IJCRR
Vol 05 issue 17
Section:General Sciences
Category: Case Study
Received on: 01/06/12
Revised on: 27/06/13
Accepted on: 19/07/13

COMPARISON OF CLASSIFICATION RULES FOR TWO UNIVERIATE POPULATIONS

Hashimu Bulus

Department of Business Administration, University of Jos, Nigeria

E-mail of Corresponding Author: hash_bulus@yahoo.com

ABSTRACT

Three procedures for classifying an entity into one of the two predetermined univariate populations π_1 and π_2 were derived, evaluated and compared. This paper proposes Unspecified structure of the variance, Regression Discriminant (RD) and Elongated Discriminant (ED) procedures for the classification using k repeated observations collected on each entity j at time t (t = 1,2,…, k;j = 1,2,…,n_i; i = 1,2). Mean arterial pressure which is a function of systolic and diastolic blood pressures were collected sequentially in time from two sampled populations π_1(survivors) and π_2 (non-survivors), of hypertensive patients admitted at the Jos University Teaching Hospital (J.U.T.H).Three techniques: re-substitution, leave – one out and partitioning of samples are used to construct and evaluate the sample based classification rules. Probabilities of misclassification obtained from the confusion matrices produced by these techniques are used to compare the performances of these rules. The analysis reveals that the procedures compare favourably with one another and the Fisher's commonly used rule. The classification rule obtained using Elongated Discriminant procedure performs better with lower error rates. This is followed by unspecified structure of the variance and regression discriminant procedure in that order.

Keywords: Classification Rule, Elongated Discriminant, Regression Discriminant, Unspecified variance –covariance structure and univariate populations

INTRODUCTION

Consider the classification rule for classifying an entity into one of the two predetermined univariate normal populations π_1 and π_2 based on observations collected at a single point in time.

Let $\pi_i \sim N_p(\mu_i, \sigma^2)$, i = 1, 2 ; be the two predetermined univariate normal populations with mean μ_i and a common variance σ^2. When μ_i and σ^2 are known, the optimal classification rule is:

$$R_{opt}: \; Classify \;\; e(Y_0) \text{ int} o \begin{cases} \pi_1 \; if \;\; [Y_0 - (\mu_1 + \mu_2)]\sigma^2(\mu_1 - \mu_2) \geq c \\ \pi_2 \; otherwise. \end{cases} \quad \cdots 1$$

Where, c = In q_2 [C(1/2)-C(1/1)] / q_2[C(2/1)-C(2/2)]; q_i ,i = 1, 2 is the prior probability that the entity is from population π_i and c(i/l), i ≠ l = 1, 2 ; is the cost of misclassifying a member of π_i into π_l. When q_1 = q_2 = ½, c(1/2) – c(1/1) = c(2/1) – c(2/2) and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ then c = 0 in (1) above

The plug-in method is used to obtain sample based classification rules for each procedure. This method entails replacing the unknown parameters by their estimates. The Plug-in method is widely used because of its simplicity, Ching – tu and Chien-Pai (1982).To generate data for estimating the unknown parameters, training sample of sizes n_1 and n_2 are drawn from population π_1 and π_2 respectively.

Suppose on the j[th] entity of population π_i, observation is collected from each of k equally spaced point in time t. Denote the k observations vector by:

$$Y_{ij} = \left[Y_{ij1}\ Y_{ij2}\ .\ .\ .\ Y_{ijt}\ .\ .\ .\ Y_{ijk}\right]' \quad \cdots 2$$

where;

i = 1, 2; indicating that the observation is obtained from population 1 or 2.

j =1,2, …, n_i ;indicates the observation is obtained from entity j.

t =1,2,…k ; indicates the observation is obtained at time t

Let $Y_1 = \left[Y_{1(n_1 x k)}\right]$ *and* $Y_2 = \left[Y_{2(n_1 x k)}\right]$ be the random matrices containing the observations collected sequentially in time from population π_1 and π_2 respectively.

