KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

Performance Evaluation of Classification Methods: the case of Equal Mean Discrimination.

By

Michael Asamoah-Boaheng

(Bsc. Statistics)

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF MPHIL MATHEMATICAL STATISTICS

May 28, 2014

Declaration

I hereby declare that this submission is my own work towards the award of the MPhil degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgement had been made in the text.

Michael Asamoah-Boaheng	....	....
Student (PG8064712)	Signature	Date

Certified by:

Dr. (Mrs.) A.O. Adebanji	....	....
Supervisor	Signature	Date

Certified by:

Prof. S.K. Amponsah	....	....
Head of Department	Signature	Date

Dedication

To my dear Mum

(Rose Akosua Boateng)

and loved ones

Abstract

This study considered the equal mean discrimination problem by evaluating the performance of Bartlett and Please method, Bayesian Posterior Probability Approach, the Quadratic Discriminant Function (QDF) and the Absolute Euclidean Distance Classifier method (AEDC) under equal and unequal prior probabilities and non normality contamination. Stocks (1933) twin data recorded in London on 832 children based on ten selected measurements was used because it satisfies the assumptions of the equal mean populations. Four discriminant functions were derived and their error rate estimates determined using the Cross Validation (CV) and Balanced Error Rate (BER) methods. Results from equal prior probability showed the Bayesian Posterior Probability classifier performing better than the three other classifiers, thus it provides maximum separation with a recorded mean error rate of 0.149. Under the unequal prior probability situation, Bartlett and Please method outperformed both the QDF and the Bayesian posterior probability classifiers under the sampling ratios, 1:2, 1:3 and 1:4. For non-normality, all four classification methods recorded higher mean error rates indicating abysmal performance of the methods. However, Bartlett and Please method was found to be very sensitive to outliers. We recommend the Bayesian approach and the AEDC methods for classifying observations with equal prior probabilities, Bartlett and Please method under unequal prior probabilities and QDF for non-normal contamination with equal prior probabilities.

Acknowledgements

I am very grateful and highly indebted to the Most High God who is able to do abundantly and exceedingly beyond what i always ask for. I wish to express my appreciation to my lovely parents especially my mum "Rose Akosua Boateng"for their support and prayers towards my Education. My utmost appreciation goes to my supervisor, Dr (Mrs) A. O. Adebanji for the immense supervisory support i received from her. In fact i have learnt a lot from her constructive criticisms on my work. To all the lecturers and classmates in the department of Mathematics, KNUST, I say thank you for the knowledge you shared and imparted on me during my postgraduate studies. Last but not the least, i appreciate every effort of all my five siblings. God bless you all.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The problem of equal mean classification 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 comparing the performance of Bartlett and Please method, Bayesian rule classification method, the Quadratic Discriminant function (QDF) and the Absolute Euclidean Distance classifier method (AEDC) under equal and unequal prior probabilities and non normality contamination. This chapter takes a look at the background of the study, the problem statement, research questions and objectives, research methodology, justification of the study as well as the organisation of the study.

1.1 Background of the study

Discriminant analysis is a multivariate approach for identifying the features that separate known groups or populations. In other words, discrimination is a multivariate technique concerned with separating distinct sets of observations and it is exploratory in nature. (Johnson and Wichern, 2007).

The problem of discrimination was first initiated by Fisher (1936) in his paper titled "the use of multiple measurements in taxonomic problems"in which equal covariance matrices was assumed with or without normality assumption. Fisher's approach to classification with two populations was based on arriving at a linear classification function that gave maximum separation between groups without assuming normality.

Several investigations mainly with respect to multidimensional normal populations with common and unequal covariance matrices have been carried out by other authors. The equal mean discrimination problem has been one of the aspects of discriminant analysis that is still going through several investigations to uncover the best classification rule for the provision of better classification into one of the known groups. For the equal mean discrimination case, the optimal discriminant rule derived has to be based on the differences between the group-covariance matrices with some adopting uniform nature and some non-uniformity assumed. (McLachlan, 2004). Bartlett and Please (1963) were the first researchers to address this specific problem of zero-mean uniform discrimination. Following the work of Bartlett and Please, Bayesian analysis of the predictive zero mean discrimination was researched into by (Desu and Geisser,1968) for various assumptions involving the uniform covariance structure. They further discovered that whenever a linear predictive discriminant function was obtained, their associated errors of misclassification were also obtained as well. Some of the researchers who adopted the Bayesian approach in deriving their classification rules in the case of equal mean vectors include Geisser and Desu (l968), Desu and Geisser (1973), and Lee (1975).

McLachlan (2004) was in agreement of the fact that, the Bayesian approach to discriminant analysis is based on the concept of predictive densities of feature observational vector, where classification is based on the observations which give a higher estimate of the posterior probabilities. Similarly, the Bayesian approach based on the minimum estimates of the Expected Cost of Misclassification (ECM) was derived by Johnson and Wichern (2007) and the classification rule derived depended on the observation that gives the highest posterior probability. Lachenbruch (1975) and Ganeslingam et al. (2006) introduced outliers/noise into their working data to check the performance of the deduced functions. Some worked better after the contamination; some distorted the data causing poor performance whilst in some situations the functions were not perturbed after the introduction of the outliers. According to Sharma (1996), to classify a given observation, first the classification function for each group are used to compute the classification scores and the observation is assigned to the group that has the highest classification score. The posterior probability of an observation belonging to a given group can also be computed. The observation is assigned to the group with the highest posterior probability. In evaluating the classification functions, one of the ways of judging the performance of several classification procedures is to calculate their error rates or misclassification probabilities. Some of the error rates are the Optimal Error Rates (OER), the Apparent Error Rate (APER), the Balanced Error Rate (BER) and the Leaving-One-Out method (LOO). (Johnson and Wichern, 2007). The performance of the various discriminant functions depend mainly on the error estimates using some of the error rate estimators such as OEA, APER, LOO or cross-validation etc. The function or the classification rule with the minimum estimates of the rates or mean rates as a result of the above listed estimators gives the best function for classification. This confirms what Johnson and Wichern (2007) stated generally that a good classification procedure should always result in few misclassifications. That is the probabilities of misclassification should be small or minimum. The discrimination procedure still remains a problem, since there is still more room for researchers to come up with classification rules that will give maximum separation among the groups. In the equal mean case, the researcher looks at the problem of discriminating between two known groups on assumption that, the mean vectors are assumed to be equal. What therefore happens and which parameters can be use to provide maximum separation in the case of the above listed assumptions?

1.2 Statement of the Problem

The problem of discrimination/discriminant analysis has always been based on developing a reliable discriminant function whether standardised or unstandardised and classifying the observations into their respective known populations/groups. The problem of statistical discrimination involving two multivariate normal distributions with equal means and with equal (or unequal) covariance matrices has been considered by many researchers. Several works have focused on discriminating between two groups with equal and unequal covariance matrices, with an assumption of equal mean difference between the two groups. Some researchers including Okamoto (1961), Bartlett and Please (1963), have focused on addressing the problem of discrimination with common mean vectors among the populations. Desu and Geisser (1973) also in furtherance to the work of the above mentioned researchers tailored their work based on the situation where several methods including the Bayesian approach, the classical approach and other methods were used to analyse stocks twin data in case of the equal means or zero mean vector without specific assumption about the distributions. They further addressed the problem with reference of assuming normal distribution with a common mean vector with differing covariance matrices from the classical viewpoint approach. Lanchenbruch (1975) also came out with a technique of tackling the problem of discrimination by taking the absolute values of the observations of the discriminating variables and solved the problem using what he termed as Absolute linear Discriminant analysis. Ganeslingam et al. (2006) also researched into equal mean case by comparing the QDF and that of an absolute Euclidean distance classifier (AEDC) with an application to a three dimensional twin data. However all these researchers were able to show that, even though the mean vectors were equal across the groups, other distributional properties were able to successfully discriminate or classify each observation into their specified groups. Notwithstanding the work of these researchers,the effect of unequal prior probabilities for the known groups/populations on the separation of observations based on the discriminant functions and their classification rules are yet to be investigated extensively to compare their performance with functions developed under equal groups prior probabilities.

Also limited application of the AEDC has been highlighted by some researchers like Ganeslingam et al. (2006). But their study was only applied to a three dimensional data structure. This study will apply the AEDC on the ten dimensional structure of Stocks twin data and compare its classification rule with other classification rules based on their estimated error rates.

Several studies in discrimination have overlooked at the situation where the Mahalanobis distance is used as a classifier in case of unequal covariance matrices as well as equal mean vector situation. Another problem also shows how in some cases, discriminant functions are affected by outliers whilst others do not. Not much extensive work/ studies have been conducted in the equal mean case when the data are always contaminated with outliers. However more work has to be done to come out with the best discriminating function or equation which will further help in solving the problem of discrimination with equal mean assumption.

1.3 Objectives of the study

The main objective of the study is to compare the performance of the discriminant functions based on unequal group sizes and contaminated data.

1.3.1 Specific objectives

1. To obtain discriminant functions for the equal mean case based on the Bartlett and Please approach, the Bayesian approach, the QDF and that of the AEDC approach.

2. To derive discriminant functions as well as their classification rules under unequal prior probabilities and under non-normality(contaminated data).

3. To compare the performance of all the four discriminant classification rules by using the various estimated error rates.

1.4 Methodology

This study which is aimed at comparing and applying methods of discriminant functions with contaminated data and with unequal population size was approached with several methods. In the case of equal mean, the Bartlett and Please (1963) approach, the Bayesian approach, the Quadratic discriminant function approach and that of the AEDC were employed. In other words, four discriminant functions were evaluated based on the above listed methods. Secondly, the sizes of the two groups were taken differently at several times in multiples of five, and the four discriminating methods were derived from the unequal population sizes situation. Lastly, two estimators of error rates were used to estimate the mean error rates of the four discriminant functions. The error estimators include the cross-validation (CV) method and the (BER). Stocks (1933) twin data on recorded measurements as well as age and sex on 832 children in Elementary Schools in London was used.

As used by Bartlett and Please (1963), 30 pairs of female twins with ten selected measured characteristics were taken from the Stocks twin data (1933), with 15 pairs belonging to both the Monozygotic and Dizygotic twin groups. The first four discriminant functions were obtained from the 30 pairs of twin data. This was obtained after 10 replications through simple random sampling technique without replacement. In the second case, the sample sizes were taken differently by multiples of five and four discriminant functions were obtained. In the third case, the equal population/group size data was contaminated and in this case also the four functions were obtained. The discriminating functions in each of the three cases were compared to ensure their performance based the error estimates.

1.4.1 Data

The analyses was based on the Percy Stocks (1933) twin data on recorded measurements as well as the age and sex of 832 children in the Elementary and central schools of London during the period 1925 to 1927. He recorded fourteen (14) measurements on each of the children.

1.4.2 Analytical software used

The researcher used R-console version 2.15.1 to write the numerous functions for the development of the discriminant functions. The MINITAB version 14 statistical software was used to generate the scatter plots as well as other plots in this study.

1.5 Justification of Work

This study will contribute immensely to knowledge since equal mean discrimination is one of the interesting aspects of discriminant analysis. The application of the AEDC method for a larger dimensional data set will also give an insight to researchers on the reliability of the criterion when applied to large data set. The study will serve as literature for future researchers since there are always limited literatures when it comes to discriminating with equal mean vectors. The study would therefore provide the needed statistical evidence to justify the best classification rule based on the derived discriminant functions.

1.6 Limitations of the study

This study would have been much more encompassing if the various methods were applied in cases where the population/groups are more than two. This was due to data availability constraint. The study also encountered the problem of limited literature where few literatures obtained were directly linked to the problem of equal mean discrimination, hence making it difficult to do much more extensive work on the research topic.

1.7 Organisation of the study

This thesis is organised in five chapters. Chapter one is the introductory chapter to the entire study. It takes a look at the general background of discriminant analysis and classification and narrowed down to the equal mean discrimination situation. The problem statement, research questions and objectives, research methodology, justification of the study are presented in this chapter. Chapter two reviews related literature (from articles and books) based on the thesis objectives and preferred models to be used in achieving these objectives. Expected outcome of the study and other comparative results of similar studies are also discussed in this chapter. Chapter three describe the theory of history model to be used, formulations and methods of solution. Chapter four is dedicated to data collection, analysis and results. Chapter five concludes the entire study by stating specific recommendations to stakeholders based on the major findings made in the study.

Chapter 2

Literature Review

This chapter discusses the literature available on discrimination in case of equal mean population vectors. It also looks at summary of abstracts on various literatures with regard to the model being used and the general working title. The Chapter comprises both empirical and theoretical literature.

