

# Performance Evaluation of some Zero Mean Classification Functions under Unequal Misclassification Cost

Michael Asamoah-Boaheng<sup>1</sup> and Atinuke. O. Adebajji<sup>2</sup>

<sup>1</sup>*School of Graduate Studies, Research and Innovation,*

*Kumasi-Polytechnic, Box 854, Kumasi, Ghana and*

<sup>2</sup>*Department of Mathematics, Kwame Nkrumah*

*University of Science and Technology, Kumasi, Ghana.*

\*Corresponding author: [asboaheng@yahoo.com](mailto:asboaheng@yahoo.com), +233 543160192

## Abstract

*This study considered the performance of Minimum Expected Cost of Misclassification method (MECM) and Quadratic Discriminant Function approach (QDF) for equal mean discrimination under unequal misclassification cost. Discriminant functions were derived under each of the following misclassification cost ratios; 1: 1, 1: 2, 1: 3 and 1: 4 and their error rate estimates determined using the Cross Validation (CV) and Balanced Error Rate (BER) methods. Least Mean error rates were recorded under QDF method as compared to that of MECM. The error rate estimates showed the QDF out performing the MECM in the provision of maximum separation between the two groups. Also the two classification rules were found to be sensitive to misclassification cost ratios exceeding 1:2.*

**Keywords:** *Classification rules, Quadratic Discriminant Function, Error Rates, Discriminant Functions, Minimum Expected Cost of Misclassification.*

## Introduction

The problem of equal mean classification or zero mean difference has posed a challenge to researchers for a long time and several attempts have been made at deriving parsimonious rules that address this hurdle. This study considered the equal mean discrimination problem by evaluating the performance of MECM and the QDF under the case of unequal misclassification cost. The problem of discrimination was first initiated by Fisher (1936) in his paper titled “the use of multiple measurements in taxonomic problems”. Fisher’s approach to classification with two populations was based on arriving at a linear classification statistics using an entirely different argument. Fisher’s idea was based on transforming multivariate observations ‘x’ to univariate ‘y’ observations such that ‘y’ being derived from either population one or population two were truly separated as much as possible. (Johnson and Wichern, 2007). Hyodo and Kubokawa (2014) studied a variable selection criterion for linear discriminant rule and its optimality in high dimensional data where a new variable procedure was developed for selecting the true variable set.

Okamoto (1961) first studied the problem of discrimination with common mean vector and with different covariance structure of two multidimensional normal populations.

Bartlett and Please (1963) considered the specific problem of zero-mean uniform discrimination. The problem was investigated for the case when the variance covariance matrices for the two groups were known to have uniform structures, with an assumption of equal and unequal correlation coefficients ( $\rho$ ). A discriminant function based on equal correlation coefficient as well as uniform covariance matrix was obtained for classificatory purposes. Desu and Geisser (1973) studied the same problem of common mean discrimination by focusing on the case where the difference of the means of the two groups are the zero vector. Geisser and Desu based their discrimination on several methods in order to compare the performance of each of them after their application. They employed the classical approach, semi Bayesian and complete Bayesian approach, Discrimination with equal and unequal covariance matrices. Lachenbruch (1975) derived a simple model for discriminating among equal mean data for two populations and the problem of assigning observations to one of the two populations by an investigation into the covariance matrices of populations using the absolute deviations. He observed that the Absolute Linear Discriminant Function (ALDF) was almost as good as the QDF when the two groups are closer. In concluding, the derived absolute linear discriminant rule performed slightly worse than the quadratic discriminant rule as used earlier by Bartlett and Please (1963) and Desu and Geisser (1973). His absolute linear rule performed reasonably well, after the data was contaminated by the introduction of outliers into the two groups/populations. On the other hand, the quadratic discriminant rule performed poorly after the contamination.

Marco et al. (1987) studied the asymptotic expected error rate for the equal-mean uniform-covariance discrimination problem. They approximated the unconditional expected error rate of the sample discriminant function up to the second order term and compared with the Monte Carlo simulation evaluated at several combinations of the parameters to ensure accuracy of the approximation of the error rate. Ariyo and Adebajji (2010) compared the performance of both Linear and Quadratic classifier under unequal cost of misclassification and concluded that both classifiers are insensitive to some specified cost ratios. They found that the procedures were insensitive to cost ratio exceeding ratio 1:2. Adebajji et al. (2008) investigated the performance of the homoscedastic discriminant function (HDF) under the non-optional condition of unequal group representation (prior probabilities) in the population and the asymptotic performance of the classification function. The misclassification of observation from the smallest group increased when the sample size ratio exceeded the sampling ratio 1:2 and this resulted in increases in error rate and was not corrected by increases in the sample size.