Three procedures Unspecified Structure of the Variance (USV), Regression Discriminant (RD) and Elongated Discriminant (ED) were proposed for the classification based on the above data setting. The Unspecified Structure of the Variance uses the average mean regression to derive the Discriminant function that classifies new entities into one of two univariate populations π_1 and π_2..The regression discriminant procedure entails obtaining a temporal trend for each population $\pi_i,\ i = 1,2.$ Autoregressive model of lag one [A R (1)] is used to describe the temporal trend and a linear discriminant function based on the regression coefficients is derived to classify new entities. The Elongated discriminant procedure uses the k repeated observations drawn from two predetermined populations π_i; i = 1, 2 to derive the discriminant function that classifies the new entities.

Three techniques (Re-substitution technique, Leave- one out (LOO) and Partition of sample) of computing probability of misclassification were employed to evaluate the performances of the classification rule.

LITERETURE REVIEW

McLachlan (1992) stated that classification is an outright assignment of an object with characteristic vector $\underset{\sim}{Y_0}$ to one of g (g>2) possible populations $\pi_{i;}$ i = 1, 2, … , g.

McLachlan (1992) stated 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. He added that 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 $0<\rho<1$.

McLachlan (1992) also said that Lawoko and McLachlen (1985) compared the performance of the simple normal linear discriminant rule with the plug-in sample version formed using the maximum likelihood estimates of the group mean and variances appropriate for an AR(k) model using a univariate training data that follows a stationary auto regressive process of order k AR(k)

According to McLachlan (1992), Lawoko and McLachlan (1985) observed that there was no first order difference between the overall unconditional error rates, but that there are first order changes of varying degrees between the group-specific error rates.

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.

THE MODELS

This section derives the classification rules based on a single variable obtained sequentially in time. Supposing on every entity j; k repeated measurements are made on a continuous random variable Y at an equally spaced fixed point in time t, denoted by Y_{ijt}; t =1, 2,…,k; j = 1,2...,n_i; i = 1,2. Sample based classification rules for two predetermined univariate normal populations are derived in sections 3.1, 3.2 and 3.3 using the unspecified structure of the variance $\left(\hat{R}_{us}^{(1)}\right)$, regression discriminant procedure $\left(\hat{R}_{rd}^{(1)}\right)$ and elongated discriminant procedure $\left(\hat{R}_{ed}^{(1)}\right)$.

Let $Y_1 = \left[\underset{\sim}{Y'_{11}}\ \underset{\sim}{Y'_{12}} \ldots \underset{\sim}{Y'_{1j}} \ldots \underset{\sim}{Y'_{1n_1}}\right]'$ *and* $Y_2 = \left[\underset{\sim}{Y'_{21}}\ \underset{\sim}{Y'_{22}} \ldots \underset{\sim}{Y'_{2j}} \ldots \underset{\sim}{Y'_{2n_1}}\right]'$

be the matrices containing the observations obtained from training samples of population π_1 and π_2 respectively.

UNSPECIFIED STRUCTURE OF THE VARIANCE

Let $\bar{Y}_{ij.}^{(1)}$ *and* $\bar{Y}_{ij.}^{(2)}$ be the means observation of the j^th^ entity of population π_i, and $\underset{\sim}{\bar{Y}_{i.}^{(1)}}$ *and* $\underset{\sim}{\bar{Y}_{i.}^{(2)}}$ an n_i –vector containing the sample means of all entities in populations π_i Again let, $\bar{Y}_{i..}^{(1)}$ *and* $\bar{Y}_{i..}^{(2)}$ be the overall sample means of population π_i

Let, $s_{ij}^{(11)}$ *and* $s_{ij}^{(21)}$ be the sample corrected sum of squares and products obtain from the j^th^ entity of population π_i and are assumed equal for all entities of both populations.

