_{y_1y_2}M_{y_1}M_{y_2} + I_{y_2y_2}M_{y_2}^2 - I_{x_1x_1}M_{x_1}^2 - 2I_{x_1x_2}M_{x_1}M_{x_2} - I_{x_2x_2}M_{x_2}^2 - \ln \frac{W_x}{W_y}$$

$$=(.01285968)(2736.34)+(.00011598)(3109.31)$$
$$+(.00707802)(3533.11)-(.01527119)(3607.20)$$
$$-(-.00471944)(4336.93)-(.00618080)(5214.28)-1_n \frac{11258.74}{10986.87}$$

$$=-6.29027969-1_n(1.0247)$$
$$=-6.29027969-(0.0106)(2.3026)$$
$$=-6.31468725$$

The Quadratic Discriminant so computed is given as-

$$Dq = -0.00241151X_1^2 + 2(.00241771)X_1X_2 + .00089722X_2^2$$
$$+2(.07065550)X_1 - -2(.11916018)X_2-6.31468725$$

We use this discriminant to classify 278 students in the data. The test criterion used for classification is as follows:

$Dq \geqslant 0$ , the individual is classified as belonging to Pass Group

$Dq \leqslant 0$ , the individual is classified as belonging to Fail Group

In using this Quadratic Discriminant to classify 278 students in the data, 21 errors of type I were observed, while 55 errors of type II were observed. Though there is some improvement in type I error, type II error is increased, compared to two variable discriminant.

We further study a particular case of Quadratic Discriminant, in which the correlation between $x_1$ and $x_2$ being very small, is assumed to be equal to zero.

Particular Case: $V_{x_1x_2} = V_{y_1y_2} = 0$

Then-

$$V_{xx} = \begin{bmatrix} V_{x_1x_1} & V_{x_1x_2} \\ V_{x_1x_2} & V_{x_2x_2} \end{bmatrix} = \begin{bmatrix} 69.5880 & 0.0 \\ 0.0 & 171.9344 \end{bmatrix}$$

Similarly-

$$V_{yy} = \begin{bmatrix} V_{y_1y_1} & V_{y_1y_2} \\ V_{y_1y_2} & V_{y_2y_2} \end{bmatrix} = \begin{bmatrix} 77.7653 & 0.0 \\ 0.0 & 141.2876 \end{bmatrix}$$

The corresponding inverse matrices are-

$$I_{xx} = \begin{bmatrix} 0.01437029 & 0.0 \\ 0.0 & 0.00581617 \end{bmatrix}$$

and

$$I_{yy} = \begin{bmatrix} 0.01285921 & 0.0 \\ 0.0 & 0.00707776 \end{bmatrix}$$

$$\alpha_{x_1x_1} = I_{y_1y_1} - I_{x_1x_1} = -0.00151108$$

$$\alpha_{x_2x_2} = I_{y_2y_2} - I_{y_1y_1} = 0.00126159$$

$$\alpha_{x_1} = I_{x_1x_1} M_{x_1} - I_{y_1y_1} M_{y_1}$$

$$= 0.19041434$$

$$\alpha_{x_2} = I_{x_2 x_2} M_{x_2} - I_{y_2 y_2} M_{y_2}$$

$$= -\ 0.00071642$$

$$\alpha = I_{y_1 y_1} M^2_{y_1} + I_{y_2 y_2} M^2_{y_2} - I_{x_1 x_1} M^2_{x_1} - I_{x_2 x_2} M^2_{x_2} - l_n \left(\frac{W_x}{W_y}\right)$$