2.1 Theoretical Literature

This part of the literature considers the theories and definitions of the various terms and concept in discriminant analysis in general as well as a review of related studies. Discriminant analysis is a multivariate approach for identifying the features that separate known groups or populations. In other words, discrimination is a multivariate technique concerned with separating distinct sets of observations and it's exploratory in nature. Classification on the other hand, allocates new objects and observations to previously observed group based on the allocatory rule deduced from discriminant functions which provides maximum separation. Classification is a term that is normally seen as less exploratory since it leads to well-defined rule which are normally used for assigning new observations into one of the several known groups. Classification rule is based on features that separate the groups and sometimes the two terms (discrimination and classification) overlap. Classification involves more problem structure than that of discrimination. The terminology was first introduced by R. A Fisher in his study of first modern treatment of separative problems. (Johnson and Wichern, 2007). Discriminant analysis is based on the three objectives;

1. By identifying the variables that discriminate best between the two or more groups.

2. The identified variables are then used to develop an equation or a function that is used for computing new index that will represent the difference between the two or more known groups parsimoniously.

3. A classificatory rule is developed by the usage of the identified variables and the computed index to classify future observations into one of the two or more known groups. (Sharma (1996).

According to Johnson and Wichern (2007), Fishers approach to classification with two populations is 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. He continued by taking the linear combinations of 'x' to create 'y's because they can be handed easily since they have simple enough functions. Fisher assumes equal variance covariance matrices since a pooled estimate of a common covariance matrix is used, but did not assume the distribution of the populations to be necessarily normal. He further stated emphatically that, classification rules become more complicated when the population covariance matrices are unequal which leads to quadratic discriminant function classification rule and becomes awkward in more than two populations. Generally, a good classification procedure should always result in few misclassifications. That is the probabilities of misclassification should be small. Some of the additional features that an optimal classification rule possesses include; one population having a greater likelihood of occurrence than the other group, because one of the two populations has a greater likelihood of occurrence. Johnson and Wichern further derived several methods of classification. These includes; classification with two multivariate normal population, classification with normal populations with equal covariance matrices, the estimated minimum expected cost of misclassification rule for two normal populations, classification of normal populations when the covariance matrices are unequal (quadratic classification rule).

In evaluating the classification functions, one of the ways of judging the performance of several classification procedures is to calculate their error rates or misclassification probabilities. Some of the error rates includes the optimal error rates (OER), the apparent error rate (APER), the balanced error rate (BER) and the leaving-one-out method (LOO). (Johnson and Wichern, 2007).

2.2 Equal mean discrimination

Applications of discriminant analysis in which the group means are assumed to have the same mean vectors have been researched by various academicians. In the case for multivariate normal group conditional distributions, the optimal discriminant rule has to be based on the differences between the group-covariance matrices with some adopting uniform nature and some non-uniform depending on the nature of their research. (McLachlan, 2004). Most research conducted into this problem focuses mostly on two groups or populations. This section outlines the various studies conducted which involves the problem of discriminating among two populations when the means of the two populations are equal. The section presents the various articles and papers in line with equal mean discrimination conducted by several researchers.

Okamoto (1961) first studied the problem of discrimination with common mean vector with different covariance structure of two multidimensional normal populations. In his paper of developing the theory for discrimination, a biometrical application was used. As at 1961, Okamota assumed that, the reason why the problem of equal mean vector for discrimination has not been tackled before might have been as a result of the complex nature of the theory and the scantiness of its application to the equal mean vectors. The two populations with a known common mean as well as the covariance matrices were specified. Further assumptions were made in the case where the populations were not completely specified. Also information provided by the random sample for the unknown parameters were taken from each population. The theories and methods were then applied to a twin data where estimates of Eigen vectors for minimax discrimination function method and that of the Bayesian discrimination function case were obtained. Examples of this application includes the following; image processing, process control with same but different variation patterns and fluid dynamics.

Bartlett and Please (1963) widely investigated the problem of discrimination when the common mean vector is a zero vector. They investigated this problem for the case when the variance covariance matrices for the two groups are known to have uniform structures, with an assumption of equal and unequal correlation coefficients (ρ). A discriminant function based on equal correlation coefficient as well as uniform covariance matrix was obtained for classificatory purposes. The discriminant function was obtained for the case of the multivariate normal populations with the same variance covariance matrix by using the likelihood ratio with the omission of the additive constant. However the best boundary for equal risk of misclassification was derived to serve as a cut-off point for assigning observations into one of the two known groups. The above discriminating methods was applied to the Stocks twin data for both males and females, where the estimates of the covariance matrices were obtained as a well as their respective values. Assuming common correlation coefficient between the variates, their estimates for males and females were obtained and a linear discriminant function was obtained for each of the sexes. The discriminant functions obtained, was able to misclassify four (4) observations from Monozygotic group and seven (7) from the Dizygotic for the males and among the females, no observation from the Monozygotic group was misclassified, but 2 individual were misclassified from the Dizygotic population. They concluded that, discrimination is much better for female liked sex twins with low values/estimate of the correlation coefficient (ρ).

Geisser and Desu (1968) in their study of "Predictive zero-mean uniform discrimination" developed a predictive Bayesian approach to the problem of assigning an individual to one of the two multivariate normal populations, on an assumption of a null mean vector for both populations and a uniform covariance structure as derived and used by Bartlett and Please (1963). They obtained a Bayesian analysis of the problem of zero-mean vector for the populations, and also presented linear predictive discriminant functions and their associated errors of misclassification. These approaches were applied to the Stocks twin data on physical measurements/characteristics of female twins and obtained linear discriminant functions as well as their corresponding results under the Bayesian approach with their predictive densities. Two classification rules were obtained under an assumption that the correlation coefficients for the two groups are unequal ($\rho_1 \neq \rho_2$). Both rules classified the data in a similar manner, with no apparent difference in the discriminatory power with two individuals being misclassified from the Monozygotic group and only one from the Dizygotic group. The data was further analysed with an assumption of equal correlation coefficient for the two groups ($\rho_1 = \rho_2$) as used by Bartlett and Please (1963). A linear discriminant function was obtained under this assumption, but no results for the Bayesian approach were obtained, hence comparison about the discriminatory power not possible. However, two observations/individuals from the Monozygotic group were misclassified under this rule.

Desu and Geisser (1973) studied the 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 following methods in their analysis; the use of classical approach, semi Bayesian and complete Bayesian approach, Discrimination with equal and unequal covariance matrices. They observed that, both the classical approach by Bartlett and Please method and the Semi-Bayesian approach gives the same rule and both rules were able to misclassify only one of the additional five observations in the Monozygotic group which was not part of the ten (10) which were used for developing the discriminant function. Similarly the same classification rule also further misclassified only one of the five observations which were not included in developing the discriminant function in the Dizygotic group under the assumption that, the estimated correlation coefficients for the two groups are unequal. Also, Bartlett and Please (1963) rule for the case when the common mean vector is zero vector and with uniform covariance matrices was used by Geisser and Desu. Complete Bayesian approach used by Desu and Geisser (1973) for equal mean discrimination under the general covariance matrix, recorded correct classification of all the twenty observations from both groups which were used in obtaining the discriminant function. However the remaining ten which were excluded in developing the function, misclassified three observations from the Monozygotic group whilst none where misclassified in the Dizygotic group. They concluded from their study that, the pattern of misclassification shows that, discriminating with uniform covariance structure is more appropriate for the stocks twin data than the arbitrary covariance matrix, although both structures misclassified the same number of individuals.

Lachenbruch (1975) went through some long process to derive 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. His new method of taking the absolute values of deviation from the mean aided in avoiding writing long programs for calculating a quadratic function and to protect against long term contamination existence in the data. He observed that the absolute linear discriminant function (ALDF) was almost as good as the quadratic discriminant function (QDF) when the two groups are closer. Lachenbruch applied the methods outlined above to the well-known stocks data with ten measured variables. In his research, he developed a criterion for variable selection where the variable with the minimum error was considered to give the best classification. By this criterion, the following variable including their height, head circumference and interpupillary distance were the variables to give better classification, but does better when combined with other variables. Three discriminant functions were used based on the variables, including the one with the full set of the ten variables. The apparent error rate (APER) was used to assess the performance of the functions as well as the cross validation method. The APER gave worse bias for the QDF than that of the linear functions based on the large number of estimators. The LOOM on the other hand provided a less biased estimate of the error rates. A six variable function was preferable to the full ten variable discriminant function obtained. The case of ALDF saw the three variable rule performing as well as the six variable QDF rule performs. In ensuring performance, the ALDF performed slightly poorer than that of the QDF for the six-variable case and a bit better for the three-variable case. Lachenbruch further contaminated the data and realized that, there exist sequential biases of the APER for QDF. An increase in the mean error rate using the LOOM when compared with the uncontaminated data was observed. The increased mean error rates after the contamination for the ALDF was quiet lower than that of the QDF for the three-variable and the six-variable case but slightly higher in the 10-variable case. In concluding, Lachenbruch (1975) found out that, his 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. (1987a) in their paper titled "asymptotic expansion and estimation of expected error rate for equal mean discrimination with uniform covariance structure", 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. The asymptotic expansion for the unconditional expected error rate was then compared with the Monte Carlo simulation evaluated at several combinations of the parameters to ensure accuracy of the approximation of the error rate. In furtherance to their work, a deduction of an evaluation of the accuracy of the expansion was made, and was done by comparing the values of the expansion and the Monte Carlo estimates of the expected error rates. Their results showed that, the first order asymptotic expansion for the expected error rate gives an excellent agreement with the Monte Carlo estimates. Furthermore, the researchers realized that, the expected error rates for the two estimators performed poorly when and n=15. The discrepancy arose when with a moderately small sample size and the results yielded estimates of rho (ρ) exceeding one (1). This also affected the classification algorithm to perform quiet poorly, and thereby increased the probability of misclassification disproportionally. However, the asymptotic expansion of the error rate for the equal mean, uniform-covariance-matrix yielded reasonably accurate approximations with the problem of estimating when . The derived estimators for the error rate and that of the variance were applied to the classical example of the well-known Stocks data, and they concluded that the equal-mean classification algorithm applied to the Stocks data appears to yield excellent results.

McFarland and Richards (2001) continued the work of Marco et al. (1987), and they came out with an idea that, there is the need for an investigation of the exact distributions of the plug-in discriminant functions for the equal mean vectors with uniform covariance matrix. They therefore developed four normal quadratic discriminant functions, Monte Carlo simulations of the exact distribution function from the new stochastic representation and the two methods were observed to be more efficient than a direct Monte Carlo simulation of the discriminant statistics itself. A stochastic representation for the exact distribution of the four discriminant functions was derived after providing preliminaries relating to the multivariate gamma functions and Bassel functions of the matrix argument. The stochastic representation was then applied to the discriminant functions to study the corresponding probability of misclassification. McFarland and Richard then further applied the results obtained to the Stocks twin data and Rencher's data on football players head size to estimate their respective misclassification probabilities. Two of the four discriminant functions were derived from minimum distance for equal mean case. The minimum distance discriminant functions provided the highest estimate of the probability of misclassifying an observation which belongs to the Monozygotic population. However the application of the methods to the Rencher's data estimated the probability of misclassifying observation to Monozygotic and that of the Dizygotic population to be relatively closed as expected.

Hosseini and Armacost (1992) presented a study on two group discriminant problem with equal group mean vectors with several methods and mathematical formulations. The researcher specifically focused on the performance of a broad mathematical technique with respect to the equal mean vector discriminant problem. The study employed the Monte Carlo simulation experiment as well as two linear programming (LP) formulations and four non LP models. For comparative purposes, both Fishers linear discriminant function (FLDF) and that of Quadratic Discriminant function (QDF) were used. The shapes of the population distributions and the variance covariance matrices as well as the availability of the unusual noise in the data were the factors manipulated. Misclassification ratios were used to judge the performance of the various models under various factor levels. The first section of the methodology was based on the description of the discriminant analysis model and the description of the Monte Carlo simulation experiment. The presentation of the analysis was done in the second section and it was expectant that, both the QDF and that of the FLDF will show a poor performance under equal population mean vectors as compared to that of the LP methods. Applying the methods to some practical data, yielded the following outcomes. Almost majority of the methods performed better in the case of multivariate non-normal distributions than compared to that of the one generated from a multivariate normal distribution. All the various discriminatory methods performed better generally when the covariance matrices for the two populations were assumed to be unequal. Also, less favorable performance was observed for FLDF, QDF with presence of outliers than when there is absence of outliers/noise. Hosseini further discovered that, the property of equality of variance-covariance matrices had a higher significant effect on misclassification rates than the other properties. In concluding, the involvement of outliers in the data did not influence the performance of the methods significantly.

Young et al. (1988) looked at the robustness of the equal mean discrimination rule with uniform covariance matrix using a serially correlated training data. Young and friends focused their study on investigating the effect of the serially correlated data on an estimated expected error rate with the involvement of the equal mean classifier with an assumption of a uniform covariance structure. After their application to the serially correlated data, it was realized that, the expected error rates was quite minimal thereby reducing the number of misclassifications of observations into one of the two groups using the equal mean classifier.

Lee (1975) also did some studies on the problem of discrimination involving the equal mean case, where his study was titled "a note on equal mean discrimination". The main method employed for developing the discriminating function and for classification in the equal mean case was based on the Bayesian approach to ensure minimizing the loss incurred in misclassifying an observational vector from one of the two populations.