Let $s_i^{(11)}$ *and* $s_i^{(21)}$ be the corrected sum of squares and products for population π_I respectively. Therefore the pooled sum of squares and product are:

$$s^{(11)} = \tfrac{1}{2}\left(s_1^{(11)} + s_2^{(11)}\right) \quad and \quad s^{(21)} = \tfrac{1}{2}\left(s_1^{(21)} + s_2^{(21)}\right)$$

Now let $\hat{Y}_{ij}^{(21)}$ be the sample mean regression of entity j in population π_i; *i* = 1, 2; j = 1,2, …,n_i. Thus

$$\hat{Y}_{ij}^{(21)} = \bar{Y}_{ij.}^{(2)} + \left(s^{(21)}/s^{(11)}\right)\left(Y_{i(t-1)j} - \bar{Y}_{ij.}^{(1)}\right) \qquad \cdots 3$$

Let $\bar{\hat{Y}}_i^{(21)}$ be average mean regression of population π_i.

On the entity to be classified, compute $\hat{Y}_0^{(21)}$ as in equation 3 above. Naturally, a new entity with mean regression $\hat{Y}_0^{(21)}$ from population π_0, believed to be a member of π_1 or π_2 can be classified as:

$$\hat{R}_{NS}^{(1)} : classify\ e\left(\hat{Y}_0^{(21)}\right) \text{ int} o \begin{cases} \pi_1 \text{ if } \hat{Y}_0^{(21)} < \tfrac{1}{2}\left(\bar{\hat{Y}}_1^{(21)} + \bar{\hat{Y}}_2^{(21)}\right) \text{ when } \bar{\hat{Y}}_1^{(21)} < \bar{\hat{Y}}_2^{(21)} \\ \pi_2 \text{ otherwise} \end{cases} \qquad \cdots 4$$

REGRESSION DISCRIMINANT (RD) PROCEDURE

The regression function of Y_{itj} on $Y_{i(t-1)j}$ is used to describe the temporal trend of a sampled entities j in population π_I, given as:

$$\underset{\sim}{Y_{itj}} = H_{ij}\,\underset{\sim}{\hat{\beta}_{ij}} + \underset{\sim}{\varepsilon_{itj}} \qquad \cdots 5$$

Where,

$\underset{\sim}{Y_{itj}}$ is a (k-1)- vector consisting of all observations collected on sampled entity j, of population π_i at time t, t = 2,3,…k, j = 1,2,…n_i; i = 1, 2

H_{ij} is a (k-1) X 2 matrix consisting of ones in the first column and observations $Y_{i(t-1)j}$ collected from sample entity j of population π_i ;at time (t-1) t=2, 3,…k

$\underset{\sim}{\hat{\beta}_{ij}}$ is a 2x1 vector consisting of the regression coefficients of entity j in population π_i.

$\underset{\sim}{\varepsilon_{itj}}$ is a (k-1) - vector consisting of the error terms of j^{th} entity in population π_i; at time t. *t= 2, 3 ,. . . , k; j = 1,2, . . .,n_i ; i = 1, 2.*

ASSUMPTIONS

(i) $E[Y_{itj}/t] = H_{ij}\,\underset{\sim}{\beta_{ij}}$;that is, $\hat{Y}_{itj} = H_{ij}\,\underset{\sim}{\hat{\beta}_{ij}}$. (ii) $\mathrm{Var}(Y_{it}/t) = \sigma^2$; where σ^2 is assumed to be identical for both populations at all times. (iii) ε_{it} follows a general autogression process of lag one, satisfying the stationary condition.

The model considers Y_t and Y_{t+h} observations obtained at time t and t + h respectively as a realization of two random variables Y_t and Y_{t+h}.

Let; $\mathrm{cov}(Y_{it}, Y_{it+h}) = \sigma^2\Gamma$

Where,

Γ is a (k-1) X (k-1) positive definite matrix whose elements are $1/(1-\varphi^2)$ on the diagonal and $\rho_h/(1-\varphi^2)$, $h \neq 0$ off the diagonal.