## Materials and Methods

The concept of Discrimination and Classification are concerned with separating objects from different populations into different groups and with allocating new observations to one of these groups. Discriminant analysis is rather exploratory in nature and used as a separative procedure which is normally employed on a one time basis. This section outlines the various methods and materials employed in obtaining the results for the study.

## Data Used

Stocks (1933) recorded the measurements he made on 832 children in Elementary and Central Schools of the London County Council during the period 1925 to 1927. Of these 832 children, 392 boys and 440 girls, 563 were members of twin pairs, including 106 like-sexed pairs and the

remainder were mostly siblings of twins or pairs of siblings unconnected with twins. As well as the age and sex, Stocks recorded the following measurements for most of the children; Height, Weight, Head length, Head breadth, Head circumference, Interpupillary distance, Systolic blood pressure, Diastolic blood pressure (DBP), Pulse interval, Respiration interval, Strength of grip right hand, Strength of grip left hand, Visual acuity right eye (VARE) and Visual acuity left eye (VALE). For the twin pairs he also recorded their finger print and facial resemblance and derived an empirical criterion, based on the finger prints, height and head measurements, for separating the like sexed twins into Monozygotic and Dizygotic pairs. 30 pairs of female liked-sexed twins were selected with 15 pairs of twin observations selected from both the Monozygotic and Dizygotic twins after ten (10) replications using simple random sampling without replacement. Ten (10) of the measurements were included, while measurements DBP, VARE and VALE being omitted owing to the large numbers of missing values. The difference between the first and the second recorded twin was taken as an observation for each variate. R-console version 2.15.1 was used to analyse the data.

### ***Discrimination and Classification of Two Populations***

Let  $\pi_1$  and  $\pi_2$  be two  $p$ -variate groups with density functions  $f_1(x)$  and  $f_2(x)$  respectively. Now consider an observed value  $X = (x, \dots, x)^T$ . The observed vector  $X$  must be assigned to either population  $\pi_1$  or  $\pi_2$ . Denote  $\Omega$  the sample space (collection of all possible outcomes of  $X$ ) and partition the sample space as  $\Omega = R_1 \cup R_2$  where  $R_1$  is the subspace of outcomes which we classify as belonging to population  $\pi_1$  and  $R_2 = \Omega - R_1$  the subspace of outcomes classified as belonging to  $\pi_2$ . It follows therefore that the (conditional) probability of classifying an object as belonging to  $\pi_j$  when it really comes from  $\pi_i$  equals;

$$P(i | j) = P(X \in R_j | X \in \pi_i) = \int_{R_j} f_i(x) dx, \quad \forall i, j \quad i \neq j \quad (1)$$

The conditional probabilities can also be obtained for  $i = j$  when  $i, j = 1, 2$ . Prior class probabilities are obtained when we want to obtain the probability of correctly and incorrectly classifying an observation/objects. We denote the following; Let  $P_i = P(X \in \pi_i)$ ,  $i = 1, 2$  be the prior probability of  $\pi_i$  where  $P_1 + P_2 = 1$ . The overall probabilities of correctly and incorrectly classifying observations are:

$$P(\text{object is correctly classified as } \pi_i) = P(X \in R_i | X \in \pi_i) = P(X \in \pi_i) = P(i | i)p_i \quad \text{where } i = 1, 2$$

$$P(\text{object is misclassified as } \pi_i) = P(X \in R_i | X \in \pi_j) = P(X \in \pi_j) = P(i | i)p_j \quad \text{where } i \neq j$$

### ***Cost of Misclassification***

Denote  $c(i | j)$  as the cost of classifying an object/observation into  $\pi_i$  when actually belongs to  $\pi_j$ . Where the ECM is derived as

$$ECM = c(2 | 1)P(2 | 1)p_1 + c(1 | 2)P(1 | 2)p_2 \quad (2)$$

Where  $p_1$  and  $p_2$  are the prior probabilities for the two populations. The two regions  $R_1$  and  $R_2$  are used to minimized the expected cost of misclassification.

$$R_1 = \left\{ x \in \Omega; \frac{f_1(x)}{f_2(x)} \geq \left( \frac{c(1 | 2)}{c(2 | 1)} \right) \left( \frac{p_2}{p_1} \right) \right\} \quad (3)$$

$$R_2 = \left\{ x \in \Omega; \frac{f_1(x)}{f_2(x)} < \left( \frac{c(1|2)}{c(2|1)} \right) \left( \frac{p_2}{p_1} \right) \right\} \quad (4)$$

(Johnson and Wichern, 2007).

