IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 12, Issue 3 Ver. V (May. - Jun. 2016), PP 38-47
www.iosrjournals.org

On The Application of Linear Discriminant Functions for Equi and Auto Correlated Time Dependent Data

HashimuBulus
PhDDepartment of InsuranceFaculty of Management SciencesUniversity of Jos, Plateau State, Nigeria

Abstract: *This paper considers two special cases of correlation matrices, when k repeated observation on each of the p – variatewere collected from entity j at time t(t = 1, 2, …, k; p = 1,2,3; j = 1, 2, …, n_i; i = 1,2)have equal correlation and auto correlation. Case 1 assumed constant equi-correlation matrix for all variables of both populations with a common variance covariance matrix while case 2 assumed auto correlation matrix. Two discriminant procedures (Regression Discriminant and elongated data discriminant) were employed in constructing the sampled based discriminant function.Systolic and diastolic blood pressures and heart rate were collected sequentially in time from two sampled population* $\pi_1(\text{survivors})\text{and}\,\pi_2(\text{nonsurvivors})$ *of hypertensive patients admitted at Jos University Teaching Hospital.Three techniques: re-substitution, leave-one out and partition of samples were used toconstruct and evaluate the sample based discriminant function. Probabilities of misclassification obtained from the computed confusion matrices of each procedure under the two special cases of correlations (equi and auto correlations) were used to compare the performance of these functions. From the analyses, the two discriminartfunctions compare favourably with each other and the Fisher's commonly used rule; though the Elongated Data Discriminant function outperform the Regression Discriminant function having a lower probability of misclassification (0.308 when Pen Rose correlation is assumed and 0.4083 when auto-correlation is assumed). For both cases (equi and auto correlations) the re-substitution technique performs betterthan the other two techniques(leave – one out and partition of sample techniques)with low probability of misclassification.*

Keywords: *Disciriminant Function, Elongated Data Discriminant, Regression Discriminant, Equicorrelation, Autocoreelation*

I. Introduction

Suppose p- vector observation is collected k time on entity j of population π_i , then the n_i x pkdata matrix denoted Y_i is:

$$Y_i = \begin{bmatrix} \underset{\sim}{y}_{i11} & \underset{\sim}{y}_{i12} & \cdots & \underset{\sim}{y}_{i1t} & \cdots & \underset{\sim}{y}_{i1k} \\ \underset{\sim}{y}_{i21} & \underset{\sim}{y}_{i22} & \cdots & \underset{\sim}{y}_{i2t} & \cdots & \underset{\sim}{y}_{i2k} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots\vdots\vdots & \vdots \\ \underset{\sim}{y}_{ij1} & \underset{\sim}{y}_{ij2} & \cdots & \underset{\sim}{y}_{ijt} & \cdots & \underset{\sim}{y}_{ijk} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots\vdots\vdots & \vdots \\ \underset{\sim}{y}_{in_i1} & \underset{\sim}{y}_{i_in_i2} & \cdots & \underset{\sim}{y}_{in_it} & \cdots & \underset{\sim}{y}_{i1n_ik} \end{bmatrix} \quad \ldots 1$$

Denote the j[th] row which is the focus of the research by $Y_{ij} = \begin{bmatrix} \underset{\sim}{y}_{ij1} & \underset{\sim}{y}_{ij2} & \cdots & \underset{\sim}{y}_{ijt} & \cdots & \underset{\sim}{y}_{ijk} \end{bmatrix}$ whose elements are the p-vector of observation obtained at time t, on the j[th] entity in population π_i ; and the i[th]column by $Y_{it} = \begin{bmatrix} \underset{\sim}{y}_{i1t} & \underset{\sim}{y}_{i2t} & \cdots & \underset{\sim}{y}_{ijt} & \cdots & \underset{\sim}{y}_{in_ik} \end{bmatrix}$ whose elements are the p-vector of observation obtained at time t collected from all sampled entities in population π_i i = 1,2; t = 1,2,…,k.

Let $\pi_i \sim N\left(\underset{\sim}{\mu_i}, \Sigma\right)$ be two predetermined multivariate normal populations with mean μ_i and a common variance covariance matrix $\sum$. For known μ_i and $\sum$ the discriminant function is

$$u = \left[\underset{\sim}{y_0} - \tfrac{1}{2}\left(\underset{\sim}{\mu_1} + \underset{\sim}{\mu_2}\right)\right]\Sigma\left(\underset{\sim}{\mu_1} - \underset{\sim}{\mu_2}\right) \qquad \ldots 2$$

The optimal classification rule (R_{opt}) based on the discriminant function is:

$$(R_{opt}): \text{Classify } \underset{\sim}{y_0} \text{ into } \begin{cases} \pi_1 \text{ if } u \geq 0 \\ \pi_2 \text{ otherwise} \end{cases} \qquad \ldots 3$$

McLenchlan (1992)stated the sample analogue of u as

$$u = \left[\underset{\sim}{y_0} - \tfrac{1}{2}\left(\underset{\sim}{\bar{y}_1} + \underset{\sim}{\bar{y}_2}\right)\right]\Sigma\left(\underset{\sim}{\bar{y}_1} - \underset{\sim}{\bar{y}_2}\right) \qquad \ldots 4$$