Schwemer and Mickey (1980) conducted a study on a note on the linear discriminant function when group means are equal. They therefore derived the expected error rates of the linear discriminant rule obtained and their application to the data was made using two groups with the same means vectors but proportional covariance matrices.

The Journal of Statistics and Management Studies published a paper in 2006 titled, "comparison of quadratic discriminant function with discriminant function based on the absolute deviation from the mean" by Ganeslingam et al. (2006). The problem of statistical discrimination involving two multivariate normal distributions with equal means and different covariance matrices was considered. They therefore employed the method of Absolute Euclidean Distance Classifier (AEDC) which is derived from the absolute values of the components of the observation vectors used in Euclidean Distance Classifier (EDC). Hence it was expectant that, it might do well with the situation of high dimensional setting. Their main objective for their study was to compare the performance of the AEDC and QDF in case of equal population mean vectors with unequal covariance matrices. They discovered the AEDC and that of the QDF as an alternative to that of the Fishers linear discriminant function (FLDF) in situations where the population means among the groups are the same and with an assumption of unequal covariance matrices. AEDC forms a discriminant function basically on the absolute deviation, where the information about the covariance matrices is always ignored. The objective was to compare the two functions using the error rates obtained from the leaving-one-out (LOO) method and that of the apparent error rate (AER). Theoretical derivation of the distributional function of the vector of absolute values in the trivate normal populations was given. Allocation rule was derived using the absolute deviations. The methods were then carried on 89 pairs of male twins with 49 being Monozygotic and 40 being Dizygotic with six variables for each pair of twins from anthropology study conducted in Germany and reported by Flurry (1997). Three variables were used after taking the difference between the first and the second twin and hence making the study focused on a three dimensional problem. The following result was realized after the application of the AEDC and QDF to the twin data. The AEDC outperformed the QDF in 89 of the 99 cases after a Monte Carlo simulation study was carried out for the bivariate case. The AEDC outperformed the QDF in the case of the contaminated data with obviously higher estimates of variance than that of the uncontaminated case. The overall actual error rate of the QDF was found to be higher than that of the AEDC by 3.5 percent. The cross validation estimates gave lower overall error rates in comparison to the estimates of the actual error rates. Further, they discovered that, the QDF are very useful in cases where the covariance matrices are nonsingular and this causes its inferior performance to that of the AEDC in higher dimensions. They concluded that, the AEDC is highly useful for the case of equal mean discriminating problem for two populations with different covariance matrices.

Earlier, the study conducted by Raudy's and Pikelis (1980) as cited in Ganeslingan et al. (2006) aimed at the comparison of the performance of the EDC and that of the LDF when p is relatively large and they concluded that the sample EDC outperformed the sample LDF when classifying individuals from two spherical and non-spherical normal populations and covariance structure.

Marco et al. (1987b) as cited in the work of Ganeslingan (2006) investigated discrimination problem under situations in which the EDC becomes a Bayes classifier in terms of the error rates. They suggested that the sample EDC performs better than the sample LDF in many practical situations, since the EDC involves only fewer estimated parameters.

2.3 Bayesian discrimination and Classification

Many researchers in the already reviewed papers in the above section have used the Bayesian method one way or the other in solving the equal mean discrimination problem. Geisser (l964), Geisser and Desu (1968), Desu and Geisser (1973), and Lee (1975) are some of the researchers who investigated the problem of equal mean vector discrimination using Bayesian approach where some was based on the predictive densities whilst others on their posterior probabilities. This section focuses not only on application of Bayesian method in equal mean discrimination, but rather on studies conducted generally on discriminant analysis with application of Bayesian approach as well as Mahalanobis distance for classification purposes.

The concept of the Bayesian approach to discriminant analysis is based on the concept of predictive densities of feature observational vector. The predictive density of observational vector x within a particular group was derived, including the likelihood function and some weighting function. The posterior probability was obtained based on the likelihood as well as the group's prior probabilities. Classifying an observation into one of the known groups depends on the one with the highest posterior probability value. (McLachlan, 2004).

Similarly, Johnson and Wichern, 2007 derived the Bayesian approach to classifying an observation to one of the known groups as based on the minimum ECM rule with equal misclassification cost with respect to their density/likelihood functions as well as the group's prior probabilities. The posterior probability was given as the ratio of the product of the density functions and the groups prior probability over the total probability involving the prior and the density functions of observations. The classification rule was identical to the observation that maximizes the posterior probability.

Rigby (1997) undertook a study aimed at solving the problem of discrimination using the Bayesian discrimination between two multivariate normal populations with equal covariance matrices. Rigby developed an estimate for the Bayesian and classical estimates of the probability that a new observation belongs to one of the two multivariate normal populations with an assumption of equal variance covariance matrices. The researcher compared the Bayesian and the classical approach of the probability P that an observation came from population given that, it is in that same population. Upon realization, the Bayesian estimates P_B is easy to calculate and their estimates are superior to the classical estimate Pc which sometimes provides extreme estimates of P in a transformed space. Contours of the estimates of P_B and Pc were obtained and he highlighted the case in which the Bayesian and the Classical estimates and their respective allocation rule differs. As a result, the Posterior distribution of P and its interval were found. The methods after its application resulted in the following outcomes. There were differing outcomes based on the estimates of Bayesian and that of the classical method when they are in a state of dimensionally reduced transformed space. The Bayesian and the classical estimated allocation rule differs hugely when their respective constants in turn differs greatly from the prior probability that an observation belongs to the first population. It was highly recommended that the Bayesian estimates leads to the provision of less extreme and a more reliable estimates of P.

Lavine and West (1992) discussed the Bayesian analysis of the traditional normal-mixture model for classification and discrimination. They adopted the application of an iterative re-sampling approach to Monte Carlo inference. They therefore performed the discrimination and classification with several normal-mixture components having different variance covariance matrices. The exact posterior classification probabilities for the observed data and that of the classification of future observations as well as the discriminant functions based on the posterior probabilities were computed. The largest among the three posterior probabilities was identified and corresponded to components 2 and 1 respectively.

2.4 Non-Normality

Several studies have been conducted on the effects of non-normality on classification rules based on the some well known methods such as the Quadratic Discriminant Function (QDF) and the Linear Discriminant Function (LDF). Lachenbruch et al. (1973) considered the robustness of LDF. They considered three specific distributions and the case of independent variables. These distributions were considered to be non-normal and were generated from the normal distributions using the Johnson system of transformations (i.e log normal, inverse hyperbolic sine normal and logit normal distribution). They observed considerable decline in performance of the LDF (the log normal distribution used had extremely large skewness and kurtosis). They conducted Monte Carlo experiments to investigate the robustness when parameters are estimated. Based on their results, Fisher's LDF was greatly affected by non-normality in the population. It was also noted that, if the error rates with LDF were greatly different in the two populations when cut-off point is zero, then presence of non-normality is indicated. They concluded that, the use of Fisher's LDF under non-normality contamination situations could be badly misleading and recommended that the data be transformed to approximate normality prior to the use of the LDF.

Lachenbruch et al. (1977) investigated the effects of non-normality on the QDF. They assumed that the data were transformable to normality. After the transformation, the variables were found to be independent with proportional covariance matrices. They generated random samples from non-normal distributions in order to study the effect of non-normality on QDF. The samples were transformed into components by using Johnson's system of transformation. Their results indicated that the actual error rates were considerably larger than the optimal rates in the case of zero mean difference. Also, non-normal samples generally under the QDF did not do substantially worse than when applied to normal samples.

2.5 Error Rates

An assessment of error rate estimators was studied by Krzanowski and Hand (1997) paying special attention to the leave-one-out method. The leave-one-out rule seeks to overcome the drawback of re-substitution by process of cross-validation. The estimator was investigated in simulation study, both in absolute terms and in comparison with a popular bootstrap estimator. From their results, Bayes procedure was found to give unreliable estimates of leave-out-two which performed better than the leave-one-out method. They found that results of the leave-one-out method was even worse than expected. By extension of the leave-one-out method, the leave-out-two was looked at, by considering their variances. As a result, they observed a slight variance reduction relative to the leave-one-out method.

2.6 Other related studies on discrimination

This section will present summaries of articles on discrimination in general. Some of the reviewed paper as presented below.

Noh et al. (2010) studied fluid dynamics model for low rank discriminant analysis. They considered the problem by reducing the dimension of the labelled data for classification. The data was considered as an interacting fluid in the high dimensional space, derived from the Bhattacharyya coefficient, which has shown to be closely related to the Bayes classification error. The main aim of their study was to find low dimensional subspaces, where classification is better as compared to other subspaces with its classification error being minimized. The projected distributions were found to be the Gaussian/normal with means and covariance matrices with known parameters of the dimensions of the original space and that of the projected space. A model was obtained for dimensionality reduction. The methods were evaluated at several datasets. Fishers Discriminant analyses (FDA) and Fukunaga's equal mean analysis were used to project the data onto the subspaces found. The Bayes classification was performed on the subspaces with an assumption of a normal/Gaussian distribution. The results showed that, the fluid analysis slightly outperformed the FDA and the Fukunaga's equal mean analysis.

Sajobi et al. (2010) aimed at investigating the effects of the repeated measures mean and its covariance structure of bias and mean square error in the discriminant function coefficients for procedures of discriminant analysis based on parsimonious covariance structures and the mean. First and foremost, the Discriminant functions coefficients (DFCs) in DA procedures were estimated for repeated measures data involving two populations/groups. The misspecification of the repeated measures means and the covariance structure resulted in an increase in the in the bias and errors of DFCs of the investigated procedures. The average bias and the root mean square error caused by the misspecified repeated measures was found to be greater than the bias and the error variation caused by the misspecification of repeated measure covariance structure for all the investigated procedures. Negligible biases were observed from the DA based on the parsimonious repeated measure mean and covariance structure. After sampling the data from the population with non-constant mean configuration, it was realized that, the DA had the smallest bias. They therefore suggested a DA procedure based on unstructured mean vectors and covariance matrices should be recommended.

Rausch and Kelley (2009) investigated the comparison of linear and mixture models for Discriminant analysis under non-normality assumption. Various methods of discriminant analysis derived were compared through the employment of Monte Carlo simulation method to ensure accuracy among the methods under non-normality assumption. The following DA methods were obtained; linear discriminant analysis based on the raw scores and on rank scores, logistic regression and a mixture of discriminant analysis. The highest rate of classification accuracy was produced by the linear discriminant analysis based on the ranks of the scores. However they failed to produce a practical important advantage over the other competing methods. The method with relatively small number of components found in each group attained the highest rate of classification accuracy and this was very useful to conditions with skewed independent variables having small values of kurtosis.

Chapter 3

Methodology

3.1 Discrimination and Classification

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. Classification procedures on the other hand are less exploratory in the sense that, they lead to well-defined rules, which can be used for assigning new objects. The goals thus are; (1)To describe (graphically or algebraically) the difference between objects from several known populations. We construct "discriminants" that have numerical values which separate the different collections as much as possible; (2) To assign objects into several labelled classes. We derive a "classification" rule that can be used to assign (new) objects to one of the labelled classes. Discriminant analysis is used in situations where the clusters are known a priori. The aim of discriminant analysis is to classify an observation, or several observations, into these known groups. The first goal corresponds to the term discrimination, in which the terminology was introduced by R.A. Fisher in the first modern treatment of separative problems. A More descriptive term for the first goal, however is 'separation' and the second goal referred to as classification or allocation. A function that separates objects may sometimes serve as an allocator, and conversely, a rule that allocates objects may suggest a discriminatory procedure. However, both terms, discrimination and classification frequently overlap and the distinction between separation and allocation becomes blurred.

A good classification procedure should result in few misclassifications. In other words, the chances, or probabilities, of misclassification should be small. Also there are additional features that an optimal classification rule should possess.(John son and Wichern, 2007). Ḧardel and Simar (2007) used the example in credit scoring. Where in credit scoring, a bank knows from past experience that there are good customers (who repay their loan without any problems) and bad customers (who showed difficulties in repaying their loan). When a new customer asks for a loan, the bank has to decide whether or not to give the loan. The past record of the bank provides two data sets: multivariate observations X_i on the two categories of customers (including for example age, salary, marital status,the amount of the loan, etc.). The new customer is a new observation X with the same variables. The discrimination rule has to classify the customer into one of the two existing groups and the discriminant analysis should evaluate the risk of a possible "bad decision".

3.2 Discrimination and Classification of Two Populations

This section focusses on separating objects from two classes and assigning new objects to one of these two classes. The classes will be labelled π_1 and π_2. Each object consists of measurements for p random variables, $X_1, ..., X_p$ such that the observed values differ to some extend from one class to the other. The distributions associated with both populations will be described by their density functions f_1 and f_2 respectively. Now consider an observed value $x = (x, ..., x)^\tau$. Then $f_1(x)$ is the density in x if x belongs to π_1. And $f_2(x)$ is the density in x if x belongs to population π_2. The object x must be assigned to either population π_1 or π_2. Denote Ω, the sample space (collection of all possible outcomes of X) and partition the sample space as Ω=R_1 U R_2 where R_1 is the subspace of outcomes which we classify as belonging to population π_1 and R_2= Ω-R_1 the subspace of outcomes classified as belonging to π_2. It follows therefore that the (conditional) probability of classifying an object as belonging to π_2 when it really comes from π_1 equals

$$P(2|1) = P\left(X \in R_2|X \in \pi_1\right) = \int_{R_2} f_1(x)dx \tag{3.1}$$

and the (conditional) probability of assigning an object to π_1 when it actually comes from π_2 is given as;

$$P(1|2) = P\left(X \in R_1|X \in \pi_2\right) = \int_{R_1} f_2(x)dx. \tag{3.2}$$

The conditional probabilities can also be obtained for $P(1|1)$ and $P(2|2)$.