### **Classification of Normal Populations when $\Sigma_1 \neq \Sigma_2$ (Quadratic Classification Rule)**

The regions of minimum ECM and minimum Total Probability of Misclassification (TPM) depends on the ratio of the densities. Hence substituting the normal densities with different covariance matrices in equation 3 and 4 after taking natural logarithm gives the following classification regions. Allocate to  $\pi_1$  or otherwise to  $\pi_2$  if,

$$-\frac{1}{2}x'(\Sigma_1^{-1} - \Sigma_2^{-1})x + (\mu_1' \Sigma_1^{-1} - \mu_2' \Sigma_2^{-1})x - k \geq h \left[ \left( \frac{c(1|2)}{c(2|1)} \right) \left( \frac{p_2}{p_1} \right) \right] \quad (5)$$

$$\text{Where } k = \frac{1}{2} h \left( \frac{|\Sigma_1|}{|\Sigma_2|} \right) + \frac{1}{2} (\mu_1' \Sigma_1^{-1} \mu_1 - \mu_2' \Sigma_2^{-1} \mu_2)$$

### **The minimum Expected Cost of Misclassification Method (MECM)**

Let  $f_i(x)$  be the multivariate normal density associated with population  $\pi_i, i = 1, 2, \dots, g$ . Let  $p_i$  be the prior probability of population  $\pi_i, i = 1, 2, \dots, g$ .  $c(k|i)$  = the cost of allocating an item to  $\pi_k$  when in fact it belongs to  $\pi_i$  for  $k, i = 1, 2, \dots, g$ . For  $k=i$ ,  $c(i|i) = 0$ . Finally, let  $R_k$  be the set of  $x$ 's classified as  $\pi_k$  and

$P(k|i) = P(\text{classifying item as } \pi_k | \pi_i) = \int_{R_k} f_i(x) dx$  for  $k, i = 1, 2, \dots, g$ . Hence the Expected Cost of Misclassification (ECM) becomes:

$$ECM = \sum_{i=1}^g p_i \left( \sum_{\substack{k=1 \\ k \neq i}}^g p(k|i)c(k|i) \right) \quad (6)$$

We seek rules that minimize the ECM. This leads to an optimal classification rule: classify an object ( $x_m$ ) into  $\pi_1$  if

$$\frac{f_1(x_m)c(2|1)p_1}{f_2(x_m)c(1|2)p_2} > 1 \quad (7)$$

Also  $x_D$  to  $\pi_2$  if

$$\frac{f_1(x_D)c(2|1)p_1}{f_2(x_D)c(1|2)p_2} > 1 \quad (8)$$

Where  $f_1(x)$  and  $f_2(x)$  are the density functions for both the Monozygotic and Dizygotic twin groups respectively.  $x_M$  and  $x_D$  being the monozygotic and dizygotic twin observations (Johnson and Wichern, 2007).

## Error Rate Estimation

The performance of any classification procedure is based on the error rates or misclassification probabilities.

### Balanced Error Rate (BER)

The Balanced Error Rate (BER) statistic is the average of the misclassification rates on samples drawn from populations  $\pi_1$  and  $\pi_2$  as shown in the table below. Where a, b, c, d are the entries in the confusion matrix.

Table 3.1: Confusion matrix for the two groups

		Prediction	
		■	■
True Pop	■	a	b
	■	c	d

The balanced error rate is given mathematically as  $BER = \frac{1}{2} \left[ \frac{b}{a+b} + \frac{c}{c+d} \right]$  (9)

### Cross Validation

Let  $n_{1M}^e$  and  $n_{2M}^e$  be the number of left out observations misclassified in group 1 and 2

respectively and it's given by  $\mathcal{C} = \frac{n_{1M}^e + n_{2M}^e}{n_1 + n_2}$  (10)

(Johnson and Wichern, 2007).

## Results and Discussion

After the application of the methods in section 2 above, the following results were obtained in studying the effect of unequal misclassification cost on the two classification functions namely QDF and Minimum Expected Cost of Misclassification (MECM).

The equality of the two mean vectors for Monozygotic and Dizygotic twin groups were tested with Hotelling  $T^2$  to ensure that the equal mean assumption is not violated. Based on the results, the test proved to be insignificant indicating that there is no significant difference between the two mean vectors.