Lanchenbruch (1975), Giri (1977) and McLenchlan (1992) obtained the error rates that will occur if u is used to classify a given entity into one of π_i i = 1,2

Application to real life

In some practical situations repeated measurements are made on the entity to be classified. For instance in medical practice, a patient may be recalled for further repetition of the same clinical variables with the intention of providing a firmer basis for diagnosis or prognosis of his or her condition. These repeated characteristics could be collected at an equally or unequally spaced fixed point in time. These repeated characteristic have to be correlated. Two special cases of correlation were considered where in the multivariate normal group – conditional densities, the means are equal and the uniform variance covariance structure is
The classification rules were evaluated and compared with one another under the following two special cases of correlation matrices:

i) the Penrose structure (equicorrelation): $\hat{R} = \begin{bmatrix} 1 & r & r & \cdots & r & \cdots & r \\ r & 1 & r & \cdots & r & \cdots & r \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r & r & r & \cdots & 1 & \cdots & r \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r & r & r & \cdots & r & \cdots & 1 \end{bmatrix}$

ii)Near neighbour correlation: $\hat{R} = \begin{bmatrix} 1 & r & 0 & \cdots 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ r & 1 & r & \cdots 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots r & 1 & r & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots 0 & \vdots & 0 & 0 & 1 & r \\ 0 & 0 & 0 & \cdots 0 & \cdots & 0 & 0 & r & 1 \end{bmatrix}$

II. Review Of Literature

The study by Choi (1972) and subsequent ones by Gupta (1980) considered when the k repeated characteristics of each P-dimensional continuous random variable are equicorrelated. The characteristic vector $\underset{\sim}{y}$ is considered as a ρk dimensional random variable. Basu and Odell (1974) distinguished between simple equicorrelated and equicorrelated training data because they felt that the error rates of the sample linear discriminant function were unaffected by simple equicorrelation. However, Mclachlan (1976) showed that the error rates do change in the presence of simple equicorrelation; indeed the error rates asymptotically increase in size for positive correlation.

The repeated measurements y_{ijk} are said to be equicorrelated if

$cov(y_{ijt}y_{ijl}) = \Gamma_i i = 1, 2 ; j = 1, 2, \dots n_i, t, l = 1, 2, \dots, K$

where, Γ_i is a symmetric matrix.

The same observations are said to be simply equicorrelated if $\Gamma_i = \rho_i\Sigma_i$,

where $0 < \rho < 1$ and Σ_i is the common variance covariance matrix of the observation from population $\pi_i i = 1, 2$,

Gupta (1986) has obtained expression for the error rates of the Bayes rule in the case of a single feature variable (p=1). Gupta and Logan (1990) studied the case of multiple variables (p>1).

Weber and Baldessari (1988), studied the effect of correlated training data on the performance of sample normal linear discriminant rule. Earlier, Basu and Odell (1974) studied the case when the usual assumption of independent training data is violated. When evaluating the performance of the sample linear discriminant rule, Basu and Odell (1974) observed that the actual error rates were higher than the theoretically anticipated values under the assumption of independence. Basu and Odell (1974) investigated the effect of equicorrelated training data that classified with respect to two populations π_1 and π_2. Basu and Odell (1974) assumed the same variance covariance structure within each group, where $\rho_1 = \rho_2 = \rho$ and $-(k-1)^{-1} < \rho < 1$.

McLachlan (1992) stated that Tubbs (1980) investigated the effect of serially corrected training data on the unconditional error rates of the sample normal linear discriminant rule under homoscedastic normal model $[y \sim N(\mu_i, \Sigma)]$. He considered when n_i training data y_{ij} (i = 1, 2 ; j =1, 2, …. n_i) from population π_i follow a simple multivariate stationary autoregressive model of order one AR(1) where

$$cov(y_{ij}, y_{ik}) = \rho_d\Sigma \quad t \neq l = 1, 2, \dots, k$$

$$\text{for } d = |t-l| \text{ and } 0 < \rho_d < 1.$$

McLachlan (1992) added that Lawoko and McLachlan (1983, 1985, 1986, 1988 and 1989) considered the effect of correlated training data on the simple normal linear discriminant rule under the homoscedestic normal model. Lawoko and Mclachlan (1983) derived the asymptotic expansion of the group – specific unconditional error rates of the sample normal linear discriminant rule based on some model considered by Tubbs (1980). Lawoko and McLachlan (1983) demonstrated the magnitude of the increase in the unconditional error rates for positively correlated training data by evaluating the asymptotic expansion for a univariate stationary autoregressive process of order one, for which $\rho_d = \rho^d$ where $0<\rho<1$.

Lawoko and McLachlan (1988) according to McLachlan (1992) showed that the optimism of the plug-in method of estimating the error rates is magnified by positively correlated training data following a stationary AR(1) model. Early, Lawoko and Mclachlan (1986) showed asymptotically that the Ƶ – statistic defined as $n_1 / (n_1 + n_2)$ is preferable to the simple normal linear discriminant function for positively correlated univariate training data following a stationary AR(1) process.Lawoko and McLachlan (1989) studied the effect of correlated training data on the estimates of the mixing proportions obtained by an application of the sample normal linear discriminant rule, as in area estimation via remote sensing.