Prior class probabilities are obtained when we want to obtain the probability of correctly and incorrectly classifying an observation/objects. We denote the following;

$p_1 = (X \in \pi_1)$=the prior probability of π_1 and $p_2 = (X \in \pi_2)$ being the prior probability of π_2 in which $p_1 + p_2$=1.

Following the above, the overall probabilities of correctly and incorrectly classifying an object can be estimated and some of them are given below;

P(Object is correctly classified as π_i)= $P(X \in R_i|X \in \pi_i)P(X \in \pi_i) = P(i|i)p_i$ where $i = 1, 2$.

P(Object is misclassified as π_i)= $P(X \in R_i|X \in \pi_j)P(X \in \pi_j) = P(i|j)p_j$, where $i \neq j$.

3.2.1 Cost of Misclassification

In considering the cost of misclassification, denote $c(i|j)$= The cost of classifying an object from π_j as π_i.

A classification rule is obtained by minimizing the Expected Cost of Misclassification (ECM). It is therefore derived by;

$$ECM := c(2|1)P(2|1)p_1 + c(1|2)P(1|2)p_2 \tag{3.3}$$

To minimize the expected cost of misclassifications, the regions R_1 and R_2 are used and they are given by the following;

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

and

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

Note that the two regions depends on the following ratios and they are often much easier to determine than the exact values of the components.

$\frac{f_1(x)}{f_2(x)}$ = Density ratio

$\left(\frac{c(1|2)}{c(2|1)}\right)$ = Cost ratio

$\left(\frac{p_2}{p_1}\right)$= Prior probability ratio

Generally we assign x_0 to π_1 if

$$\frac{f_1(x_0)}{f_2(x_0)} \geq \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right) \tag{3.4}$$

From the above equation, the following special cases for classification can be deduced;

1. Equal (or unknown) prior probabilities: Compare density ratio with cost ratio;

   R_1: $\frac{f_1(x)}{f_2(x)} \geq \frac{c(1|2)}{c(2|1)}$ and R_2: $\frac{f_1(x)}{f_2(x)} < \frac{c(1|2)}{c(2|1)}$

2. Equal (or undetermined) misclassification cost: Compare density ratio with prior probability ratio:

R_1: $\dfrac{f_1(x)}{f_2(x)} \geq \dfrac{p_2}{p_1}$ and R_2: $\dfrac{f_1(x)}{f_2(x)} < \dfrac{p_2}{p_1}$

3. Equal prior probabilities and equal misclassification cost or $\dfrac{p_2}{p_1} = \left(\dfrac{c(1|2)}{c(2|1)}\right)$

R_1: $\dfrac{f_1(x)}{f_2(x)} \geq 1$ and R_2: $\dfrac{f_1(x)}{f_2(x)} < 1$

3.3 Classification with Two Multivariate Normal Populations

Classification procedures based on normal populations predominate in statistical practice because of their simplicity and reasonably high efficiency across a wide variety of population models.

We now assume that f_1 and f_2 are multivariate normal densities with respectively mean vectors μ_1 and μ_2 and the covariance matrices Σ_1 and Σ_2.

3.3.1 Classification when Σ_1=Σ_2 = Σ

The density of population π_i $(i = 1, 2)$ is now given by;

$$f_i(x) = \frac{1}{(2\pi)^{p/2} det(\Sigma)^{1/2}} \exp -\frac{1}{2}(x - \mu_i)^\tau \Sigma^{-1}(x - \mu_i) \tag{3.5}$$

If the populations π_1 and π_2 both have multivariate normal densities with equal covariance matrices,then the classification rule corresponding to minimizing ECM becomes: classify x_0 as π_1 if

$$(\mu_1 - \mu_2)^\tau \Sigma^{-1} x_0 - \frac{1}{2}(\mu_1 - \mu_2)^\tau \Sigma^{-1}(\mu_1 + \mu_2) \geq \left[\ln\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \tag{3.6}$$

and classify x_0 as π_2 otherwise.

In practice, the population parameters μ_1 and μ_2 and Σ are unknown and have to be estimated from the data. Suppose we have n_1 objects belonging to π_1 (denoted as $x_1^{(1)}$, . . . , $x_{n_1}^{(1)}$) and n_2 objects from π_2 (denoted as $x_1^{(2)}$, . . . , $x_{n_2}^{(2)}$) with $n_1 + n_2 = n$ the total sample size. The sample mean vectors and covariance matrices of both groups are estimated using their sample estimators.
Since both populations have the same covariance matrix Σ we combine the two sample covariance matrices S_1 and S_2 to obtain a more precise estimate of Σ.
Replacing μ_1 , μ_2 and Σ with $\bar{x}_1$, $\bar{x}_2$ and S_{pooled} in equation (3.6), then the sample classification rule is obtained as ; classify x_0 as π_1 if

$$(\bar{x}_1 - \bar{x}_2)^\tau S_{pooled}^{-1} x_0 - \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^\tau S_{pooled}^{-1}(\bar{x}_1 + \bar{x}_2) \geq \left[\ln\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \quad (3.7)$$

and classify x_0 as π_2 otherwise.
Below gives the special case of equation (3.7) where the prior probabilities and the misclassification cost are equal:

$$\left[\ln\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] = \ln(1) = 0 \quad (3.8)$$

such that we assign x_0 to π_1 if

$$(\bar{x}_1 - \bar{x}_2)^\tau S_{pooled}^{-1} x_0 \geq \frac{1}{2}(\bar{x}_1 - \bar{x}_2)^\tau S_{pooled}^{-1}(\bar{x}_1 + \bar{x}_2) \quad (3.9)$$

Denote $a = S_{pooled}^{-1}(\bar{x}_1 - \bar{x}_2) \in \Re^p$ and the above equation can be rewritten as; $a^\tau x_0 \geq \frac{1}{2}(a^\tau \bar{x}_1 + a^\tau \bar{x}_2)$(Johnson and Wichern,2007).

3.3.2 Classification when $\Sigma_1 \neq \Sigma_2$

The density function of the two populations π_1 and π_2 is given by;

$$f_i(x) = \frac{1}{(2\pi)^{p/2} det(\Sigma)^{1/2}} \exp -\frac{1}{2}(x-\mu_i)^\tau \Sigma^{-1}(x-\mu_i)$$

If both populations π_1 and π_2 have multivariate normal densities with mean vectors and covariance matrices μ_1 , $\varSigma_1$ and μ_2 , $\varSigma_2$ respectively, then the classification rule corresponding to minimizing ECM becomes:
classify x_0 as π_1 if

$$-\frac{1}{2}x_0^\tau(\Sigma_1^{-1} - \Sigma_2^{-1})x_0 + (\mu_1^\tau\Sigma_1^{-1} - \mu_2^\tau\Sigma_2^{-1})x_0 - k \geq \left[\ln\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \quad (3.10)$$

where the constant k is given by $k = \frac{1}{2}\ln\left(\frac{det(\Sigma_1)}{det(\Sigma_2)}\right) + \frac{1}{2}(\mu_1^\tau\Sigma_1^{-1}\mu_1 - \mu_2^\tau\Sigma_2^{-1}\mu_2)$ and classify x_0 to π_2 otherwise.

The classification regions are defined by quadratic functions of x. When $\Sigma_1 = \Sigma_2$, the quadratic term $-\frac{1}{2}x^\tau(\Sigma_1^{-1} - \Sigma_2^{-1})x$, disappears, and the regions defined earlier on, reduces. (Johnson and Wichern, 2007).

3.3.3 Quadratic Classification Rule(Normal Populations with Unequal Covariance matrices)

Allocate x_0 to π_1 if

$$-\frac{1}{2}x_0^\tau(S_1^{-1} - S_2^{-1})x_0 + (\bar{x}_1^\tau S_1^{-1} - \bar{x}_2 S_2^{-1})x_0 - k \geq \left[\ln\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \quad (3.11)$$

The awkward nature of the Quadratic discriminant function occurs in more than two dimensions and can lead to some strange results.

3.4 Classification with Several Populations

This section considers the more general situation of separating objects from g ($g \geq 2$) classes and assigning new objects to one of these g classes. For $i = 1, ..., g$ denote

1. f_i the density associated with population π_i

2. p_i the prior probability of π_i

3. R_i the subspace of outcomes assigned to π_i

4. $c(j|i)$ the cost of misclassifying an object to π_j when it is from π_i.

5. $P(j|i)$ the conditional probability of assigning an object of π_i to π_j.

The (conditional) expected cost of misclassifying an object of population π_1 is given by;

$$ECM(1) = P(2|1)c(2|1) + ... + P(g|1)c(g|1) = \sum_{i=1}^{2} P(i|1)c(i|1) \tag{3.12}$$

The expected cost of misclassifying objects of population $\pi_2, ..., \pi_g$. Hence the overall ECM becomes ;

$$ECM = \sum_{j=1}^{g} p_j ECM(j) = \sum_{j=1}^{g} \sum_{(i=1)(i\neq j)}^{g} P(i|j)c(i|j) \tag{3.13}$$

The classification rule that minimizes the ECM assigns each object x to the population π_i for which

$$\sum_{(j=1)(j\neq i)}^{g} p_j f_j(x)c(i|j)$$

is smallest. If the minimum is not unique then x can be assigned to any of the populations for which the minimum is attained.

In some special cases, we assign x to population π_i for which $\sum_{(j=1)(j\neq i)}^{g} p_j f_j(x)$ is the smallest or equivalently for which $p_i f_i(x)$ is the largest when all the misclassification costs are equal (or unknown).We therefore obtain;

classify x as π_i if

$$p_i f_i(x) > p_j f_j(x), \forall_{j\neq i} \tag{3.14}$$

3.5 Classification with Normal Populations

The density function of population $\pi_i (i = 1, ..., g)$ is given as;

$$f_i(x) = \frac{1}{(2\pi)^{p/2} det(\Sigma_i)^{1/2}} \exp -\frac{1}{2}(x-\mu_i)^\tau {\Sigma_i}^{-1}(x-\mu_i)$$

If all misclassification costs are equal (or unknown) we assign x to the population π_i if the (quadratic) score $d_i(x) = max_{j=1}^{g} d_i(x)$ where the scores are given by the equation below;

$$d_j(x) = -\frac{1}{2}\ln(\det(\Sigma_j)) - \frac{1}{2}(x-\mu_j)^\tau \Sigma_j^{-1}(x-\mu_j) + \ln(p_j), j = 1, ..., g \quad (3.15)$$

In practice, the parameters μ_j and Σ_j are unknown and will be replaced by the sample means $\bar{x}_j$ and covariance S_j which yields the sample classification rule; classify x as π_i if the (quadratic) score $\hat{d}_i(x) = max_{j=1}^{g} \hat{d}_i(x)$ where the scores are given by;

$$d_j(x) = -\frac{1}{2}\ln(\det(S_j)) - \frac{1}{2}(x-\bar{x}_j)^\tau S_j^{-1}(x-\bar{x}_j) + \ln(p_j), j = 1, ..., g \quad (3.16)$$

3.5.1 Estimated Minimum TPM Rule for Equal-Covariance Normal Populations

On assumption that, the covariance matrices are equal;$\Sigma_j = \Sigma$ for $j = 1, ..., g$, then the quadratic scores d_j becomes

$$d_j(x) = -\frac{1}{2}\ln(\det(\Sigma)) - \frac{1}{2}x^\tau \Sigma^{-1}x + \mu_j^\tau \Sigma^{-1}x - \frac{1}{2}\mu_j^\tau \Sigma^{-1}\mu_j + \ln(p_j).$$

The first two terms from the above equation are the same for all $d_j(x)$ so they can be left out, which yields the (linear) scores

$$d_j(x) = \mu_j^\tau \Sigma^{-1} x - \frac{1}{2}\mu_j^\tau \Sigma^{-1}\mu_j + \ln(p_j) \tag{3.17}$$

Based on the above equation, its sampled classification rule is obtained by classifying x as π_i if the linear (score) is $d_i(x) = max_{j=1}^{g} d_i(x)$ where the scores are given by their sample estimates as;

$$\hat{d}_j(x) = \bar{x}_j^\tau S_{pooled}^{-1} x - \frac{1}{2}\bar{x}_j^\tau S_{pooled}^{-1}\bar{x}_j + \ln(p_j), j = 1, ..., g \tag{3.18}$$

That is from above equation, we allocate x to π_i if $\hat{d}_j(x)$ is maximum. In the case of equal covariance matrices, the scores $d_j(x)$ can also be reduced to:

$$d_j(x) = -\frac{1}{2}(x-\mu_j)^\tau \Sigma^{-1}(x-\mu_j) + \ln(p_j) = -\frac{1}{2}d_\Sigma^2(x,\mu_j) + \ln(p_j) \tag{3.19}$$

where $d_\Sigma^2(x,\mu_j) = (x-\mu_j)^\tau \Sigma (x-\mu_j)$.