### Application with QDF

From the methodology section the QDF obtained under equal prior probabilities ( $p_1 = p_2 = 0.5$ ) and equal misclassification cost ( $c(1|2) = c(2|1) = 1$ ) was derived as: Allocate  $x$  to  $\pi_1$ , otherwise to  $\pi_2$  if

$$\mathbf{h} \left\{ \frac{|\Sigma_1|}{|\Sigma_2|} \right\} - \left\{ (x - \mu_1)' \Sigma_1^{-1} (x - \mu_1) \right\} + \left\{ (x - \mu_2)' \Sigma_2^{-1} (x - \mu_2) \right\} > 0 \quad (11)$$

The QDF as well as the quadratic classification rules for equal and unequal misclassification ratios  $\left( \frac{c(1|2)}{c(2|1)} \right)$  for  $\pi_1 : \pi_2$  in the order 1:1, 1:2, 1:3 and 1:4 were obtained as follows:

$$\{(x - \mu_1)' \Sigma_1^{-1} (x - \mu_1)\} - \{(x - \mu_2)' \Sigma_2^{-1} (x - \mu_2)\} < 0.192 \tag{12}$$

$$\{(x - \mu_1)' \Sigma_1^{-1} (x - \mu_1)\} - \{(x - \mu_2)' \Sigma_2^{-1} (x - \mu_2)\} < -0.500 \tag{13}$$

$$\{(x - \mu_1)' \Sigma_1^{-1} (x - \mu_1)\} - \{(x - \mu_2)' \Sigma_2^{-1} (x - \mu_2)\} < -0.906 \tag{14}$$

$$\{(x - \mu_1)' \Sigma_1^{-1} (x - \mu_1)\} - \{(x - \mu_2)' \Sigma_2^{-1} (x - \mu_2)\} < -1.193 \tag{15}$$

The following QDF were obtained when the cost ratios  $\frac{c(2|1)}{c(1|2)}$  for the two groups were alternated as  $\pi_2 : \pi_1$  in the order of the following misclassification cost ratios: 1:2, 1:3, and 1:4.

$$\{(x - \mu_1)' \Sigma_1^{-1} (x - \mu_1)\} - \{(x - \mu_2)' \Sigma_2^{-1} (x - \mu_2)\} < 0.885 \tag{16}$$

$$\{(x - \mu_1)' \Sigma_1^{-1} (x - \mu_1)\} - \{(x - \mu_2)' \Sigma_2^{-1} (x - \mu_2)\} < 1.291 \tag{17}$$

$$\{(x - \mu_1)' \Sigma_1^{-1} (x - \mu_1)\} - \{(x - \mu_2)' \Sigma_2^{-1} (x - \mu_2)\} < 1.579 \tag{18}$$

Based on the above functions under the two misclassification cost ratios, the following discriminant scores were obtained in Table 1.

Table 1: Discriminant scores for the various misclassification cost ratios using QDF