III. The Model

This section constructs and evaluates sample based classification rules for two multivariate populations. Here data on the three clinical variables (p = 3); systolic blood pressure (Y_1), diastolic blood pressure (Y_2) and heart rate (Y_3) obtain at k (k = 12) points in time (6 am and 6 pm) from n sampled ($n = n_1 + n_2 = 120$) hypertensive patients of known survival status admitted at a Hospital were used to construct the sample based classification rules. For each of the two procedures used in constructing the sample based classification rules, confusion matrices and probabilities of misclassification were obtained. Two correlation structures (Pen Rose and near neighbour correlation matrices) were assumed.

3.1 Sample Based Descriminant Functions For Equi-Correlated Data (Pen Rose Correlation Structure)

This section considers first special case of correlation matrix, when the k repeated observations on each of the p- variate have equal correlation. It assumes constant equi-correlation matrix for all variables of both populations, with a common variance-covariance matrix. Confusion matrices and probabilities of misclassification were obtained for each of the two procedures (regression discriminant and elonogated data discriminant procedures) used in constructing the sample-based classification rules,.

1. **Regression Discriminant Procedure:** The sample based classification rules, confusion matrices and probabilities of misclassification were obtained here using the re-substitution technique, Jackknife methods of leave-one out and partitioning of samples.

i) Re-substitution technique: Here all sampled objects were use, in estimating the following the parameters.

$$\hat{R}^{-1} = \begin{bmatrix} 5.509177 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 \\ -0.672426 & 5.509177 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 \\ -0.672426 & -0.672426 & 5.509177 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 \\ -0.672426 & -0.672426 & -0.672426 & 5.509177 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 \\ -0.672426 & -0.672426 & -0.672426 & -0.672426 & 5.509177 & -0.672426 & -0.672426 & -0.672426 & -0.672426 \\ -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & 5.509177 & -0.672426 & -0.672426 & -0.672426 \\ -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & 5.509177 & -0.672426 & -0.672426 \\ -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & 5.509177 & -0.672426 \\ -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & -0.672426 & 5.509177 \end{bmatrix}$$

The regression coefficients are:

$$\overline{\hat{B}}_1 = \begin{bmatrix} 108.3508 & 79.39059 & 83.89045 \\ 0.2280981 & 0.128107 & 0.02061279 \end{bmatrix} \qquad \overline{\hat{B}}_2 = \begin{bmatrix} 128.1734 & 85.21008 & 79.89485 \\ 0.1339035 & 0.1093184 & 0.09419183 \end{bmatrix}$$

$$\underset{\sim}{\gamma_1} = [108.3508 \quad 0.2280981 \quad 79.39059 \quad 0.128107 \quad 83.89045 \quad 0.02061279]'$$

$$\underset{\sim}{\gamma_2} = [128.1734 \quad 0.1339035 \quad 85.21008 \quad 0.1093184 \quad 79.89485 \quad 0.09419183]'$$

$$S = \begin{bmatrix} 1325.272000 & 654.206200 & 13.026930 \\ 654.206200 & 543.150400 & 12.395630 \\ 13.026930 & 12.395630 & 302.543900 \end{bmatrix} \qquad S^{-1} = \begin{bmatrix} 0.001861 & -0.002242 & 0.000012 \\ -0.002242 & 0.004544 & -0.000090 \\ 0.000012 & -0.000090 & 0.003308 \end{bmatrix}$$

Hence the discriminant function is:

$$\hat{T} = \begin{bmatrix} \hat{\beta}_{011} - 118.2621 \\ \hat{\beta}_{021} - 0.1810008 \\ \hat{\beta}_{012} - 82.30033 \\ \hat{\beta}_{022} - 0.1187127 \\ \hat{\beta}_{013} - 81.89265 \\ \hat{\beta}_{023} - 0.05740231 \end{bmatrix}' H' R^{-1} H \otimes \begin{bmatrix} 0.001861 & -0.002242 & 0.000012 \\ -0.002242 & 0.004544 & -0.000090 \\ 0.000012 & -0.000090 & 0.003308 \end{bmatrix}^{-1} \begin{bmatrix} -19.82267 \\ 0.09419461 \\ -5.819489 \\ 0.01878863 \\ 3.995598 \\ -0.07357904 \end{bmatrix}$$