The parameters $\hat{\beta}_i$ *and* V_i are unknown. They are estimated from training samples. Thus the estimate of V_i is

$$V_i = \frac{1}{n_i - 1}\left[\sum_{j=1}^{n_i} \underset{\sim}{Y_{ij}}\,\underset{\sim}{Y'_{ij}} - n_i\,\underset{\sim}{\bar{Y}_i}\,\underset{\sim}{\bar{Y}_i}'\right] \qquad \cdots 6$$

Hence,

$$v_{t(t+h)} = \frac{\sigma^2\rho_h}{1-\varphi^2} = \frac{1}{(k-h-1)(n_i-1)}\left[\sum_{t=1}^{k-h-1}\sum_{j=1}^{n_i} Y_{itj}Y_{i(t+h)j} - n_i(k-h-1)\,\bar{Y}_{it}\,\bar{Y}_{i(t+h)}\right] \qquad \cdots 7$$

Let A and C be the estimates of $\hat{V}_1$ and $\hat{V}_2$ respectively; and S the estimator of the pooled variance covariance matrix. Since $n_1 = n_2$ in this study, S = ½ (A + C), is a (k-1) x (k-1) common auto-covariance matrix

Since the auto-covariance matrix (S) is not equal to $\sigma^2 I_{(k-1)}$, the ordinary least square (OLS) estimator of β_i will never be the best linear unbiased estimator (BLUE). However, Timm (1979) stated that a straightforward transformation adjusts the variance covariance matrix so as to use the ordinary least square (OLS).

Let $E[\varepsilon] = 0 \quad and \quad \mathrm{var}[\varepsilon] = S$.

Let the transformed regression model be:

$$b = S^{-1/2} H_i \beta_i + \Omega \qquad \cdots 8$$

where, $b = S^{-1/2} Y_{it}$; and $\Omega = S^{-1/2}\varepsilon$. Therefore $\mathrm{var}(\Omega a = \mathrm{var}(S^{-1/2}\varepsilon) = S^{-1}\mathrm{var}(\varepsilon a = S^{-1} S = I$

With this transformation, the model satisfies the assumption of the Gauss-Markov theorem. Therefore, applying this theorem, the best linear unbiased estimator of $\hat{\beta}_{ij}$ is obtained.as:

$$\hat{\beta}_{ij} = \left(H'_{ij} S^{-1} H_{ij}\right)^{-1} H'_{ij} S^{-1} \underset{\sim}{Y_{itj}} \qquad \cdots 9$$

.Now consider, $\hat{\beta}_i$ as a 2 x n_i matrix, consisting of the estimated regression coefficients of all the entities in population π_i ;i = 1,2. Post multiplying $\hat{\beta}_i$ by $1_{n_i}\left(1_{n_i}' 1_{n_i}\right)^{-1}$, we obtain a 2 x1 vector consisting of the means of the estimated regression coefficients of population π_i ;i = 1,2. $\underset{\sim}{1_{n_i}}$ is an n_i – vector whose elements are all unity.

The common variance is estimated by: $\mathrm{var}(\hat{\beta}) = \left(H' S^{-1} H\right)^{-1}$

On the entity to be classified, let $\underset{\sim}{\hat{\beta}_0}$ be the regression coefficients, given as:

$$\underset{\sim}{\hat{\beta}_0} = \left[\hat{\beta}_{00} \quad \hat{\beta}_{01}\right]'$$

Using the plug –in method, the sample based regression discriminant classification rule is

$$\hat{R}_{rd} : Classify \; e\left(\underset{\sim}{\beta_0}\right) \text{ into} \begin{cases} \pi_1 \; if \; \hat{\omega} = \left[\underset{\sim}{\beta_0} - \left(\underset{\sim}{\beta_1} + \underset{\sim}{\beta_2}\right)\right]' \left(\underset{\sim}{\beta_1} - \underset{\sim}{\beta_2}\right) \geq 0 \\ \pi_2 \;\; otherwise \end{cases} \qquad \cdots 10$$