c(1 2):c(2 1)		c(1 2):c(2 1)						c(2 1):c(1 2)					
1:1		1:2		1:3		1:4		1:2		1:3		1:4	
D <sub>M</sub>	D <sub>D</sub>	D <sub>M</sub>	D <sub>D</sub>	D <sub>M</sub>	D <sub>D</sub>	D <sub>M</sub>	D <sub>D</sub>	D <sub>M</sub>	D <sub>D</sub>	D <sub>M</sub> <	D <sub>D</sub> ≥	D <sub>M</sub> <	D <sub>D</sub> ≥
<0.192	>0.192	<-0.5	>-0.5	<-0.91	≥-0.91	<-1.19	≥-1.19	<0.88	≥0.88	1.29	1.29	1.58	1.58
-2.16	<b>-0.88</b>	-2.16	<b>-0.88</b>	-2.16	-0.88	-2.16	-0.88	-2.16	<b>-0.88</b>	-2.16	<b>-0.88</b>	-2.16	<b>-0.88</b>
-1.02	3.72	-1.02	3.72	-1.02	3.72	<b>-1.02</b>	3.72	-1.02	3.72	-1.02	3.72	-1.02	3.72
-1.06	2.84	-1.06	2.84	-1.06	2.84	<b>-1.06</b>	2.84	-1.06	2.84	-1.06	2.84	-1.06	2.84
0.69	9.32	<b>0.69</b>	9.32	<b>0.69</b>	9.32	<b>0.69</b>	9.32	0.69	9.32	0.69	9.32	0.69	9.32
-2.49	0.54	-2.49	0.54	-2.49	0.54	-2.49	0.54	-2.49	<b>0.54</b>	-2.49	0.54	-2.49	<b>0.54</b>
1.59	<b>-2.8</b>	<b>1.59</b>	<b>-2.8</b>	<b>1.59</b>	<b>-2.8</b>	<b>1.59</b>	<b>-2.8</b>	<b>1.59</b>	<b>-2.8</b>	<b>1.59</b>	<b>-2.8</b>	<b>1.59</b>	<b>-2.8</b>
-1.82	3.81	-1.82	3.81	-1.82	3.81	-1.82	3.81	-1.82	3.81	-1.82	3.81	-1.82	3.81
-3.92	<b>-1.51</b>	-3.92	<b>-1.51</b>	-3.92	<b>-1.51</b>	-3.92	<b>-1.51</b>	-3.92	<b>-1.51</b>	-3.92	<b>-1.51</b>	-3.92	<b>-1.51</b>
-6.08	2.49	-6.08	2.49	-6.08	2.49	-6.08	2.49	-6.08	2.49	-6.08	2.49	-6.08	2.49
-4.28	0.75	-4.28	0.75	-4.28	0.75	-4.28	0.75	-4.28	<b>0.75</b>	-4.28	<b>0.75</b>	-4.28	<b>0.75</b>
0.95	8.87	<b>0.95</b>	8.87	<b>0.95</b>	8.87	<b>0.95</b>	8.87	<b>0.95</b>	8.87	0.95	8.87	0.95	8.87
-1.38	3.45	-1.38	3.45	-1.38	3.45	-1.38	3.45	-1.38	3.45	-1.38	3.45	-1.38	3.45
-3.57	6.29	-3.57	6.29	-3.57	6.29	-3.57	6.29	-3.57	6.29	-3.57	6.29	-3.57	6.29
1.32	1.97	<b>1.32</b>	1.97	<b>1.32</b>	1.97	<b>1.32</b>	1.97	<b>1.32</b>	1.97	1.32	1.97	1.32	1.97
0.65	2.02	<b>0.65</b>	2.02	<b>0.65</b>	2.02	<b>0.65</b>	2.02	0.65	2.02	0.65	2.02	0.65	2.02

The bolded figures indicates the misclassified observations; D<sub>M</sub> and D<sub>D</sub> being the discriminant scores for Monozygotic and Dizygotic groups respectively.

From Table 1, we observed no misclassified observations from the monozygotic twin group whilst three twin pairs of observations were misclassified from the dizygotic group. The proportion of correct classification was recorded as 0.80. Five (5) and three (3) observations were misclassified from monozygotic and dizygotic twin groups representing approximately 73

percent of correct classification when the cost of misclassifying an observation as monozygotic twin was twice the cost of the misclassifying observation as dizygotic twin. For ratio 1:3, the correct proportion of classification was 0.77 since 5 and 2 twin observations were misclassified from monozygotic and dizygotic groups. Seven (7) and nine (9) were misclassified from monozygotic and dizygotic groups with 70 percent correct classification of observations. After alternating the cost ratios, 0.73, 0.83 and 0.80 were the correct proportion of classification for 1:2, 1:3 and 1:4 misclassification cost respectively (*See Table 1 above*).

### **Results for the various misclassification cost ratios using MECM classification rule**

The optimal classification rule derived in the methodology section (equations 10 and 11) were used to derive the discriminant scores as shown in Table 2 and 2a below.