This was then use in classifying a new object o(γ_o) into one the two predetermined population π_1 and π_2 as:

$$\hat{R}^{(p)}_{rd}: \text{classify } o(\gamma_o) \text{ into } \begin{cases} \pi_1 & \text{if } \hat{T} \geq 0 \\ \pi_2 & \text{otherwise} \end{cases} \qquad \ldots 5$$

ii) When the leave-one out technique is use, the probability of misclassification obtained using its confusion matrix was:

$$\text{PMC} = \frac{65}{120} = 0.5417$$

(iii) When half of the training sample ($n_1 = n_2 = 25$) was used to computed the Drscriminant function and the other half to validate, we get estimate of the following parameters as:

$$R^{-1}=\begin{bmatrix} 5.028993 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 \\ -0.612137 & 5.028993 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 \\ -0.612137 & -0.612137 & 5.028993 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 \\ -0.612137 & -0.612137 & -0.612137 & 5.028993 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 \\ -0.612137 & -0.612137 & -0.612137 & -0.612137 & 5.028993 & -0.612137 & -0.612137 & -0.612137 & -0.612137 \\ -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & 5.028993 & -0.612137 & -0.612137 & -0.612137 \\ -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & 5.028993 & -0.612137 & -0.612137 \\ -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & 5.028993 & -0.612137 \\ -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & -0.612137 & 5.028993 \end{bmatrix}$$

The computed regression coefficients are:

$$\hat{B}_1 = \begin{bmatrix} 116.6719 & 82.02781 & 81.6575 \\ 0.1551614 & 0.07939636 & 0.04227237 \end{bmatrix} \qquad \hat{B}_2 = \begin{bmatrix} 124.837 & 86.49452 & 74.21841 \\ 0.1736604 & 0.08802477 & 0.1450099 \end{bmatrix}$$

So that,

$$\gamma_1 = [\ 116.6719 \quad 0.1551614 \quad 82.02781 \quad 0.07939636 \quad 81.6575 \quad 0.04227237\ ]'$$

$$\gamma_2 = [\ 124.837 \quad 0.1736604 \quad 86.49452 \quad 0.08802477 \quad 74.21841 \quad 0.1450099\ ]'$$

Therefore the discriminant function is

$$\hat{T} = \begin{bmatrix} \hat{\beta}_{o11} - 120.7544 \\ \hat{\beta}_{o21} - 0.1644109 \\ \hat{\beta}_{o12} - 84.26117 \\ \hat{\beta}_{o22} - 0.08371057 \\ \hat{\beta}_{o13} - 77.93796 \\ \hat{\beta}_{o23} - 0.09364112 \end{bmatrix} H' R^{-1} H \otimes S^{-1} \begin{bmatrix} -8.1651 \\ -0,01849902 \\ -4.466713 \\ -0.08628406 \\ 7.439087 \\ -0.1027375 \end{bmatrix}$$

The discriminating rule is:

$$\hat{R}^{(p)}_{rd} : classify \quad o(\gamma_o) \text{ int } o \begin{cases} \pi_1 \ if \ \hat{T} \geq 0 \\ \pi_2 \ otherwise \end{cases} \qquad \ldots 6$$

Again the probability of misclassification from the confusion matrix was:

$$\text{PMC} = \frac{29}{60} = 0.4833$$

\When Pen Rose correlation was assumed the confusion matrices computed for each technique using Regression Discriminant Function was summarized in table 1 while the probability of misclassification obtained from these confusion matrices was presented in table 3

Table 1: Confusion Matrices forThree Classification Techniques Using Regression Discriminant Function

Allocate to Population	Classification techniques								
	Re-substitution			Leave one out			Partition of data		
	Actual Population		Total	Actual Population		Total	Actual Population		Total
	π_1	π_2		π_1	π_2		π_1	π_2	
π_1	32	27	59	28	33	61	16	15	31
π_2	28	33	61	32	27	59	14	15	29
Total	60	60	120	60	60	120	60	60	120

2. Elongated (**Conbined) data Discriminant Procedure:** The discriminant function obtained here used the elongated data discriminant without summary.

i) Re-substitution technique: Using all the sampled data to construct and evaluate the classification function, we have:

$\hat{X}$ = -0.012257y_{o1} + 0.067051y_{o2} - 0.018176 y_{o3} - 0.038290 y_{o4} + 0.027579 y_{o5} - 0.043416 y_{o6} + 0.033168 y_{o7} + 0.012814 y_{o8} - 0.010160 y_{o9} - 0.026007 y_{o10} + 0.093093 y_{o11} - 0.133003 y_{o12} 0.027416 y_{o13} + 0.047536 y_{o14} -0.081336 y_{o15} + 0.086531 y_{o16} - 0.077718 y_{o17} + 0.004061 y_{o18} + 0.000945 y_{o19} + 0.037812 y_{o20} + 0.045595 y_{o21} + 0.045675y_{o22} - 0.014522 y_{o23} -0.068450 y_{o24} + 0.081725 y_{o25} - 0.070112 y_{o26} - 0.005356 y_{o27} + 0.084241 y_{o28} + 0.026188 y_{o29} - 0.136745 y_{o30}+ 1.784809

Hence the classification rule is:

$$\hat{R}_{sd}^{(p)} : \text{classify o}\left(\underset{\sim}{\gamma'_o}\right) \text{into} \begin{cases} \pi_1 \text{ if } \hat{X} \geq 0 \\ \pi_2 \text{ otherwise} \end{cases} \quad \ldots 7$$

where

$$\underset{\sim}{\gamma_0} = \begin{bmatrix} \underset{\sim}{y'_{o1}} & \underset{\sim}{y'_{o2}} & \underset{\sim}{y'_{o3}} \end{bmatrix}'$$

$$= \begin{bmatrix} y_{o11} & y_{o12} & \cdots & y_{o110} & y_{o21} & y_{o22} & \cdots & y_{o210} & y_{o31} & y_{o32} & \cdots & y_{o310} \end{bmatrix}'$$

is a column vector consisting of the observations on the three clinical variables (p = 3) obtain at ten points (k = 12) in time from the object to be classified.