The total probability of misclassification for the sample based classification rule obtained using this procedure is:

$$P\hat{M}C = \Phi\left(CD_{\hat{\beta}}^{-1} - \tfrac{1}{2} D_{\hat{\beta}}\right) + \Phi\left(CD_{\hat{\beta}}^{-1} + \tfrac{1}{2} D_{\hat{\beta}}\right), \text{ where, } D_{\hat{\beta}} = \left(\bar{\hat{\beta}}_1 - \bar{\hat{\beta}}_2\right)' H'H\left(\bar{\hat{\beta}}_1 - \bar{\hat{\beta}}_2\right) s^{-2}$$

is the Mahalobis square distance based on the regression coefficients obtained from training data. s^2 is the sample variance.

ELONGATED DISCRIMINANT PROCEDURE

Let Y_i be an $n_i \times k$ matrix consisting of observations collected sequentially in time from n_i sampled entities as presented in 4

Let $\underset{\sim}{\bar{Y}_i} = \left[\bar{Y}_{i1}\ \bar{Y}_{i2}\ \ldots\ \bar{Y}_{it}\ \ldots\ \bar{Y}_{ik}\right]'$ $\cdots$ 11

be a k-vector consisting of the estimated means at k time points.

Let Ω be a k $\times$k positive definite matrix estimated as: $\Omega_i = s_i^2 R_i$

The pooled estimate for the two populations case is; $\Omega = \frac{1}{2}\left(\Omega_1 + \Omega_2\right)$

Now let $\underset{\sim}{Y_0} = \left[Y_{01}\ Y_{02}\ \ldots\ Y_{0t}\ \ldots\ Y_{ok}\right]'$ be a k-vector consisting of the observations collected sequentially in time from the entity to be classified.

The sample based classification rule is:

$$\hat{R}_{ed}\ \text{:Classify } e\left(\underset{\sim}{\bar{Y}_o}\right) \text{ into} \begin{cases} \pi_1 \text{ if } \hat{\lambda} = \left[\underset{\sim}{\bar{Y}_o} - \left(\underset{\sim}{\bar{Y}_1} + \underset{\sim}{\bar{Y}_2}\right)\right]'\left(\underset{\sim}{\bar{Y}_1} - \underset{\sim}{\bar{Y}_2}\right) \geq 0 \\ \pi_2 \text{ otherwise} \end{cases} \qquad \cdots 12$$

With total probability of misclassification as obtained above.

APPLICATION TO REAL LIFE SITUATION

Mean arterial pressure of 120 sampled hypertensive patients admitted at the (J.U.T.H.) was collected and used for constructing and evaluating the sample based classification rules. Sixty (60) patients were randomly sampled from each of the two predetermined populations,π_1 (survivors) and π_2 (non- survivors). Ten (10) repeated measurements were collected at an equally spaced fixed point in time (6 am and 6pm) for five days. The clinical data are treated as samples of sizes n_{it} drawn from $N_{it}\left(\mu, \sigma_{it}^2\right)$; i = 1,2; t = 1, 2 ,. . . 10.The sample sizes n_{it} are equal at any point in time t. That is $n_{11} = n_{12} = \ldots = n_{110} = n_{21} = n_{22} = \ldots = n_{210} = 60$

Pooled variance is used as an estimator of the common variance. Therefore, test of homogeneity of variances is carried out to confirm or reject the hypothesis that the variances are equal.

Re- substitution technique: In this method the training sample used to construct the classification function and is re- used to evaluate the function. It is used to compute the apparent error rate (APERA), defined as a fraction of the entities in the training sample that are misclassified by the sample based classification rule. The major disadvantage of this technique is that the actual error rate (AER) tends to be underestimated.