Table 2: Scores obtained from the optimum classification rule based on the MECM

C(2 1): C(1 2)= 1:1		C(1 2): C(2 1)= 1:2		C(1 2):c(2 1)= 1:3		C(1 2): C(2 1)= 1:4	
D <sub>M</sub>	D <sub>D</sub>	D <sub>M</sub>	D <sub>D</sub>	D <sub>M</sub>	D <sub>D</sub>	D <sub>M</sub>	DD
2.43E+00	2.04E+00	4.86E+00	4.09E+00	7.29E+00	6.13E+00	9.71E+00	8.17E+00
4.55E+00	<b>5.18E-01</b>	9.10E+00	1.04E+00	1.36E+01	1.56E+00	1.82E+01	2.07E+00
1.28E+00	1.76E-01	2.55E+00	<b>3.52E-01</b>	3.83E+00	<b>5.28E-01</b>	5.11E+00	<b>7.03E-01</b>
<b>9.63E-01</b>	5.98E-02	1.93E+00	<b>1.20E-01</b>	2.89E+00	<b>1.79E-01</b>	3.85E+00	<b>2.39E-01</b>
2.12E+00	1.78E+00	4.23E+00	3.55E+00	6.35E+00	5.33E+00	8.46E+00	7.11E+00
<b>8.09E-01</b>	1.53E+00	1.62E+00	3.06E+00	2.43E+00	<b>4.59E+00</b>	3.24E+00	6.12E+00
3.05E+00	<b>1.10E-01</b>	6.09E+00	<b>2.21E-01</b>	9.14E+00	<b>3.31E-01</b>	1.22E+01	<b>4.42E-01</b>
5.50E+00	<b>3.31E-01</b>	1.10E+01	<b>6.63E-01</b>	1.65E+01	<b>9.94E-01</b>	2.20E+01	1.33E+00
1.90E+01	<b>3.89E-01</b>	3.81E+01	<b>7.79E-01</b>	5.71E+01	1.17E+00	7.62E+01	1.56E+00
7.95E+00	<b>7.13E-01</b>	1.59E+01	1.43E+00	2.39E+01	2.14E+00	3.18E+01	2.85E+00
1.59E+00	<b>1.27E-03</b>	3.18E+00	<b>2.53E-03</b>	4.77E+00	<b>3.80E-03</b>	6.36E+00	<b>5.07E-03</b>
2.13E+00	<b>1.11E-01</b>	4.27E+00	<b>2.21E-01</b>	6.40E+00	<b>3.32E-01</b>	8.53E+00	<b>4.43E-01</b>
3.66E+00	<b>6.27E-03</b>	7.32E+00	<b>1.25E-02</b>	1.10E+01	<b>1.88E-02</b>	1.46E+01	<b>2.51E-02</b>
1.38E+00	1.09E+00	2.76E+00	2.18E+00	4.14E+00	3.27E+00	5.52E+00	4.37E+00
1.00E+00	<b>2.55E-01</b>	2.01E+00	<b>5.10E-01</b>	3.01E+00	<b>7.66E-01</b>	4.02E+00	1.02E+00

From Table 2, the discriminant scores generated under equal cost of misclassification (i.e. with cost ratio  $c(1|2):c(2|1)=1:1$ ) was able to misclassify 2 and 9 twin pair observations from both the monozygotic and the dizygotic groups respectively with proportion of correct classification as 0.633. The proportion for correct classification of twin observations under the cost ratios 1:2, 1:3 and 1:4 were obtained as 0.70, 0.70 and 0.80 respectively. Hence as the cost of misclassifying a twin observation into dizygotic twin group increases, better and maximum separation were achieved since few observations were misclassified hence with a recorded least error rates. The misclassification cost ratios were alternated in the order  $\pi_2 : \pi_1$  and the effect of the misclassification cost on the classification rule was assessed. Based on this, the proportion of correct classification as shown in Table 2a (*See appendix*) for the misclassification costs 1:2, 1:3, 1: 4 were obtained as 0.33, 0.20 and 0.133. This results indicates that, as we increase the cost of allocating a twin observation into monozygotic group when it actually belongs to the dizygotic group, the proportion of correct classification reduces in that manner, hence increases the number of misclassified observations.

### ***Evaluating the classification rules of QDF and MECM under various misclassification costs***

The performance of two classification functions namely QDF and MECM were evaluated by estimating their error rates as a result of the misclassified observations with CV and BER methods. From Table 3, we observed generally that, the error estimates obtained using CV error estimator was quite higher than the estimates of that of the Balanced Error Rate (BER). However the mean error rates obtained for the QDF method under the cost ratios in the order  $c(1|2):c(2|1)$  recorded low error rates generally as compared to when the cost ratios were alternated in the order of  $c(2|1):c(1|2)$ . This means that the discriminant functions performs better in the provision of maximum separation between the two twin groups when the associated misclassification cost assigned to misclassifying an observation to population 2 (dizygotic group) increases. Also, the function performed slightly poor when then cost ratios exceeded 1:1 and 1:2 and somehow having no effect on the classification rule, under both misclassification cost situations. This shows some slight conformation with the research work of Adebajji et al (2008) in which they discovered that the misclassification of observation from the smallest group increased when the sample size ratio exceeded the sampling ratio 1:2 and this resulted in increases in error rate and was not corrected by increases in the sample size. Also Ariyo and Adebajji (2011) asserted that their three derived linear classifiers were insensitive under the cost ratios exceeding 1:2.