(ii) Technique of leave-one out:

(iii) When half of the training sample (n_1 = n_2 =30) to compute the classification rule and the remaining to evaluate, we obtained the discriminant function $\hat{\Lambda}$ as:

$\hat{\Lambda}$ = 0.070154y_{o1} + 0.060172 y_{o2} - 0.025645 y_{o3} - 0.029440 y_{o4} + 0.003717 y_{o5} - 0.073324 $_{o6}$ + 0.079218 y_{o7} - 0.010908 y_{o8} + 0.038892 y_{o9} - 0.124846 y_{o10} + 0.086198 y_{o11} - 0.022818 y_{o12} + 0.019232 y_{o13} - 0.024001 y_{o14} -0.070115 y_{o15} + 0.098787 y_{o16} - 0.247972 y_{o17} + 0.051866 y_{o18} - 0.104802 y_{o19} + 0.216457 y_{o20} + 0.053348 y_{o21} + 0.098384 y_{o22} + 0.063102 y_{o23} -0.012679 y_{o24} - 0.092118 y_{o25} - 0.115671 Yy_{o26} + 0.075021 y_{o27} + 0.159265 y_{o28} - 0.051068 y_{o29} - 0.181996 y_{o30} + 0.07860756.

Therefore the sample based classification rule, confusion matrix and the corresponding probability of misclassification are respectively obtain as:

$$\hat{R}_{sd}^{(p)} : \text{classify o}\left(\underset{\sim}{\gamma_o}\right) \text{into} \begin{cases} \pi_1 \text{ if } \hat{\Lambda} \geq 0 \\ \pi_2 \text{ otherwise} \end{cases} \quad \ldots 8$$

Similarly, the confusion matrices computed for each technique under this procedure assuming the Pen Rose type of correlation was presented in table 2and the probability of misclassification obtained from the confusion matrices was summarized in table 3

Table 2: Confusion Matrices for Three Classification Techniques Using Elongated Data Discriminant Function

Allocate to Population	Classification techniques								
	Re-substitution			Leave one out			Partition of data		
	Actual population		Total	Actual population		Total	Actual population		Total
	π_1	π_2		π_1	π_2		π_1	π_2	
π_1	43	20	63	34	33	67	15	13	28
π_2	17	40	57	26	27	53	15	17	32
Total	60	60	120	60	60	120	30	30	60

Comparison of Procedures by Techniques for Equi-correlated Data

From the confusion matrix, the empirical estimate of the corresponding probability of misclassification refer to as error rates for the two procedures were presented in table 1

Table 3: Probabilities Of Misclassification Assuming Pen Rose Correlatiom Structure (Equicorrelation)

Classification Procedures	Evaluation techniques		
	Re-substitution	Leave- one out	Partition of samples
Regression discriminant	0.4583	0.5417	0.4833
Combined data disciminant	0.308	0.4917	0.4667

From table1 it can be observed that when Pen Rose correlation structure (equicorrelation) is assumed, the elongated data discriminant procedure performs better for all the three techniques. The re-substitution technique for instance indicates that the probability of misclassification of the combined data discriminant procedure is 0.308.This is better than the regression discriminant procedure with probability of misclassification 0.4583.

3.2 Sample Based Classification Rules Assuming Near Neighbour Correlation

This section considers the second special case of correlation matrix (near neighbour correlation structure) given in section 1 as:

$$|\rho_{t(t+h)}| = \begin{cases} 1 & for\ \ h = 0 \\ |r| < 1 & for\ \ h = \pm 1 \\ 0 & elswhere \end{cases} \qquad t = 1,2,\ldots,k$$

Here the near neighbor correlation matrix is assumed to be constant for all variables of both populations.
Base on this structure of the correlation matrix, sample based discriminant function for two multivariate normal populations are constructed and evaluated using the following discriminant procedures.

1. **Regression Discriminant Procedure:** Sample based discriminant function was obtained when using all sampled entities, and the same entities were employed to construct the confusion matrices that producedthe probabilities of misclassification.