Here we employed all the clinical data collected on the sampled entities to compute the classification rule for each procedure and a confusion matrix to evaluate the rule as follows:

i) Unspecified structure of the variance: In this procedure we compute the following:

$\bar{\bar{y}}_1^{(1)} = \mathbf{111.7846}$ $\bar{\bar{y}}_1^{(2)} = \mathbf{109.6439}$ $\bar{\bar{y}}_2^{(1)} = 114.3407$ $\bar{\bar{y}}_2^{(2)} = \mathbf{113.0543}$

The pool sum of squares (s_{11}) and cross products (s_{21}) are: $s_{11} = 365630.4$ $s_{21} = 266789.1$

Therefore, the mean regression for population π_1 and π_2

$$\hat{\bar{y}}_{1t(t-1)} = 109.6439 + \frac{266789.1}{365630.4}\left(\bar{y}_{1(t-1)} - 111.7848\right); \text{ when } \bar{y}_{1k} = 104.3294 \quad \text{then } \hat{\bar{y}}_{1k(k+1)} = 104.1974$$

Similarly,

$$\bar{\bar{y}}_{2t(t-1)} = 113.0545 + \frac{266789.1}{365630.4}\left(\bar{y}_{2(t-1)} - 114.3047\right); when\ \bar{y}_{2k} = 107.2515\ then; \bar{\hat{y}}_{2k(k+1)} = 107.908$$

The cut-off point is: Z = ½(104.1947 + 107.9088) =106.0518 if $\bar{\hat{y}}_{1k(k+1)} < \bar{\hat{y}}_{2k(k+1)}$ the classification rule is:

$$\hat{R}_{Ns}^{(1)}\textbf{:classify } o\left(\hat{y}_{ok(k+1)}\right)\textbf{into} \begin{cases} \pi_1 \ if\ \hat{y}_{ok(k+1)} \le 106.0518 \\ \pi_2\ otherwise. \end{cases}$$

ii) Regression Discriminant Procedure: Here the following parameters were computed to construct the regression discriminant function used in deriving the sample based classification rule.

$$\hat{\beta}_{\tilde{1}} = \left[104.1132\ \ 0.03442515\right]' \qquad \hat{\beta}_{\tilde{2}} = \left[114.9184 - 0.01352404\right]'$$

The classification rule is obtained as: $\hat{R}_{rd}^{(1)}: classify\ \ e(\hat{\beta}_{\tilde{o}})\ \mathrm{int}\ o \begin{cases} \pi_1\ if\ \ \hat{\omega} \ge 0 \\ \pi_2\ otherwise \end{cases}$

where $\hat{\omega} = \left[\hat{\beta}_{01} - 109.5158 \quad \hat{\beta}_{02} - 0.01045055\right] H' \Gamma^{-1} H \begin{bmatrix} -10.80519 \\ 0.04794919 \end{bmatrix} (0.0014718)$

H is a 9 x 2 random matrix consisting the observation collected from the object to be classified at time t-1,t = 2,3,…,10 ; and Γ^{-1} is the inverse of Γ

iii).**Elongated Discriminant Procedure:** The classification function obtained here uses the single discriminant procedure.

$$\hat{R}_{sd}^{(1)} : classify\ \ e\left(\underset{\sim}{Y_0'}\right)\ \mathrm{int}\ o \begin{cases} \pi_1\ if\ \ \hat{\lambda} \le 0 \\ \pi_2\ otherwise \end{cases}$$

where. $\hat{\lambda}$ = $0.1093Y_{01} - 0.0634\ Y_{02} + 0.0050\ Y_{03} - 0.0159\ Y_{04} - 0.0494Y_{05} + 0.0264\ Y_{06} - 0.03.64\ Y_{07} + 0.0391Y_{08} - 0.0\ 244\ Y_{09} - 0.0029Y_{010} + 0.6691$

From the confusion matrix the empirical estimate of the corresponding probability of misclassification (PM.C), referred to as error rates for the three procedures are presented in Table 1.