Similarly, the classification rule obtained under MECM performed slightly better when the cost of misclassifying an observation into the dizygotic group was increased beyond one in the ratio order  $c(1|2):c(2|1)$  than when it was alternated.

In comparing the two methods based on their mean error rates in Table 3, we discovered that, the error rate for the MECM was appreciably higher ranging from approximately 17% to 82%. Whilst QDF recorded least error rates ranging from approximately 17%-28% under the various misclassification cost ratios. Hence the QDF outperformed the MECM in the provision of maximum separation between the two twin observations. The results conforms to the work of Lachenbruch (1975) where his derived absolute linear discriminant rule performed slightly worse than the quadratic discriminant rule as used by Bartlett and Please (1963) and Desu and Geisser (1973).

Table 3: Error rate estimates for QDF and MECM under unequal misclassification cost ratios

Classification methods	Error rates		Mean Error
	CV	BER	
QDF			
$c(1:2):c(2:1)$			
1:1	0.266	0.100	0.183
1:2	0.233	0.107	0.170
1:3	0.166	0.233	0.199
1:4	0.200	0.295	0.248
$c(2:1):c(1:2)$			
1:2	0.233	0.267	0.250
1:3	0.233	0.166	0.199
1:4	0.366	0.200	0.283
MECM			

c(1:2):c(2:1)			
1:1	0.333	0.367	0.350
1:2	0.200	0.300	0.250
1:3	0.200	0.300	0.250
1:4	0.133	0.200	0.167
c(2:1):c(1:2)			
1:2	0.500	0.667	0.584
1:3	0.700	0.800	0.750
1:4	0.767	0.867	0.817

## Conclusion

Two classification methods were studied; QDF and MECM when the assumptions of equal misclassification cost were violated. Generally, most of the twin observations were correctly classified under the various misclassification cost ratios in the order of  $c(1|2):c(2|1)$  than when it was alternated. Hence increasing the cost of misclassifying a dizygotic observation provided maximum separation than when the cost of misclassifying an observation into the monozygotic twin group was increased. Also, both classification methods were found to be sensitive when the misclassification cost ratios exceeded 1:2. Least mean error rates were recorded for the QDF based on the misclassified observations while the MECM recorded high mean error rates, thus outperforming the MECM. Hence maximum separation between the two twin groups (monozygotic and Dizygotic) with equal group vectors assumed were provided by the Quadratic Discriminant Function (QDF) classification method.

## References

### Journal

- Adebanji, A.O., S. Adeyemi, and J.O.Iyaniwura. (2008). Effect of sample size Ratio on the Performance of the Linear Discriminate function: *International Journal of Modern Mathematics* 3 (1), 97-108.
- Fisher, R .A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7, 179-188.
- Ganeslingam, S., A. Nanthakumar, and S. Ganesh. (2006). A comparison of quadratic discriminant function with discriminant function based on absolute deviation from the mean. *Journal of Statistics and Management Studies*, 9:441-457.
- Geisser, S. and M. M. Desu. 1968. Predictive zero-mean uniform discrimination. *Biometrika*, 55, No. 3:519-524.
- Hyodo, M. and T. Kubokawa. 2014. A variable selection criterion for linear discriminant rule and its optimality in high dimensional and large sample data. *Journal of multivariate Analysis*, 123, 364-379.
- Lachenbruch, P. A. 1975. Zero-mean difference discrimination and the absolute linear discriminant function. *Biometrika*, 62(2):397-401.
- Marco, V. R., D. M. Young and D. W. Tubne. (1987). Asymptotic expansions and estimation of the expected error rate for equal-mean discrimination with uniform covariance structure. *Biometrika*, 29(1):103-111.

**Conference Proceedings**

Ariyo, O. S and A. O. Adebajji. 2010. Effect of Misclassification Costs on the Performance Functions. Paper Presented at the 45<sup>th</sup> Annual Conference of the Science Association of Nigeria (SAN) held Niger Delta University, Wilberforce, Island , Bayelsa State.

Bartlett, M. S. and N. W. Please. 1963. Discrimination in the case of zero mean differences. *Biometrika*, 50(1/2):17-21.

**Book**

Desu, M. M. and S. Geisser. 1973. Methods and applications of equal-mean discrimination. *Discriminant Analysis and Applications*, pages 139-161.

Johnson, R. A. and D. W. Wichern. 2007. Applied Multivariate Statistical Analysis, (Sixth ed). NJ: Pearson Education, Inc. 575-621.