$$R^{-1} = \begin{bmatrix} 9.510392 & -10.152820 & 2.576874 & 7.117750 & -10.921680 & 5.783354 & 4.076608 & -10.762080 & 8.918080 \\ -10.152820 & 12.112210 & -3.074186 & -8.491408 & 13.029460 & -6.899486 & -4.863355 & 12.839050 & -10.639180 \\ 2.576874 & -3.074186 & 1.080144 & 2.983534 & -4.578022 & 2.424198 & 1.708784 & -4.511119 & 3.738175 \\ 7.117750 & -8.491407 & 2.983535 & 4.972946 & -7.630633 & 4.040645 & 2.848196 & -7.519120 & 6.230779 \\ -10.921680 & 13.029460 & -4.578022 & -7.630633 & 13.511490 & -7.154733 & -5.043274 & 13.314040 & -11.032780 \\ 5.783353 & -6.899485 & 2.424198 & 4.040645 & -7.154732 & 4.418312 & 3.114408 & -8.221910 & 6.813151 \\ 4.076608 & -4.863354 & 1.708785 & 2.848197 & -5.043275 & 3.114409 & 1.350487 & -3.565230 & 2.954357 \\ -10.762080 & 12.839050 & -4.511120 & -7.519120 & 13.314040 & -8.221910 & -3.565230 & 12.603620 & -10.444090 \\ 8.918079 & -10.639180 & 3.738176 & 6.230779 & -11.032780 & 6.813151 & 2.954356 & -10.444090 & 9.654582 \end{bmatrix}$$

The regression coefficients are

$$\bar{\hat{B}}_1 = \begin{bmatrix} 154.1501 & 58.8903 & 88.53336 \\ -0.1026616 & 0.2977844 & -0.00596293 \end{bmatrix}$$

$$\bar{\hat{B}}_2 = \begin{bmatrix} 106.7337 & 123.839 & -84.8001 \\ 0.1286151 & -0.2381343 & 2.03119 \end{bmatrix}$$

Now let

$$\gamma_1 = \begin{bmatrix} 154.1501 & -0.1026616 & 58.8903 & 0.2977844 & 88.53336 & -0.00596293 \end{bmatrix}'$$

$$\gamma_2 = \begin{bmatrix} 106.7337 & 0.1286151 & 123.839 & -0.2381343 & -84.8001 & 2.03119 \end{bmatrix}'$$

$$S = \begin{bmatrix} 1325.272000 & 654.206200 & 13.026930 \\ 654.206200 & 543.150400 & 12.395630 \\ 13.026930 & 12.395630 & 302.543900 \end{bmatrix} \qquad S^{-1} = \begin{bmatrix} 0.001861 & -0.002242 & 0.000012 \\ -0.002242 & 0.004544 & -0.000090 \\ 0.000012 & -0.000090 & 0.003308 \end{bmatrix}$$

The regression discriminant function is :

$$\hat{\Theta} = \begin{bmatrix} \hat{\beta}_{o11} - 130.4419 \\ \hat{\beta}_{o21} - 0.01297672 \\ \hat{\beta}_{o12} - 91.36467 \\ \hat{\beta}_{o22} - 0.02982505 \\ \hat{\beta}_{o13} - 1.866631 \\ \hat{\beta}_{o23} - 1.012613 \end{bmatrix}' \; H R^{-1} H \otimes S^{-1} \begin{bmatrix} 47.41635 \\ -0.2312767 \\ -64.94874 \\ 0.5359187 \\ 173.3335 \\ -2.037153 \end{bmatrix}$$

Hence the discriminating rule is

$$\hat{R}^{(p)}_{rd} : classify \;\; o(\underset{\sim}{\gamma_o}) \text{ int } o \begin{cases} \pi_1 \; if \; \hat{\Theta} \geq 0 \\ \pi_2 \; otherwise \end{cases} \qquad \ldots 9$$

where

$$\underset{\sim}{\gamma_o} = \begin{bmatrix} \underset{\sim}{\hat{\beta}'_{o1}} & \underset{\sim}{\hat{\beta}'_{o2}} & \underset{\sim}{\hat{\beta}'_{o3}} \end{bmatrix}'$$

$$= \begin{bmatrix} \hat{\beta}_{o11} & \hat{\beta}_{o21} & \hat{\beta}_{o112} & \hat{\beta}_{o22} & \hat{\beta}_{o113} & \hat{\beta}_{o23} \end{bmatrix}'$$

is a pk = 6 _ dimensional column vector consisting of the regression coefficients of the object to be classified.

ii) The leave-one out technique:

iii) Using half of the training sample ($n_1 = n_2 = 30$) we computed the discriminantfunction and the other half to validate, we got:

$$R^{-1} \begin{bmatrix} 3.824769 & -3.433407 & 0.353197 & 2.971102 & -3.930041 & 1.742008 & 1.910653 & -4.117533 & 3.252652 \\ -3.433407 & 4.173184 & -0.429298 & -3.611269 & 4.776826 & -2.117349 & -2.322331 & 5.004716 & -3.953484 \\ 0.353197 & -0.429298 & 0.170913 & 1.437728 & -1.901762 & 0.842964 & 0.924573 & -1.992491 & 1.573971 \\ 2.971101 & -3.611269 & 1.437728 & 1.823146 & -2.411576 & 1.068941 & 1.172427 & -2.526626 & 1.995912 \\ -3.930040 & 4.776825 & -1.901762 & -2.411576 & 4.784369 & -2.120693 & -2.325999 & 5.012619 & -3.959728 \\ 1.742007 & -2.117348 & 0.842964 & 1.068942 & -2.120693 & 1.466998 & 1.609020 & -3.467501 & 2.739158 \\ 1.910653 & -2.322331 & 0.924573 & 1.172427 & -2.325999 & 1.609020 & 0.424500 & -0.914815 & 0.722660 \\ -4.117533 & 5.004716 & -1.992491 & -2.526626 & 5.012620 & -3.467501 & -0.914815 & 4.631205 & -3.658428 \\ 3.252652 & -3.953484 & 1.573971 & 1.995913 & -3.959728 & 2.739158 & 0.722660 & -3.658428 & 3.889982 \end{bmatrix}$$