Table 1: Probabilities of Misclassification (PM.C) for the Three Procedures (Re-substitution technique)

Procedures	Unspecified variance	Regression Discriminant	Elongated Discriminant
PMC	0.4417	0.4917	0.3667

2) Leave- one out (LOO) Technique: This technique according to Lachenbruch (1975) estimates the expected actual error rates. An entity Y_{11} with measurement vector $\left[y_{111}\ y_{112}\ ...\ y_{11t} ..\ .\ y_{11k}\right]'$ is omitted in population π_1 and the sample based classification rule is constructed using the rest of n_1-1 in π_1 and n_2 from population π_2. The classification rule is then used to classify the entity left out (Y_{11}). This is done for all sampled entities in population π_1 and π_2. This technique is better than all the methods discussed. The only disadvantage of this procedure is the cost, in the sense that $n_1 + n_2 = n$ different classification rules will be computed. This technique is used to compute the probability of misclassification as shown in Tables 2

Table 2: Probabilities of Misclassification (PM.C) for the Three Procedures (Leave- one out Technique)

Procedures	Unspecified variance	Regression Discriminant	Elongated Discriminant
PMC	0.48083	0.5000	0.4667

3) **The technique of partition of sample**: Partition of sample is another method of evaluating the performances of the sample based classification rule. Here half of the sampled entities in each population (1/2 n_1 and 1/2 n_2) are used to construct the classification rule and the other half to evaluate it. The major disadvantages of this technique include: i) It wastes data (ii) Requires large samples that are often not available in practice (iii) The classification function evaluated may not be the function of interest as some valuable information may be lost by not using all samples in constructing the function. Here we used half of the training sample ($n_1 = n_2 = 30$) to construct the classification function and the remaining half as validation sample to evaluate.

i) **Unspecified Structure of the variance**: In this procedure, the following parameters were computed using half of the data points, the results are as follows:

$$\bar{\bar{y}}_1^{(1)} = 109.1247 \quad \bar{\bar{y}}_1^{(2)} = 06.9408 \quad \bar{\bar{y}}_2^{(1)} = 115.5666 \quad \bar{\bar{y}}_2^{(2)} = 114.3222$$

The pooled sum of squares (s_{11}) and cross products (s_{21}) are:

s_{11} = 140680.6 s_{21} = 86191.92

Therefore, the mean regression for population π_1 and π_2 when this technique is used are:

$$\bar{\hat{y}}_{1t(t-1)} = 109.1247 + \frac{86191.92}{140680.6}\left(\bar{y}_{1(t-1)} - 109.1247\right);\ \text{when, } \bar{y}_{1k} = 98.7687,\ \text{then}\ \bar{\hat{y}}_{1k(k+1)} = 102.7798$$

similarly

$$\bar{\hat{y}}_{2t(t-1)} = 114.3222 + \frac{86191.92}{140680.6}\left(\bar{y}_{2(t-1)} - 115.5666\right);\ \text{when, } \bar{y}_{2k} = 107.5333,\ \text{then}\ \bar{\hat{y}}_{2k(k+1)} = 109.4004$$

The cutoff point Z is:Z = ½(102.7798 + 109.4004) = 106.0901

Since $\bar{\hat{y}}_{1k(k+1)} < \bar{\hat{y}}_{2k(k+1)}$ the classification rule and confusion matrix obtained by this technique are:

$$\hat{R}_{Ns}^{(1)}: \ \textbf{classify}\ o\left(\hat{y}_{ok(k+1)}\right)\ \textbf{into}\ \begin{cases}\pi_1 \ \ if \ \ \hat{y}_{ok(k+1)} \le 106.0901 \\ \pi_2 \ \ otherwise\end{cases}$$

ii) **Regression Discriminant Procedure:** When half of the training sample ($n_1 = n_2 = 30$) is used to construct the classification function and the remaining half as validation sample to evaluate; the following classification rule is obtained.