$$= -\ 22.05563039$$

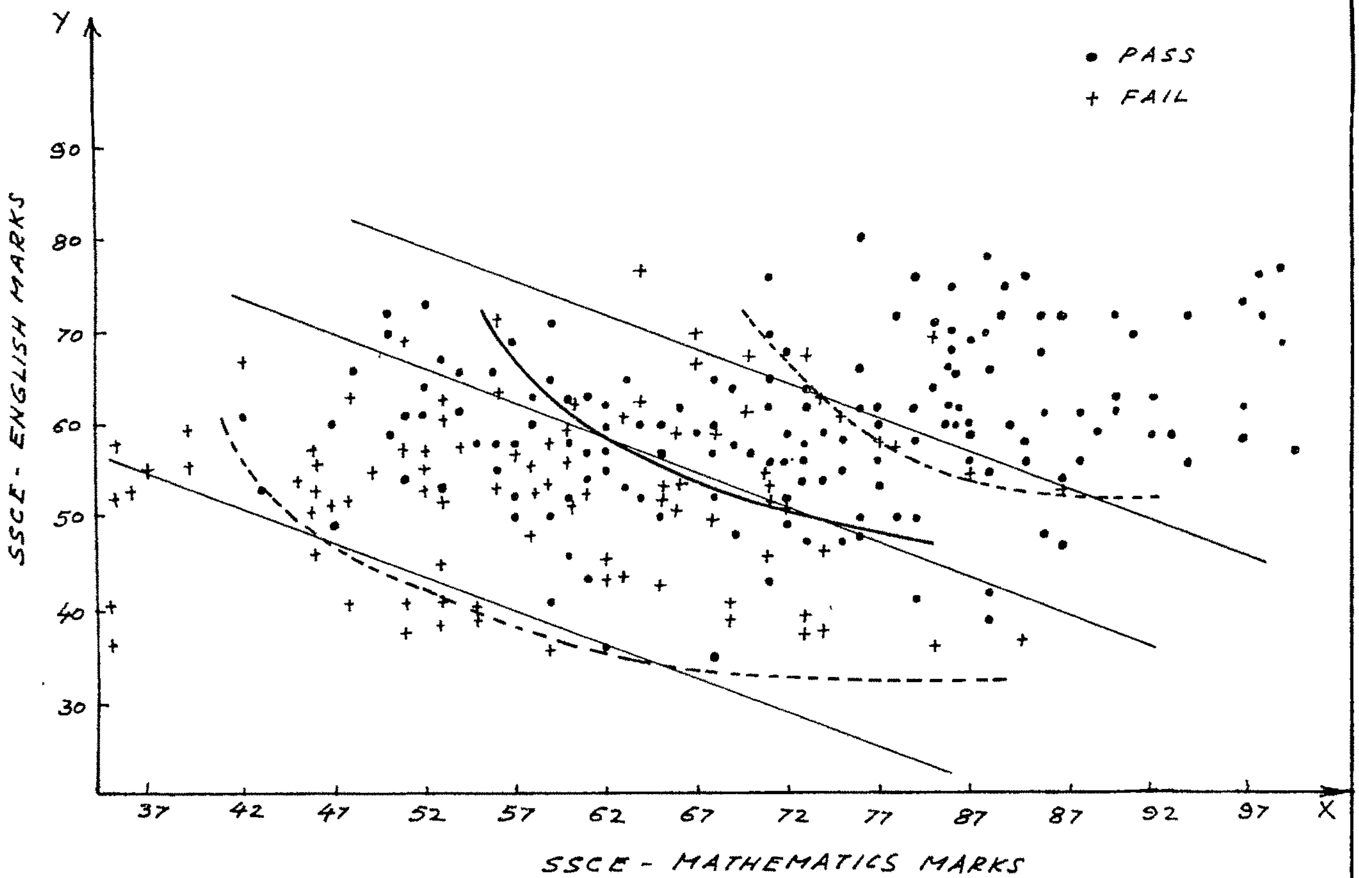
The Quadratic Discriminant, without the product term so calculated is-

$$= -0.00151108X_1^2 + 0.00126159X_2^2 + 0.38082868X_1 - 0.00143284X_2$$
$$-22.05563039$$

Results and Discussion-

In using this discriminant to classify 278 students and the same test criterion as before, 23 errors of type I were observed, while 47 errors of type II were observed. comparing these results with those found with two variable discriminant, we observe that the results are as much the same; yet in some specific situations, as shown in the following graph (Figure 4), the observed curvilinear trends seem to be more appropriate for prediction purposes.

The graph shows the points corresponding to passes and fails (by dots and cross) and the two discriminant.