$d_j(x)$ can be estimated by

$$\hat{d}_j(x) = -\frac{1}{2}d_{S_{pooled}}(x,\mu_j) + \ln(p_j)$$

If the prior probabilities are all equal (or unknown) we thus assign an object x to the closest population.

3.5.2 Estimated Minimum Total probability of Misclassification(TPM) Rule for Normal Populations with Unequal covariances

Allocate x to π_i if the quadratic score $\hat{d}_i(x)$ = largest of $\hat{d}_1(x)$,$\hat{d}_2(x)$,. . . ,$\hat{d}_1(g)$ where $\hat{d}_i(x)$ is given in equation (3.15).

3.6 Minimum ECM Classification Rule with Equal Misclassification Costs

This section presents the classification rules based on the minimum ECM, as well as the equal costs of misclassification.

Allocate x_0 to π_k if

$$p_k f_k(x) > p_i f_i(x), \forall_{i \neq k} \text{ or } \ln p_k f_k(x) > \ln p_i f_i(x), \forall_{i \neq k}$$

The classification rule in the above equation is identical to the one that maximises the "posterior" probability $P(\pi_k|x)$=P(x comes from π_k given that x was observed) where

$$P(\pi_k|x) = \frac{p_k f_k(x)}{\sum_{i=1}^{g} p_i f_i(x)} = \frac{(prior) * (likelihood)}{\sum [(prior) * (likelihood)]} \tag{3.20}$$

for $k = 1, 2, ..., g$

Generally, the minimum ECM rules have the following three components; prior probabilities p_i, misclassification costs $c(i|j)$, and density functions f_i,f_j. Always the above listed components are specified or estimated before the classification rules are implemented.

3.7 Evaluating Classification Rules

To judge the performance of a sample classification procedure, we want to calculate its misclassification probability or error rate. A measure of performance that can be calculated for any classification procedure is the apparent error rate (APER) which is defined as the fraction of observations in the sample that are misclassified by the classification procedure. Let n_{1M} and n_{2M} be the number of objects misclassified as π_1 and π_2 respectively,then,

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \tag{3.21}$$

The APER is intuitively appealing and easy to calculate. Unfortunately, it tends to underestimate the actual error rate (AER) when classifying new objects. This underestimation occurs because we used the sample to "build" the classification rule. To obtain a reliable estimate of the AER we ideally consider an independent "test sample" of new objects from which we know the true class label. This means that we split the original sample in a training sample and test sample.

3.7.1 Cross Validation Procedure

An alternative to the APER is the (leave-one-out) cross-validation or jackknife procedure or the Holdout method which works as follows:

1. Leave one object out of the sample and construct a classification rule based on the remaining $n-1$ objects in the sample.

2. Classify the left-out observation using the classification rule obtained in step 1 above.

3. Repeat the two previous steps for each of the objects in the sample.

4. Let n_{1M}^{CM} and n_{2M}^{CV} be the number of left-out observations misclassified in groups 1 and 2 respectively.

A good estimate of the actual error rate is given by:

$$A\hat{E}R = \frac{n_{1M}^{CM} + n_{2M}^{CV}}{n_1 + n_2} \tag{3.22}$$

3.7.2 The Balanced Error Rate Procedure (BER)

The Balanced Error Rate (BER) statistic is the average of the misclassification rates on samples drawn from positive and negative classes (denoted by C^+ and C^- respectively) as shown in the table below.

Table 3.1: Confusion Matrix for two class pattern recognition

	Prediction	
True Pop.	C^-	C^+
C^-	a	b
C^+	c	d

Where a, b, c and d are entries in the confusion matrix. The Balance Error Rate can be given mathematically as

$$BER = \frac{1}{2}\left[\frac{b}{a+b} + \frac{c}{c+d}\right] \tag{3.23}$$

3.8 Equal mean Discrimination

There are some applications of discriminant analysis in practice in which the groups may be assumed to have the same mean vectors. In this case for multivariate normal group conditional distributions, the optimal discriminant rule $r_o(x;\theta_U)$ has to be based on the differences between the group-covariance. matrices.

3.8.1 Comparing Two Mean Vectors (Hotelling T^2)

This section considers the case where p variables are measured on each sampling unit in two samples.We wish to test $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$.

A random sample from two populations are obtained. The random sample from the first population, correspond to $x_{11}, x_{12}, ..., x_{1n_1}$ from $N_p(\mu_1, \Sigma_1)$ and a second random sample $x_{21}, x_{22}, ..., x_{2n_2}$ from $N_p(\mu_2, \Sigma_2)$. We assume that the two samples are independent and that $\Sigma_1 = \Sigma_2 = \Sigma$ with Σ unknown. The assumption is neccessary for the T^2-test statistics to have a T^2-distribution. The sample mean vectors are estimated as $\bar{x}_1 = \sum_{i=1}^{n_1} x_{1i}/n_1$ and $\bar{x}_2 = \sum_{i=1}^{n_2} x_{2i}/n_2$. Let $\mathbf{W}_1$ and $\mathbf{W}_2$ be the sum of squares and cross product matrices for the two samples.

$$\mathbf{W}_1 = \sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)(x_{1i} - \bar{x}_1)' = (n_1 - 1)\mathbf{S}_1$$

and

$$\mathbf{W}_2 = \sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)(x_{2i} - \bar{x}_2)' = (n_2 - 1)\mathbf{S}_2.$$

$(n_1 - 1)\mathbf{S}_1$ is an unbiased estimator of $(n_1 - 1)\Sigma_1$ and $(n_2 - 1)\mathbf{S}_2$ is an unbiased estimator of $(n_2 - 1)\Sigma_2$. Then the two covariance matrices are pooled together to obtain an unbiased estimator of the common population covariance matrix, Σ.

$$S_{pl} = \frac{\mathbf{W}_1 + \mathbf{W}_2}{n_1 + n_2 - 2}$$

Thus $\mathbf{E}(S_{pl})$=Σ. By substitution and generalisation, we obtain the Hotelling T^2 as;

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\bar{x}_1 - \bar{x}_2)' S_{pl}^{-1}(\bar{x}_1 - \bar{x}_2) \tag{3.24}$$

Where it is distributed as T^2_{p,n_1+n_2-2} when $H_0 : \mu_1 = \mu_2$ is true. Therefore reject H_o if $T^2 \geq T^2_{\alpha,p,n_1+n_2-2}$. The T^2 test statistics can be transformed to F-test statistics and its given below:

$$\frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} T^2 = F_p, n_1 + n_2 - p - 1 \tag{3.25}$$

where the dimension p of the the T^2 statistics becomes the first degree of freedom parameter for the F statistics. Rencher (2002).

3.8.2 Bartlett and Please approach to Equal mean discrimination

This section considers applications of discriminant analysis in which the population mean vectors are assumed to be equal. For multivariate normal group conditional distributions, the optimal discriminant rule has to be based on the individual group-covariance matrices. Bartlett and Please (1963) considered tackling the problem of equal mean discrimination with two populations using the well-known twins data of Stocks (1933) in order to study the usefulness of the quadratic discriminant rule in discriminating between Monozygotic and Dizygotic twins. They adopted the general uniform covariance structure and its given below;

$$\Sigma_i = \sigma_i^2 \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & \vdots & \vdots & & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix} = \begin{pmatrix} \sigma_i^2 & \sigma_i^2\rho & \sigma_i^2\rho & \cdots & \sigma_i^2\rho \\ \sigma_i^2\rho & \sigma_i^2 & \sigma_i^2\rho & \cdots & \sigma_i^2\rho \\ \vdots & \vdots & \vdots & & \vdots \\ \sigma_i^2\rho & \sigma_i^2\rho & \sigma_i^2\rho & \cdots & \sigma_i^2 \end{pmatrix}$$

$$\Sigma_i = \sigma_i^2\{(1-\rho_i)\mathbf{I}_p + \rho_i \mathbf{1}_p \mathbf{1}_p'\} \tag{3.26}$$

Where $\mathbf{1}$ is a column vector of 1's and ρ is the population correlation coefficient between any two variables. This pattern of equal covariances and equal variances in Σ is variously referred to as *uniformity*, *compound symmetry*, or the *intraclass correlation model*.

We obtain the sample covariance matrix $\mathbf{S}$. Estimates of σ^2 and $\sigma^2\rho$ are given by; $s^2 = \dfrac{1}{p}\sum_{j=1}^{p} s_{jj}$ and $s^2 r = \dfrac{1}{p(p-1)}\sum_{j\neq k} s_{jk}$ respectively, where s_{jj} and s_{jk} are from $\mathbf{S}$. An average of the variances on the diagonal of $\mathbf{S}$ is given as s^2, and $s^2 r$ is an average of the off-diagonal covariances in S. Here an estimate of ρ can be obtained as $r = \dfrac{s^2 r}{s^2}$. The estimate of Σ is given below using s^2 and $s^2 r$.

$$\mathbf{S}_i = \begin{pmatrix} s^2 & s^2r & s^2r & \cdots & s^2r \\ s^2r & s^2 & s^2r & \cdots & s^2r \\ \vdots & \vdots & \vdots & & \vdots \\ s^2r & s^2r & s^2r & \cdots & s^2 \end{pmatrix} = s^2\{(1-r_i)\mathbf{I}_p + r_i\mathbf{1}_p\mathbf{1}'_p\}$$

Bartlett and Please standardised the first covariance matrix corresponding to the Monozygotic twins as;

$$\Sigma_1 = (1-\rho_1)\mathbf{I} + \rho_1\mathbf{1}_p\mathbf{1}'_p \tag{3.27}$$

They further assumed that, Σ_2 cannot be simultaneously standardized to unit variances and was given by;

$$\Sigma_2 = \sigma^2(1-\rho_2)\mathbf{I} + \rho_2\mathbf{1}_p\mathbf{1}'_p \tag{3.28}$$

That is in generality, they assumed that σ_1^2=1, with $\rho_1 = \rho_2 = \rho$ where σ^2 is obtained as the ratio of the sum of squares of the Dizygotic twins π_2 to that of the Monozygotic twins π_1.

Desu and Geisser (1973) made further assumption on the fact that the two population correlation coefficient are not the same that is, $\rho_1 \neq \rho_2$. With reference to the two uniform covariance matrices above, the inverses of Σ_1 and Σ_2 are given below. Let $\mathbf{1}_p\mathbf{1}'_p = E$, where E is the matrix with entries equal to unity.

$$\Sigma_1^{-1} = \frac{\mathbf{I}}{(1-\rho_1)} - \frac{\rho_1}{(1-\rho_1)}\frac{\mathbf{E}}{(1+(p-1)\rho_1)} \tag{3.29}$$

and

$$\Sigma_2^{-1} = \frac{1}{\sigma^2}\left[\frac{\mathbf{I}}{(1-\rho_2)} - \frac{\rho_2}{(1-\rho_2)}\frac{\mathbf{E}}{(1+(p-1)\rho_2)}\right] \tag{3.30}$$

Therefore using the log likelihood or the likelihood ratio $\dfrac{f_1(x)}{f_2(x)}$ which is one of the most efficient criterion for classification, by ignoring the additive constants gives the optimal discriminant function below in the situation where the difference between the mean vectors of the two populations is zero.

$x\Sigma_1^{-1}x - x'\Sigma_2^{-1}x$

$$= \left[\frac{1}{1-\rho_1} - \frac{1}{\sigma^2(1-\rho_2)}\right] Z_1 - \left[\frac{\rho_1}{1-\rho_1}\frac{1}{1+(p-1)\rho_1} - \frac{\rho_2}{\sigma^2(1-\rho_2)}\frac{1}{1+(p-1)\rho_2}\right] Z_2 \tag{3.31}$$

where $p =$ the number of independent variables/measurements, $Z_1 = tr(zz')$ and $Z_2 = tr(Ezz\prime)$, with z being the observational vector belonging to either π_1 or π_2.

It was observed that, the ideal discriminant function involves only Z_1 and Z_2 so that these two quantities are plotted and gives a resulting straight line boundary. After making further assumptions that, $\rho_1 = \rho_2 = \rho$,the discriminant function given by equation (3.30) becomes

$$Z_1 - \frac{\rho}{1+(p-1)\rho} Z_2 \tag{3.32}$$

Such that the coefficient of Z_2 is estimated numerically. They further developed an equation for determining the best boundary to give equal risks of misclassification for both groups/populations that may be written as

$$\frac{p(1-\rho)\log\sigma^2}{1-\sigma^{-2}} - 2c \tag{3.33}$$

where $c = \log\left(\frac{q_2}{q_1}\right)$, q_1 and q_2 being the prior probabilities of populations π_1 and π_2 respectively.