Okamoto, M. 1961. Discrimination for variance matrices. *Osaka Math.*, 13:1-39.

Stocks, P. 1933. A biometric investigation of twins, Part II. *Ann. Eugen*, 5, 1-55.

## APPENDIX

Table 1a: likelihoods obtained from the density functions from each group

$f_1(M)$	$f_1(D)$	$f_2(D)$	$f_2(M)$
4.08E-05	1.14E-05	5.58E-06	1.68E-05
1.51E-05	4.09E-07	7.89E-07	3.32E-06
3.60E-05	5.24E-06	2.98E-05	2.82E-05
1.31E-05	1.13E-08	1.89E-07	1.36E-05
1.06E-05	1.83E-08	1.03E-08	5.01E-06
3.56E-05	6.10E-06	3.99E-06	4.40E-05
3.87E-05	2.23E-10	2.02E-09	1.27E-05
3.47E-05	1.59E-07	4.80E-07	6.31E-06
1.43E-09	4.40E-06	1.13E-05	7.51E-11
3.18E-05	1.12E-05	1.57E-05	4.00E-06
1.70E-05	6.41E-15	5.06E-12	1.07E-05
1.96E-05	2.38E-07	2.15E-06	9.19E-06
4.06E-05	2.76E-10	4.40E-08	1.11E-05
3.92E-05	4.06E-08	3.72E-08	2.84E-05
4.31E-05	6.33E-06	2.48E-05	4.29E-05

Table 2a: Scores obtained as a result of the classification rule from MECM under various costs.

$C(2 1): C(1 2)=$ 1:2		$C(2 1): C(1 2)=$ 1:3		$C(2 1): C(1 2)=1:4$	
$D_M$	$D_D$	$D_M$	$D_D$	$D_M$	$D_D$
1.21E+00	1.02E+00	<b>8.10E-01</b>	<b>6.81E-01</b>	<b>6.07E-01</b>	<b>5.11E-01</b>
2.27E+00	<b>2.59E-01</b>	1.52E+00	<b>1.73E-01</b>	1.14E+00	<b>1.30E-01</b>
<b>6.38E-01</b>	<b>8.79E-02</b>	<b>4.26E-01</b>	<b>5.86E-02</b>	<b>3.19E-01</b>	<b>4.40E-02</b>
<b>4.82E-01</b>	<b>2.99E-02</b>	<b>3.21E-01</b>	<b>1.99E-02</b>	<b>2.41E-01</b>	<b>1.49E-02</b>
1.06E+00	<b>8.88E-01</b>	<b>7.05E-01</b>	<b>5.92E-01</b>	<b>5.29E-01</b>	<b>4.44E-01</b>
<b>4.05E-01</b>	<b>7.64E-01</b>	<b>2.70E-01</b>	<b>5.10E-01</b>	<b>2.02E-01</b>	<b>3.82E-01</b>
1.52E+00	<b>5.52E-02</b>	1.02E+00	<b>3.68E-02</b>	<b>7.62E-01</b>	<b>2.76E-02</b>
2.75E+00	<b>1.66E-01</b>	1.83E+00	<b>1.10E-01</b>	1.37E+00	<b>8.28E-02</b>
9.52E+00	<b>1.95E-01</b>	6.35E+00	<b>1.30E-01</b>	4.76E+00	<b>9.73E-02</b>
3.98E+00	<b>3.57E-01</b>	2.65E+00	<b>2.38E-01</b>	1.99E+00	<b>1.78E-01</b>
<b>7.94E-01</b>	<b>6.33E-04</b>	<b>5.30E-01</b>	<b>4.22E-04</b>	<b>3.97E-01</b>	<b>3.17E-04</b>
1.07E+00	<b>5.53E-02</b>	<b>7.11E-01</b>	<b>3.69E-02</b>	<b>5.33E-01</b>	<b>2.77E-02</b>
1.83E+00	<b>3.14E-03</b>	1.22E+00	<b>2.09E-03</b>	<b>9.14E-01</b>	<b>1.57E-03</b>
<b>6.90E-01</b>	<b>5.46E-01</b>	<b>4.60E-01</b>	<b>3.64E-01</b>	<b>3.45E-01</b>	<b>2.73E-01</b>
<b>5.02E-01</b>	<b>1.28E-01</b>	<b>3.35E-01</b>	<b>8.51E-02</b>	<b>2.51E-01</b>	<b>6.38E-02</b>