The regression coefficients are:

$$\hat{B}_1 = \begin{bmatrix} 690.7573 & 152.3043 & 806.7698 \\ -3.539114 & -0.456894 & -7.167589 \end{bmatrix} \qquad \hat{B}_2 = \begin{bmatrix} 146.2521 & 81.41943 & 95.45372 \\ 0.09190734 & 0.2690378 & -0.05815665 \end{bmatrix}$$

Again, let

$$\underset{\sim}{\gamma_1} = [\, 690.7573 \quad -3.539114 \quad 152.3043 \quad -0.456894 \quad 806.7698 \quad -7.167589 \,]'$$

$$\underset{\sim}{\gamma_2} = [\, 146.2521 \quad 0.09190734 \quad 81.41943 \quad 0.2690378 \quad 95.45372 \quad -0.05815665 \,]'$$

The regression, discriminant function computed here is:

$$\hat{\varpi} = \begin{bmatrix} \hat{\beta}_{o11} - 418.5047 \\ \hat{\beta}_{o21} + 1.723603 \\ \hat{\beta}_{o12} - 116.8619 \\ \hat{\beta}_{o22} + 0.0939281 \\ \hat{\beta}_{o13} - 451.1118 \\ \hat{\beta}_{o23} + 3.612873 \end{bmatrix}' H' R^{-1} H \otimes S^{-1} \begin{bmatrix} 544.5051 \\ -3.631021 \\ 70.8849 \\ -0.7259318 \\ 711.316 \\ -7.109432 \end{bmatrix}$$

Therefore the discriminant rule is:

$$\hat{R}^{(p)}_{rd} : classify \ \ o(\underset{\sim}{\gamma_o}) \text{int } o \begin{cases} \pi_1 \ if \ \hat{\varpi} \ge 0 \\ \pi_2 \ otherwise \end{cases} \qquad \ldots 10$$

where

$$\underset{\sim}{\gamma_o} = \begin{bmatrix} \underset{\sim}{\hat{\beta}'_{o1}} & \underset{\sim}{\hat{\beta}'_{o2}} & \underset{\sim}{\hat{\beta}'_{o3}} \end{bmatrix}'$$

$$= \begin{bmatrix} \hat{\beta}_{o11} \ \hat{\beta}_{o21} & \hat{\beta}_{o112} \ \hat{\beta}_{o22} & \hat{\beta}_{o113} \ \hat{\beta}_{o23} \end{bmatrix}'$$

is a pk = 6 _ dimensional column vector consisting of the regression coefficients of the object to be classified.

For the Near Neighbour type of correlation, the confusion matrices computed using the three different techniques was presented in table 4 while the probability of misclassification obtained was summarized in table 5

Table 4: Confusion Matrices for Three Classification Techniques Using Regression Discriminant Function

Allocate to Population	Classification techniques								
	Re-substitution			Leave one out			Partition of data		
	Actual population		Total	Actual population		Total	Actual population		Total
	π_1	π_2		π_1	π_2		π_1	π_2	
π_1	48	46	94	11	12	23	1	3	4
π_2	12	14	26	49	48	97	29	28	57
Total	60	60	120	60	60	120	30	30	60

2). **Combined DataDiscriminant Procedure:** The classification rules obtained here use the elongated data discriminant procedure.

(i) Using all the sampled data to construct and evaluate the classification rule we have:

$\hat{T}$ = 0.062399y_{o1} + 0.062996 y_{o2} + 0.071829 y_{o3} + 0.084921 y_{o4} + 0.099286 y_{o5} + 0.108923y_{o6} + 0.116898 y_{o7} + 0.120664 y_{o8} + 0.132692 y_{o9} + 0.144199y_{o10} + 0.060459 y_{o11} + 0.054826 y_{o12} + 0.051885 y_{o13} + 0.047535 y_{o14} + 0.036833 y_{o15} + 0.037052 y_{o16} + 0.030294 y_{o17} + 0.029386 y_{o18} + 0.024724 y_{o19} + 0.023284y_{o20} + 0.023495 y_{o21} + 0.023463 y_{o22} + 0.023476 y_{o23} + 0.023508y_{o24} + 0.023497 y_{o25} + 0.023536 y_{o26} + 0.023554 y_{o27} + 0.023543 y_{o28} + 0.023563 y_{o29} + 0.023662 y_{o30} -1.009459