The classification rule then becomes, assign z to π_1 if and only if

$$Z_1 - \frac{\rho}{1+(p-1)\rho} Z_2 < \frac{p(1-\rho)\log\sigma^2}{1-\sigma^{-2}} - 2c \tag{3.34}$$

Otherwise assign z to π_2.

If we let equation (3.30)$= U$, then we assign z to π_1 if and only if $U > c$.

3.9 Distance and Classification

There are two Mahalanobis-type distance indices that are of importance in discriminant analysis. One is an index of distance between two points where each point represents a vector of means on the p variables. Having two centroids, μ_1 and μ_2 in their respective populations,the distance between the two centroids within their respective populations becomes

$$\triangle_{12}= \left[(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)\right]^{1/2} \tag{3.35}$$

In this case, the covariance matrix, Σ for the two populations are assumed to be equal. Another type of Mahalanobis distance is appropriate when one point represents a vector of p observations on an analysis unit and the other point represents a centroid for a population. Consider i populations of interest; the distance between x_z the observation vector for unit z and μ_i, the centroid for population i may be given by;

$$\triangle_{zi}= \left[(x_z - \mu_i)'\Sigma_i^{-1}(x_z - \mu_i)\right]^{1/2} \tag{3.36}$$

where Σ_i is the covariance matrix for Population i. The Mahalanobis distance derived above is of much importance in discriminant analysis, since the main goal in classification is to classify a unit into that population to which the unit is nearest. That is the classification rule here is to classify unit z into population i if $\triangle_{zi}$ is smaller than $\triangle_{zi'}$, for $i' \neq i$ and $i, i' = 1, 2, ..., i$. That is, the equal mean vectors for the two populations can be much more applicable to the second type of the Mahalanobis distance derived in equation (3.31) above. For the univariate case, $\triangle_{zi}$ can be derived as;

$$\triangle_{zi}= \left[(x_z - \mu_i)'\frac{1}{\sigma_i^2}(x_z - \mu_i)\right]^{1/2} = \left[\frac{(x_z - \mu_i)^2}{\sigma_i^2}\right]^{1/2} \tag{3.37}$$

3.10 Bayesian Classification

This section presents the Bayesian classification rule for classifying observations into their respective groups. It consist basically of prior and posterior probabilities.

3.10.1 Prior Probability

Let π_j denote the proportion of units in the total observations/units in Population j. That is the probability that a unit u will be randomly selected from the universe and it will be from population j is π_j. We denote π_j as the prior probability of membership in population j. The word prior means that, this is a probability of population membership before x_u is known. It is reasonable that these prior probabilities be taken into consideration when arriving at values of $P(j|x_u)$. The product of the prior and the posterior probabilities, $\pi_j.P(x_u|j)$ denotes the joint probability that a randomly selected unit belongs to Population j and at the same time has a score vector close to x_u.

3.10.2 Posterior Probability

Considering the probability of unit u belonging to Group j, given that the unit has a particular observation vector,x_u. This probability, denoted by $P(j|x_u)$, is called the posterior probability of membership in Population j. The word "posterior" means that, it is a probability of population membership conditioned on knowing x_u. Therefore, a unit is assigned to the population for which $P(j|x)$, that is the posterior probability of membership, is greatest. Given the probability that a unit belongs to Population j (given an observed score vector) is equal to the ratio of the probability of its score vector in Population j to the sum of the probabilities associated with its score vector in all J groups. That is

$$P(j|X_u) = \frac{P(x_u|j)}{\sum_{j'=1}^{j} P(x_u|j')} \tag{3.38}$$

Therefore assign unit u to population j if

$$P(j|x_u) > P(j'|x_u), for j' \neq j \tag{3.39}$$

The product of the prior and the posterior probabilities are therefore use to arrive at values of $P(j|X_u)$ by employing the Bayes theorem rule developed by Rev.T Bayes. Incorporating prior probabilities, the posterior probability of unit u belonging to Population j given a score vector is given by;

$$P(j|x_u) = \frac{\pi_j.P(x_u|j)}{\sum_{j'=1}^{J} \pi_j.P(x_u|j')} \tag{3.40}$$

The Bayesian probability rule is stated as below; Assign unit u to population j if

$$P(j|x_u) > P(j'|x_u)$$

for $j \neq j'$, where $P(j|x_u)$ is defined in equation (3.39) above. Also because $P(x_u|j)$ values are proportional to $f(x_u|j)$ values, we consider j values of $\pi_j.f(x_u|j)$. Hence the posterior probability becomes;

$$P(j|X_u) = \frac{\pi_j.f(x_u|j)}{\sum_{j'=1}^{j} \pi_j.f(x_u|j')} \tag{3.41}$$

By using the above Bayesian probabilities, the total number of misclassification errors is minimized. Huberty and Olejnik (2006).

3.11 Discrimination using absolute values: The Absolute Euclidean Distance Classifier (AEDC)

The Euclidean distance classifier cannot be used when the centroids for the two populations are equal, $\mu_1 = \mu_2$. AEDC are used when the absolute values of the components of the observations are used in Euclidean Distance Classifier (EDC).

It is expected that, this approach does well in a high dimensional data set. AEDC is mostly used and applicable in situations when the $\Sigma_1 \neq \Sigma_2$ and $\mu_1 = \mu_2$. AEDC and QDF are always used as an alternative to LDF.

The Euclidean distance classifier (EDC) will allocate an observational vector X to population $1(\pi_1)$ if

$$\left\{x - \frac{1}{2}(\mu_1 + \mu_2)\right\}^T .(\mu_1 - \mu_2) > 0 \tag{3.42}$$

Otherwise to population $2(\pi_2)$

For the case of equal mean, that is taking the absolute values of the observational vectors, then $Y = |X|$, for instance using a three dimensional vector, allocate a three dimensional observational vector to π_1 if

$$y_1(\mu_1^1 - \mu_1^2) - \frac{1}{2}((\mu_1^1)^2 - (\mu_1^2)^2) + y_2(\mu_2^1 - \mu_2^2) - \frac{1}{2}((\mu_2^1)^2 - (\mu_2^2)^2) + y_3(\mu_3^1 - \mu_3^2) - \frac{1}{2}((\mu_3^1)^2 - (\mu_3^2)^2) > 0 \tag{3.43}$$

Where $\mu_i^{(k)}$ is the mean of the ith component of Y in the kth population, for instance given that $i = 1, 2, 3$ and $k = 1, 2$. Therefore the AEDC is given as:

$$y_1\left(\sqrt{\frac{2}{\pi}\sigma_{11}^{(1)}} - \sqrt{\frac{2}{\pi}\sigma_{11}^{(2)}}\right) - \frac{1}{2}\left(\frac{2}{\pi}(\sigma_{11}^{(1)} - \sigma_{11}^{(2)})\right) + y_2\left(\sqrt{\frac{2}{\pi}\sigma_{22}^{(1)}} - \sqrt{\frac{2}{\pi}(\sigma_{22}^{(2)}}\right) - \frac{1}{2}\left(\frac{2}{\pi}(\sigma_{22}^{(1)} - \sigma_{22}^{(2)})\right)$$
$$+ y_3\left(\sqrt{\frac{2}{\pi}\sigma_{33}^{(1)}} - \sqrt{\frac{2}{\pi}(\sigma_{33}^{(2)}}\right) - \frac{1}{2}\left(\frac{2}{\pi}\sigma_{33}^{(1)} - \sigma_{(33)}^{2})\right) > 0$$

Otherwise to population 2. It can be also written as, allocate X to π_1 if

$$\left[y_1\left(\sqrt{\sigma_{11}^{(1)}} - \sqrt{\sigma_{11}^{(2)}}\right) + y_2\left(\sqrt{\sigma_{22}^{(1)}} - \sqrt{\sigma_{22}^{(2)}}\right) + y_3\left(\sqrt{\sigma_{33}^{(1)}} - \sqrt{\sigma_{33}^{(2)}}\right)\right] \geq$$
$$\frac{1}{2}\sqrt{\frac{2}{\pi}}\left[(\sigma_{11}^{(1)} - \sigma_{11}^{(2)} + \sigma_{22}^{(1)} - \sigma_{22}^{(2)} + \sigma_{33}^{(1)} - \sigma_{33}^{2})\right]$$

In general , we allocate observation vector X to π_1 if

$$\sum_{i=1}^{p}\left[y_i\left(\sqrt{\sigma_{ii}^{(1)}} - \sqrt{\sigma_{ii}^{(2)}}\right) - \frac{1}{2}\sqrt{\frac{2}{\pi}}(\sigma_{ii}^{(1)} - \sigma_{ii}^{(2)})\right] > 0. \tag{3.44}$$

Chapter 4

Results and Discussion

This section presents the results of our data analyses and detailed discussion. The Chapter comprises, preliminary analysis and detailed analyses based on the objectives of the study.

4.1 Preliminary Analysis

Hotelling T^2 was used to test whether there exist differences between the mean vectors of the Monozygotic and the Dizygotic twin groups. The significance of the mean difference was tested in order to ensure that, the assumption of the equal mean vectors for the groups are not violated. The hypothesis of testing the equality of the mean vectors was stated as:

.

$$H_0 : \begin{pmatrix} \mu_{1,1} \\ \mu_{1,2} \\ \mu_{1,3} \\ \vdots \\ \mu_{1,10} \end{pmatrix} = \begin{pmatrix} \mu_{2,1} \\ \mu_{2,2} \\ \mu_{2,3} \\ \vdots \\ \mu_{2,10} \end{pmatrix} \text{ Against } H_1 : \begin{pmatrix} \mu_{1,1} \\ \mu_{1,2} \\ \mu_{1,3} \\ \vdots \\ \mu_{1,10} \end{pmatrix} \neq \begin{pmatrix} \mu_{2,1} \\ \mu_{2,2} \\ \mu_{2,3} \\ \vdots \\ \mu_{2,10} \end{pmatrix}$$

The mean vectors for the two groups (i.e the Monozygotic twins and that of the Dizygotic twins) were computed from the twin data and they are summarised below: Let the mean vectors for Monozygotic twins be denoted as μ_M and that of the Dizygotic twin group be denoted as μ_D. Therefore the estimates of μ_M and μ_D are

$$\mu_M = \begin{array}{c} Ht \\ Wt \\ HL \\ HB \\ HC \\ ID \\ BPS \\ PI \\ SGR \\ SGL \end{array} \begin{pmatrix} 0.0617 \\ -0.0197 \\ -0.239 \\ 0.035 \\ -0.205 \\ -0.134 \\ 0.078 \\ 0.352 \\ -0.259 \\ -0.106 \end{pmatrix} \text{ and } \mu_D = \begin{array}{c} Ht \\ Wt \\ HL \\ HB \\ HC \\ ID \\ BPS \\ PI \\ SGR \\ SGL \end{array} \begin{pmatrix} 0.636 \\ -0.503 \\ -0.253 \\ 0.247 \\ -0.205 \\ -0.217 \\ -0.616 \\ -0.356 \\ -0.030 \\ -0.233 \end{pmatrix}$$

respectively. The two variance covariance matrices were pooled together to assume equal covariance matrix between the two twin groups. Hence the computed pooled variance covariance matrix, Σ is given below:

$$\Sigma = \begin{array}{c} \\ Ht \\ Wt \\ HL \\ HB \\ HC \\ ID \\ BPS \\ PI \\ SGR \\ SGL \end{array} \begin{array}{c} \begin{array}{cccccccccc} Ht & Wt & HL & HB & HC & ID & BPS & PI & SGR & SGL. \end{array} \\ \begin{pmatrix} 0.98 & 0.69 & 0.70 & 0.34 & 0.63 & 0.17 & 0.11 & 0.01 & 0.20 & 0.13 \\ 0.69 & 1.36 & & & & & & & & \\ 0.70 & 0.76 & 1.26 & & & & & & & \\ 0.34 & 0.36 & 0.36 & 0.88 & & & & & & \\ 0.63 & 0.94 & 1.11 & 0.52 & 1.27 & & & & & \\ 0.17 & 0.26 & 0.27 & 0.20 & 0.34 & 0.21 & & & & \\ 0.11 & -0.08 & 0.09 & 0.38 & 0.31 & 0.22 & 1.40 & & & \\ 0.01 & -0.13 & -0.25 & 0.05 & -0.12 & -0.02 & 0.52 & 0.87 & & \\ 0.20 & 0.18 & 0.27 & 0.14 & 0.26 & 0.05 & 0.21 & 0.04 & 0.99 & \\ 0.13 & 0.32 & 0.12 & 0.21 & 0.28 & 0.11 & 0.28 & -0.02 & 0.07 & 0.49 \end{pmatrix} \end{array}$$

and the inverse of the pooled variance covariance matrix given by:

$\Sigma^{-1} =$

$$
\begin{array}{c|cccccccccc}
 & Ht & Wt & HL & HB & HC & ID & BPS & PI & SGR & SGL. \\
\hline
Ht & 2.41 & & & & & & & & & \\
Wt & -1.13 & 2.65 & & & & & & & & \\
HL & -1.95 & 1.35 & 6.05 & & & & & & & \\
HB & -0.47 & 0.27 & 0.66 & 1.78 & & & & & & \\
HC & 1.58 & -2.56 & -5.86 & -1.02 & 8.27 & & & & & \\
ID & -0.09 & -0.79 & 0.37 & -0.64 & -1.95 & 10.51 & & & & \\
BPS & -0.18 & 1.05 & 0.42 & -0.10 & -0.87 & -1.61 & 1.81 & & & \\
PI & -0.39 & -0.25 & 0.85 & 0.02 & -0.35 & 0.93 & -0.97 & 1.92 & & \\
SGR & -0.05 & -0.07 & -0.23 & -0.05 & 0.02 & 0.38 & -0.21 & 0.01 & 1.11 & \\
SGL & -0.01 & -0.84 & 0.94 & -0.19 & -0.65 & 0.39 & -0.89 & 0.68 & -0.01 & 3.26
\end{array}
$$

The table below summarises the output of the estimate of the Hotelling T^2, the $P-value$ and the $F-test$ statistics.