$$\bar{\hat{\beta}}_{\underset{\sim}{1}} = [109.2252 \ \ \text{-}0.02864476]; \ \bar{\hat{\beta}}_{\underset{\sim}{2}} = [105.683 \ \ 0.06189757] \text{ Thus;}$$

$$\hat{R}_{rd}^{(1)}: classify \ \ o(\hat{\beta}_{\underset{\sim}{o}}) \operatorname{int} o \begin{cases}\pi_1 \ if \ \hat{\omega} \ge 0 \\ \pi_2 \ otherwise\end{cases}$$

Where, $\hat{\omega} = \left[\hat{\beta}_{00} \text{ -107.4541} \quad \hat{\beta}_{01} \text{ - 0.0166264}\right] H' \ \Gamma^{-1} H \begin{bmatrix}3.54213 \\ \text{-}0.09054233\end{bmatrix}(0.00189764)$ and Γ^{-1} **is the inverse of** Γ

iii) **Elongated Discriminant Procedure:** If only half of the training sample ($n_1 = n_2 = 30$) are used to compute the classification function, then the classification rule is:

$$\hat{R}_{sd}^{(1)} : classify\ e(\underline{y}_0)\ \text{int}\,o \begin{cases} \pi_1\ if\ \ \hat{\lambda} \geq 0 \\ \pi_2\ otherwise \end{cases}$$

Where, $\hat{\lambda}$ = $0.0866Y_{01} - 0.0398Y_{02} - 0.0035Y_{03} - 0.0278\ Y_{04} - 0.0206\ Y_{05} + 0.0056\ Y_{06} - 0.0292Y_{07} + 0.0450Y_{08} - 0.0267Y_{09} - 0.0175Y_{010} + 6.8054$

The probability of misclassification for the three procedures obtained from the confusion matrix is shown in Tables 3:

Table 3: Probabilities of Misclassification (PM.C) for the Three Procedures (Partition of Sample Technique)

Procedures	Unspecified variance	Regression Discriminant	Elongated Discriminant
PMC	0.5333	0.5500	0.3667

CONCLUSSION

The analyses reveal that whichever technique is used to construct and evaluate the sample based classification rule, the Elongated Discriminant procedure out performs the other two, with minimum probability of misclassification. This is followed by the Unspecified Structure of the Variance and then the Regression Discriminant procedures. The Re-substitution Technique is found to be most appropriate when estimating the apparent error rate (APERA), as this gives the minimum error rate for all the procedures. When actual error rate is desired, the technique of Leave one out and partition of sample are most appropriate.

REFERENCES

1. Bulus, H (2008). A Heuristic classification for Repeated Measure. Bagale Journal of Pure and Applied Sciences, Yola. Vol. 6, pages 38-46.
2. Ching-Tsao Tu and Chien-Pai Han (1982). Discriminant Analysis Based on Binary and Continuous Variables. Journals of the American Statistical Association.Vol.77, no 377, pages 447 – 454.
3. Hand, D. J. (1989). Discrimination and Classification. John Wiley and sons, New York.
4. Lachenbruch, P. A. (1975). Discriminant Analysis. Hafner press, New York.
5. Lawoko, C.R.O. and McLachlan, G.J. (1983). Some Asymptotic Results on the Effect of Autocorrelation in the error Rate of the Sample Linear Discriminant Function. Pattern Recognition, Vol. 16, pages. 119 – 121.
6. McLahlan, G. J. (1992). Discriminant Analysis and Statistical Pattern Recognition. John Wiley and sons Inc, New York.
7. Timm, N. H. (1975). Multivariate Analysis with Applications in Education and Psychology. Brooks / Cole Publishing Company, Monterey, California.