GRAPH SHOWING THE OBSERVED SCATTERED POINTS (PASSES AND FAILS) AND DISCRIMINANT LINES AND CURVES

FIG. 4

The passes and fails are scattered so close to and mixed up on both sides of the discriminant line or curve that further improvement does not seem feasible with the help of either; but as we look on moving away from the dis-criminant on both sides, the curvilinear trends joining the dots as well as crosses, as shown by dotted lines, seem to provide better solutions than these given by the straight lines. For instance, in the lower extreme position, the suitability of the curve shown by the dotted line becomes more evident for a small probability (say 0.05 or 0.01) of II kind of error, as also, in the upper extreme portion for I kind of error to be equal to 0.01 or one percent. Thus the method of selection based on lower curvilinear solution ensures elimination of unfits in a better way, and can be used in situations where the number of seats for admissions are sufficiently large, and yet we want to reduce the unnecessary wastage keeping Type II error to be minimum or zero.

Usually the admission in college or university are governed by the number of seats. For institutions which maintain a very high standard, the number of seats are very limited and only a few top students are admitted. In such situations the upper curvilinear solution resulting in 0.01 error of II kind, seem to be more desirable compared to the straight line solution.

How these better solutions at the extreme situations can be explained?

It may be pointed out that the differences in variance- covariance matrices, as a result of which the Quadratic Discriminant arises, are mainly contributed by the extreme groups, while in the middle situation, these differences are not so effective.

So far we have considered the solution to the prediction in term of errors of first kind ( $\alpha_1$ ) and second kind ( $\alpha_2$ ) only and we have corresponding two regions only- one of acceptance ($R_1$) and the other region of rejection ($R_2$) and we have observed that as we change the discriminant line or curve, from one position to another, if one kind of error decreases the other kind of error increases and vice versa. If we use the optimum solution, then we have corresponding minimum errors of first and second kind that can be offerred by the data under the situation, depending upon the efficiency and reliability of the measuring instrument.

In yet another way, we may require to limit these errors of I and II kind to be smaller than some fixed values. In this case, the third region of doubtful cases (D), such that an individual belonging to this region remains unclassified, has to be introduced to keep the error of classification at a desired level. The frequency of doubtful cases decreases as the number of sufficiently accurate predictors increases. Contrary to this, a sequential test given by A. Wald and the extension of this test to multivariate analysis developed by C.R. Rao can be used to keep the wrong classification at a desired level, without making use of a large number of predictors simultaneously. We have also observed in the analysis of chapter VI, that it is not always profitable to start with a number of variables together unless each variable is sufficiently accurate to reduce percentage error in prediction. As such, the above method (due to A. Wald and C.R. Rao) can be used to advantage.

The regions $R_1$, $R_2$ and D are given as :

If $\frac{pf_1}{pf_2} \geq \frac{1-\alpha_2}{\alpha_1}$; the individual is classified as belonging to $R_1$

$\frac{pf_1}{pf_2} \leq \frac{\alpha_2}{1-\alpha_1}$, the individual is classified as belonging to $R_2$

and $\frac{\alpha_2}{1-\alpha_1} < \frac{pf_2}{pf_2} < \frac{1-\alpha_2}{\alpha_1}$; the individual is classified as belonging to D

where $\frac{pf_1}{pf_2}$ denote probability densities.

If the individual is found to belong to D, we take one or more additional test variables to classify him as belonging to $R_1$ or $R_2$. We will illustrate this method in a general case in the next chapter.

In chapter IV, we observed that, in case of aptitude measures, the quadratic relationship improves the prediction significantly over the linear combination of test variables and so, even if the results of the Quadratic Discriminant in this analysis with achievement measures are not found to be very gainful there is wide scope for Quadratic Discriminant of being useful, while dealing with aptitude measures.

Secondly, Linear Discriminant is a particular case of more general Quadratic Discriminant and if, appropriately used, Quadratic Discriminant does not harm the prediction, and is more suitable at the extremes, as observed above. Moreover, Quadratic Discriminant is based on probability density and hence can be easily geared to other more advanced methods due to A. Wald, C.R. Rao and Welch. As such, it might perhaps yield improved results with more rigorous methods and with better situations of test data and examinations.