Table 4.1: Parameter estimation of test statistics

$(HotellingT^2)$	$F-test$	$P-value$	α
15.692	1.064	0.358	0.05

From Table 4.1 above, the Hotelling T^2 statistics was found to be 15.692, $F-test$ statistics to be 1.064, 0.358 being $P-value$, and the level of significance (α) to be 0.05. By rule, the null hypothesis of no difference of the two mean vectors are rejected if $P-value < \alpha$. From Table 4.1, $P-value$=0.356 $> \alpha = 0.05$. Since the P-value is greater than the level of significance, we fail to reject the null hypothesis and conclude that, there exist no difference between the two mean vectors. Meaning there exist no significant mean difference between the Monozygotic twins and that of the Dizygotic twins constituting the two separate populations. Hence, discrimination will be solely based on the groups variance covariance matrices.

4.2 Case 1: Classification with Equal Prior Probabilities

This study focussed on classifying the twin pair observations into their respective groups based on the classification functions derived using the following four classification methods; The Bartlett and Please Method (BPM), the Bayesian Classifier based on posterior probabilities, the Quadratic Discriminant Function approach (QDF) and the Absolute Euclidean Distance Classifier approach (AEDC).

4.2.1 Bartlett and Please Classification Method for $n_1 = n_2$

This case deals with discriminating between both the Monozygotic twins and Dizygotic twins when equal prior probabilities are assumed for the groups. Bartlett and Please applied their method of discrimination with equal prior probabilities and they came out with a discriminant function and their corresponding classification rule. The discriminant function as well as the classification rule for this method was obtained from the likelihood ratios of the two normal density functions for both the Monozygotic twin group and that of the Dizygotic twin group. The method also employed the use of Uniform Covariance matrices for both groups with an assumption of equal population correlation coefficient $\rho_1 = \rho_2$ as well as the inequality of the correlation coefficient assumption $\rho_1 \neq \rho_2$. In other to come out with the boundary for the cost of misclassification and other functions, the common variance was one of the parameters used often and the estimate of the common variance (σ^2) was obtained from the ratio of the sum of squares of Dizygotic group to that of the Monozygotic group. Other estimates for ρ_1, ρ_2, ρ, Σ_1 and Σ_2 were obtained from their sample estimates from the sampled data. Fifteen pairs of twins were selected for each group with ten measured characteristics. The difference between the observations of each twin pair were taken to represent an observation. The discriminant functions deduced was based on only the first ten observations from each group. The observational vectors, z were reconstructed using $tr(zz') = Z_1$ and $tr(Ezz') = Z_2$. Values of Z_1 and Z_2 for each of the twin groups were obtained and in all 30 estimates were made, with E being a $10 * 10$ matrix with entries of **1's**. The values obtained were plotted and a scatter diagram was obtained. Ten (10) samples taken from each of the two groups were used for obtaining the discriminant functions, the classification rules as well as the cut-off point(the boundary of misclassification). The Figure below shows the scatter plot of the various observational vectors in the two populations transformed into scalar values.

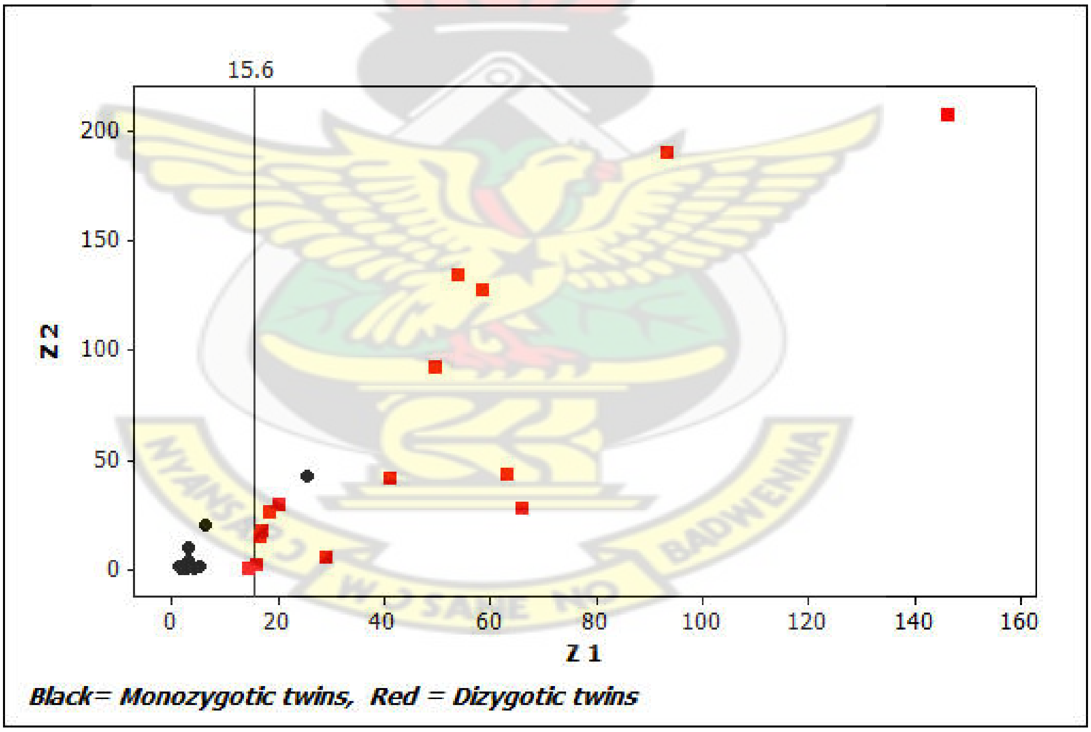

Figure 4.1: scatter plot of observations from both Mono. and Dizy. groups with equal priors

Table 4.2: Table of estimates of parameters for $n_1 = n_2$

Sample Ratios	σ^2	ρ_1	ρ_2	ρ	cut-off
$n_1 : n_2$	4.0296	0.0478	0.2194	0.1336	15.60

From Figure 4.1 and Table 4.2 above, it is observed that, the cut-off point for classification for assumed equal correlation coefficient was obtained as 15.6 and the estimate of the overall common variance, σ^2 recorded as 4.0296. Estimates of ρ_1 and ρ_2 corresponding to Monozygotic and Dizygotic twin groups were obtained as 0.0478 and 0.2194 respectively and that of the common ρ as 0.1336. It can be seen that all the observations in the Dizygotic group are widely varied, and that of the Monozygotic twins found to be less varied or closely located to each other. As evident from the plot, only one observation(individual) was misclassified from the Monozygotic twin group and two (2) observations being misclassified from the Dizygotic twin group. However, none of the remaining five observations from each of the two twin groups which were not included in deriving the function were misclassified, indicating a very good and reliable function which possibly provides maximum separation between the two groups. From the plot above, the linear discriminant function, assuming $\rho_1 = \rho_2 = \rho = 0.1336$ was obtained as;

$$Z_1 - 0.0606Z_2 \tag{4.1}$$

The discriminant scores after substituting all the fifteen individual twin observations into equation 4.1 are summarised in Table 4.1 below. Let D_M be the discriminant scores for classifying observations into the Monozygotic twin group.

Table 4.3: Discriminant scores for classifying observations as Monozygotic twins

D_M	1.53	5.03	1.58	2.83	5.05	1.60	2.42	2.91	**22.81**
	2.70	3.98	2.47	1.59	2.51	1.45			

And the discriminant scores, D_D for classifying observations into the Dizygotic twin group are described in Table 4.3 below.

Table 4.4: Discriminant scores for classifying observations as Dizygotic twins

D_D	15.87	28.77	16.65	50.77	60.40	**14.37**	64.12	38.61	**15.59**
	15.58	133.63	45.58	81.79	43.92	18.21			

The classification rule for this particular method was to assign x_1 to π_1 if the discriminant scores D_i obtained from equation 4.1, is $D_i < 15.60$ otherwise to π_2. From Table 4.3, it was observed that the score 22.81 exceeded the cut-off point 15.6, indicating a misclassified observation into Dizygotic group. The scores for the Dizygotic group from Table 4.4 also recorded three (i.e 14.37, 15.59, 15.58 are all less than 15.60) misclassified observations.

The table below gives the confusion matrix of the misclassified observations.

Table 4.5: Confusion Matrix for Bartlett and Please method with common ρ

	Classified as		
True Pop.	π_1	π_2	*Total*
π_1	14.0	1.0	15.0
π_2	3.0	12.0	15.0
Total	17.0	13.0	30.0

Where π_1 and π_2 represents a group of Monozygotic twins and a group of Dizygotic twins respectively. The probability of correct classification from the table was found to be 0.867.

Assuming $\rho_1 \neq \rho_2$

Bartlett and Please (1963) and Desu and Geisser (1973) derived another discriminant function from the likelihood ratio of the two density functions with unequal population correlation coefficient common among the individual groups. With reference to equation (3.26) and by substitution, the discriminant function below was obtained.

$$0.732Z_1 - 0.011Z_2 \tag{4.2}$$

Based on the above linear function, the following discriminant scores were obtained for each group.

Table 4.6: Discriminant scores for classifying observations as Monozygotic twins

D_M	1.16	3.73	1.20	2.21	4.35	1.17	1.78	2.20	**18.11**
	1.98	2.93	2.12	1.21	1.84	1.09			

The Discriminant scores for observations (individuals) in the Dizygotic twin groups are summarised in Table 4.7.

Table 4.7: Discriminant scores for classifying observations as Dizygotic twins

D_D	**12.19**	21.24	**13.05**	41.37	45.65	**10.53**	47.87	29.64	**11.49**
	11.89	104.67	37.77	66.14	35.20	**14.32**			

From the two tables above (Table 4.6 and 4.7), only one observation was misclassified from the Monozygotic twin group according to the computed discriminant scores with six (6) observations being misclassified from the Dizygotic twin group. Comparatively, the discriminant function derived when the two correlation coefficients are assumed to be equal provides maximum separation based on the total number of misclassified observations in each case.

In relation to existing literature, the results shows an agreement with the research work of Bartlett and Please (1963), where one and two observations were misclassified from both the Monozygotic and the Dizygotic twin groups. Its also shows a relation with the research based work of Desu and Geisser (1968, 1973) where in their study one of the five observations not used for obtaining the discriminant functions was misclassified from the Dizygotic group.

4.2.2 The Bayesian Posterior Probability Approach for $n_1 = n_2$

This particular method focusses on classifying an observation with the greatest posterior probability value. It involves the computation of likelihoods based on the density functions for two twin groups. In the application of the Bayesian Classifier based on the posterior probabilities, we assumed equal prior probabilities as well as normality.

This section considers classification of observations from both the Monozygotic twins and that of the Dizygotic twins using the Bayes rule method for classification when the prior probabilities were assumed to be equal. i.e $(0.5 + 0.5 = 1)$. Table 4.8 summarises the values for the likelihoods for each group as well as their corresponding posterior probabilities. Let $f(x|j)$ and $f(x|j')$ be the likelihood functions for the Monozygotic and Dizygotic group respectively. Denote the density functions of Monozygotic and Dizygotic groups as $f_1(x)$ and $f_2(x)$ respectively. Also let x_M and x_D be the observations selected from the Monozygotic and the Dizygotic twin groups.

The table below presents the summary of the posterior probabilities of an observational vector belonging to either Monozygotic group or the Dizygotic group. Their corresponding likelihoods are presented in Appendix A.