The classification rule for classifying a given hypertensive patient with vector of clinical data

$$\underset{\sim}{\gamma_o} = \begin{bmatrix} \underset{\sim}{y'_{o1}} & \underset{\sim}{y'_{o2}} & \underset{\sim}{y'_{o3}} \end{bmatrix}' \text{ is}$$

$$\hat{R}^{(p)}_{sd} : \text{classify } o\left(\underset{\sim}{\gamma_o}\right) \text{int } o \begin{cases} \pi_1 \text{ if } \hat{T} \ge 0 \\ \pi_2 \text{ otherwise} \end{cases} \qquad \ldots 11$$

ii) Jackknife technique of leave-one out:

iii) Using half of the training sample (n_1= n_2=30) to compute the classification function and remaining to evaluate, we obtained the following :

$\hat{\Lambda} =$ - 0.003022y_{o1} - 0.000049 y_{o2} - 0.001714 y_{o3} - 0.014467 y_{o4} + 0.002380 y_{o5} - 0.001001 y_{o6} - 0.023375 y_{o7} + 0.027053 y_{o8} - 0.024456 y_{o9} - 0.004978 y_{o10} - 0.000891 y_{o11} + 0.014365 y_{o12} - 0.016446 y_{o13} + 0.011393 y_{o14} + 0.002279 y_{o15} - 0.020153 y_{o16} + 0.037502 y_{o17} - 0.055022 y_{o18} + 0.040116 y_{o19} - 0.005597 y_{o20} + 0.013196 y_{o21} - 0.012703 y_{o22} + 0.009388 y_{o23} + 0.004931 y_{o24} - 0.021810 y_{o25} + 0.004592 y_{o26} - 0.003596 y_{o27} + 0.004851 y_{o28} + 0.014194 y_{o29} - 0.034960 y_{o30} + 7.792888

Hence a given hypertensive patient with vector of clinical data

$\underset{\sim}{\gamma_o} = \left[\underset{\sim}{Y'_{o1}} \quad \underset{\sim}{Y'_{o2}} \quad \underset{\sim}{Y'_{o3}} \right]'$ is classify as follows:

$$\hat{R}_{sd}^{(p)} : \text{classify } o\left(\underset{\sim}{\gamma_o}\right) \text{ into } \begin{cases} \pi_1 \text{ if } \hat{\Lambda} \geq 0 \\ \pi_2 \text{ otherwise} \end{cases} \qquad \ldots 12$$

Similarly, the confusion matrices computed for each technique under Elongated Discriminant procedure was presented in table 5 with the probability of misclassification obtained from the confusion matrices in table 6..

Table 5: Confusion Matrices for Three Classification TechniquesUsing Elongated Data Discriminant Function

Allocate to Population	Classification techniques								
	Re-substitution			Leave one out			Partition of data		
	Actual population		Total	Actual population		Total	Actual population		Total
	π_1	π_2		π_1	π_2		π_1	π_2	
π_1	36	25	61	33	25	58	17	15	32
π_2	24	35	59	27	35	62	13	15	28
Total	60	60	120	60	60	120	30	30	60

Comparison Procedures by Techniques for Near Neighbour Procedure

Again the confusion matrix produced the probability of misclassification (error rates) in table 6. These error rates were used to compare the performances of the two procedures and three Techniques used.

Table 6: Probabilities of Misclassification when Near Neighbour Correlation (autocorrelation) is Assumed

Classification Procedures	Evaluation techniques		
	Re-substitution	Leave- one out	Partition of samples
Regression discriminant	0.4833	0.5083	0.5167
Combined data disciminant	0.4083	0.4333	0.4667

From table 2, the three techniques indicate that the combined data discriminant procedure is better than the regression discriminant with lower probabilities of misclassification. The probabilities of misclassification show that the re-substitution technique is a better estimator of the apparent error rate.

IV. Conclusion

From the analyses it can be observed that whichever technique is employed to construct and evaluate the sample based discriminant function, the elongated discriminant function performs better than the regression disciminant function. In both procedures the Re-substitution technique was discovered to be the most appropriate technique of estimating the apparent error rate (APERA), as this gives a lower probability of misclassification.

References

[1]. Basu. P. A and Odell, P. L (1974). Effect of IntraclasssCorrelationAmongh Training Samples on Misclassification Probabilities of Bayes Procedure. Pattern Recognition. Vol. 6. Pg. 13-16.

[2]. Choi, S. C. (1972). Classification of Multiple Observed Data. Biometrika, Vol. 14 pg. 8- 11.

[3]. Gupta, K. A. (1980).On a Multivariate Classification Model.in Multivariate Statistical Analysis, R. A. Gupta (Ed) Amsterdam. North Holland. Pg 83 -93.

[4]. Gupta A. K. (1986). On Classification Rule for Multiple Measurements. Computer Mathematics Applicltion. Vol, 12a Pg.301 – 308.

[5]. Mclachlan, G. J. (1976). The Bias of Apperent Error in DiscrimirnantAnalyses. Biometrika. Vol. 63. Pg.239 – 244.

[6]. McLachlan, G. J. (1992). Discriminant Analysis and Statistical Pattern Recognition. John Wiley and Sons. New York.

[7]. Gupta, A. K. and Logan, T. P. (1990). On a Multiple Observation Model in Discriminant Analysis . Journal of Statistical computing Simulation. Vol.34. Pg 119 -132.