Table 4.8: Posterior probabilities for equal group prior probability

$P(j\|x_M)$	$P(j'\|x_M)$	$P(j\|x_D)$	$P(j'\|x_D)$
7.070e-01	2.930e-01	**6.69e-01**	**3.31e-01**
8.190e-01	1.810e-01	3.41e-01	6.59e-01
5.580e-01	4.420e-01	1.48e-01	8.52e-01
4.890e-01	**5.110e-01**	5.68e-02	9.43e-01
6.780e-01	3.220e-01	**6.43e-01**	**3.57e-01**
4.440e-01	**5.560e-01**	**6.03e-01**	**3.97e-01**
7.510e-01	2.490e-01	1.00e-01	9.00e-01
8.450e-01	1.550e-01	2.49e-01	7.51e-01
9.510e-01	4.890e-02	2.79e-01	7.21e-01
8.880e-01	1.120e-01	4.15e-01	5.85e-01
6.130e-01	3.870e-01	1.31e-03	9.99e-01
6.790e-01	3.210e-01	9.96e-02	9.00e-01
7.840e-01	2.160e-01	6.29e-03	9.94e-01
5.780e-01	4.220e-01	**5.25e-01**	**4.75e-01**
4.980e-01	5.020e-01	2.02e-01	7.98e-01

The Bayes rule of posterior probabilities stipulates that, the posterior probability of an observational vector from the Monozygotic group belonging to the Monozygotic group $P(j|x_M)$ should be greater than the posterior probability of an observational vector from Monozygotic group belonging to the Dizygotic group $P(j'|x_M)$. That is an observation x_M is more likely to belong to the Monozygotic population (j) than that of the Dizygotic population j'. Table 4.8 above, shows that, only three (3) observations were found to be misclassified from the Monozygotic group since their posterior probability values were all found to be less than their corresponding posterior probabilities of an observation x_M belonging to population j'. (i.e observations 4, 6 and 15 from the table).

On the other hand, it was observed that, four (4) values of posterior probabilities of x_D belonging to the Dizygotic group (j') were found to be less than those belonging to the Monozygotic group (j). This indicates four misclassified twin observations from the Dizygotic twin group. The four misclassified observation corresponds to the posterior probability values of observations 1, 5, 6 and 14 in Table 4.8. The Bayesian classification procedure presented in this study shows a true reflection with the research based study of Johnson and Wichern (2007).

The table below summarises the total number of true classification of observations as well as misclassified observations. From Table 4.9, the probability of correct classification was 0.767 percent.

Table 4.9: Confusion Matrix for Bayes rule approach for equal population prior probabilities

	Classified as		
True Pop.	π_1	π_2	*Total*
π_1	12.0	3.0	15.0
π_2	4.0	11.0	15.0
Total	16.0	14.0	30.0

4.2.3 Quadratic Discriminant Function Approach for $n_1 = n_2$

This section applies the Quadratic Discriminant function (QDF) approach for classifying observations into either population π_1 or π_2 for equal prior probabilities and misclassification costs. Mahalanobis Distances for each group was also used to classify the twin pair observations. As already explained in Chapter three, the QDF is normally used in the case of equal mean discrimination, when the assumptions of unequal covariance matrices as well as normality are satisfied.

Ganeslingam et al. (2006) deduced the QDF for equal prior probabilities and equal misclassification cost as;

$$\ln\left\{\frac{|\Sigma_1|}{|\Sigma_2|}\right\} - \left\{(x-\mu_1)^\tau \Sigma_1^{-1}(x-\mu_1)\right\} + \left\{(x-\mu_2)^\tau \Sigma_2^{-1}(x-\mu_2)\right\} > 0$$

Which can be rewritten as;
$\left\{(x-\mu_1)^{\tau}\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)^{\tau}\Sigma_2^{-1}(x-\mu_2)\right\} < \ln\left\{\frac{|\Sigma_1|}{|\Sigma_2|}\right\}$
For the equal prior probability case, i.e 0.5 for each twin group, 15 pairs of observations were sampled from each group. However the discriminant function obtained in this case as shown above, was composed of as the difference between the Mahalanobis distance for the two respective twin groups and their additive constant. Therefore we assigned observation x_1 to Monozygotic group π_1 when;

$$\left\{(x-\mu_1)^{\tau}\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)^{\tau}\Sigma_2^{-1}(x-\mu_2)\right\} < -0.1664 \tag{4.3}$$

Based on the above equation the discriminant scores for both groups were obtained. The table below presents the discriminant scores for the case of equal prior probabilities and equal misclassification costs.

Table 4.10: Discriminant scores for equal prior probability case

$D_M < -0.166$	$D_D > -0.166$
1.599	**-1.245**
-2.851	1.483
-0.303	3.662
0.253	5.785
-1.321	**-1.007**
0.612	**-0.672**
-2.044	4.550
-3.230	2.378
-5.767	2.064
-3.968	0.853
-0.753	13.439
-1.336	4.570
-2.417	10.291
-0.460	-0.030
0.180	**2.915**

From Table 4.10, we observed four misclassified observations from both twin groups based on the cut-off point (-0.166), representing approximately 73 percent correct classification rate. Hence approximately 73 percent of the observations from both the Monozygotic and the Dizygotic twin groups were correctly classified.

4.2.4 The AEDC Classification Method for $n_1 = n_2$

This method was applied as already explained in Chapter three when the covariance matrices for the two twin groups were assumed to be unequal with equal mean vectors across the groups. This sections applies the AEDC methods for the case of equal prior probabilities and misclassification costs.

The absolute linear discriminant function, derived in this case for a larger dimensional data set (i.e a 10 by 10) matrix was derived by;

$$y_1\left(\sqrt{\frac{2}{\pi}\sigma_{11}^{(1)}} - \sqrt{\frac{2}{\pi}\sigma_{11}^{(2)}}\right) - \frac{1}{2}\left(\frac{2}{\pi}(\sigma_{11}^{(1)} - \sigma_{11}^{(2)})\right) + y_2\left(\sqrt{\frac{2}{\pi}\sigma_{22}^{(1)}} - \sqrt{\frac{2}{\pi}(\sigma_{22}^{(2)}}\right) - \frac{1}{2}\left(\frac{2}{\pi}(\sigma_{22}^{(1)} - \sigma_{22}^{(2)})\right)$$
$$+y_3\left(\sqrt{\frac{2}{\pi}\sigma_{33}^{(1)}} - \sqrt{\frac{2}{\pi}(\sigma_{33}^{(2)}}\right) - \frac{1}{2}\left(\frac{2}{\pi}\sigma_{33}^{(1)} - \sigma_{(33)}^{2})\right) + \ldots + y_{10}\left(\sqrt{\frac{2}{\pi}\sigma_{1010}^{(1)}} - \sqrt{\frac{2}{\pi}\sigma_{1010}^{(2)}}\right) -$$
$$\frac{1}{2}\left(\frac{2}{\pi}(\sigma_{1010}^{(1)} - \sigma_{1010}^{(2)})\right) > 0$$

We therefore allocated an observation vector X to π_1 if

$$\sum_{i=1}^{10}\left[y_i\left(\sqrt{\sigma_{ii}^{(1)}} - \sqrt{\sigma_{ii}^{(2)}}\right) - \frac{1}{2}\sqrt{\frac{2}{\pi}}(\sigma_{ii}^{(1)} - \sigma_{ii}^{(2)})\right] > 0. \tag{4.4}$$

where $Y = |X|$, i.e $y_1 = |x_1|, y_2 = |x_2|, ..., y_{10} = |x_{10}|$.

Otherwise to π_2.

After simplifications of the above equation, the new classification rule was obtained as;

Assign $|X|$ to π_1 if

$$y_1\left(\sqrt{\sigma_{11}^{(1)}} - \sqrt{\sigma_{11}^{(2)}}\right) + y_2\left(\sqrt{\sigma_{22}^{(1)}} - \sqrt{\sigma_{22}^{(2)}}\right) + y_3\left(\sqrt{\sigma_{33}^{(1)}} - \sqrt{\sigma_{33}^{(2)}}\right)$$
$$+ \ldots$$
$$+ y_{10}\left(\sqrt{\sigma_{1010}^{(1)}} - \sqrt{\sigma_{1010}^{(2)}}\right) \geq -1.2368$$

Otherwise to π_2.

The derived function using the AEDC method was used to compute the discriminant scores as shown below.

Table 4.11: Discriminant scores using AEDC method for $n_1 = n_2$

$D_M < -1.24$	-1.65	-2.56	-1.25	-1.25	-3.25	**-1.18**	-1.72	-2.04	**0.32**	-2.15
	-1.40	**-1.16**	**-0.99**	-1.30	-1.37					
$D_D > -1.24$	-1.07	**-3.29**	2.85	5.38	6.29	2.82	2.82	4.09	1.76	2.74
	6.11	5.12	7.76	5.43	3.57					

With reference to Table 4.11, four (4) observations from the Monozygotic group were found to be misclassified and one (1) being misclassified from the Dizygotic group based on their scores/their Euclidean distances. The correct classification rate based on the outcome from the discriminant scores was discovered to be approximately 83 percent.

4.2.5 Performance Evaluation of the Methods for $n_1 = n_2$

The performance of the Bartlett and Please method (BPM), the Bayesian Posterior Probability approach, the QDF and the AEDC methods were evaluated based on the Cross-Validation (CV) or Leaving-One-Out error rate (LOO) and the Balance Error Rate (BER). The Table below, provides the output that summarises the error rates of the four methods for comparison.

Table 4.12: Error Rates estimates of the four methods for $n_1 = n_2$

Classification Methods	**Error Rates**		**Mean**
	CV	**BER**	
Bartlett and Please Method	0.233	0.133	0.183
The Bayesian Rule	0.066	0.233	0.149
The QDF Approach	0.333	0.233	0.283
The AEDC	0.166	0.166	0.166

From Table 4.12, we found out that, the margin of the mean error rates ranges from approximately 15 percent to approximately 28 percent. From the application of the four main methods based on equal prior probabilities, it was observed that two main error estimators namely cross validation (CV) and the Balanced Error Rates (BER) were used in estimating the errors incurred based on the misclassified observations. According to Johnson and Wichern, 2007, the best performing function or method is judged by the one with the least estimate of error rates or mean error rates. Comparatively, it was seen from Table 4.12 that, the classification method with the least estimate of Cross validation error rate was the Bayesian Classifier using posterior probabilities with the error rate of 0.067, followed by the AEDC method with CV error rate of 0.166, with the QDF approach recording the highest CV error rate of 0.333. Their corresponding BER also gave quiet similar results with the Bartlett and Please method recording the least error rate of 0.133, with the highest by recorded by both the Bayesian rule method and the QDF approach with BER of 0.233. Generally, inferences were drawn based on the mean error rates recorded by each of the four methods. it is clearly indicated from the various methods based on their mean error rates that, the Bayesian Classifier on posterior probabilities performs better in classifying the twin pair observations into their respective groups with a recorded least mean error rate of 0.149. In other words, the Bayesian classifier provided better maximum separation between the two groups as compared to the remaining methods.

The mean error estimate of the AEDC method of 0.166 was quiet observed to be closed to the BPM, performing much more better than two methods with the exception of the BPM. Conclusively, The QDF approach obtained the highest mean error rate of 0.283 indicating a poor performance in the aspect of classifying the twin pair observations into their respective groups. In other words, the Expected misclassification Error rate for QDF was observed to be higher than the remaining three classification methods. This results shows partial conformity with the research based study of Ganeslingam et al, where in their study they compared the performance of the QDF and the AEDC method and concluded that, the AEDC method outperforms the QDF method.

4.3 Case 2: Classification with Unequal Prior Probabilities

Several sample selection were used to derive the classification rules for classifying the twin pair observations into their respective groups for the assumption of unequal prior probabilities ($p_1 \neq p_2$) and equal misclassification, $c(1|2) = c(2|1)$. The sample selection were taking in multiples of five (5). The varying degrees of sample selections used in the groups ratio order of Monozygotic: Dizygotic were; $1:2$, $1:3$, $1:4$ and the ratio order of Dizygotic: Monozygotic were $1:2$, $1:3$, $1:4$. This section also employed the four methods earlier discussed in studying the effect of unequal prior probabilities and equal misclassification cost on the classification rules of each of the methods.

4.3.1 Bartlett and Please Classification Method for $n_1 \neq n_2$, $(n_1 : n_2)$

We started the varying sample selections by taking five (5) sample observations from the Monozygotic group($n_1 = 5$) and Ten (10) from the Dizygotic group ($n_2 = 10$). The estimates for ρ_1,ρ_2, ρ and σ^2 were obtained as 0.0967, 0.2194, 0.16 and 6.8933 respectively as shown in the Table below.

Table 4.13: Table of estimates of parameters for $n_1 \neq n_2$

Sample Ratios	σ^2	ρ_1	ρ_2	ρ	cut-off
$n_1 : n_2$	6.8933	0.0967	0.2194	0.16	14.96

The estimate of the common correlation coefficient ρ was used to obtain the value of the boundary for misclassification/cut-off point. From the boundary of misclassification formula given as

$$\frac{(1-\rho)p\log\sigma^2}{1-\sigma^{-2}} - 2c$$

where $c = \frac{q_2}{q_2}$, q_1 and q_1 are the prior probabilities for Monozygotic and Dizygotic twin groups respectively. That is q_1 and q_2 were estimated as 0.33 and 0.67 respectively. Hence $c = 2$. Therefore the boundary of misclassification was derived as 14.96. The Figure below presents the scatter plots for the total 15 observations. (*That is 5 from Monozygotic and 10 from Dizygotic*). However from the plot and other computations the linear discriminant function was obtained for this particular case.

$$Z_1 - 0.0655Z_2 \tag{4.5}$$

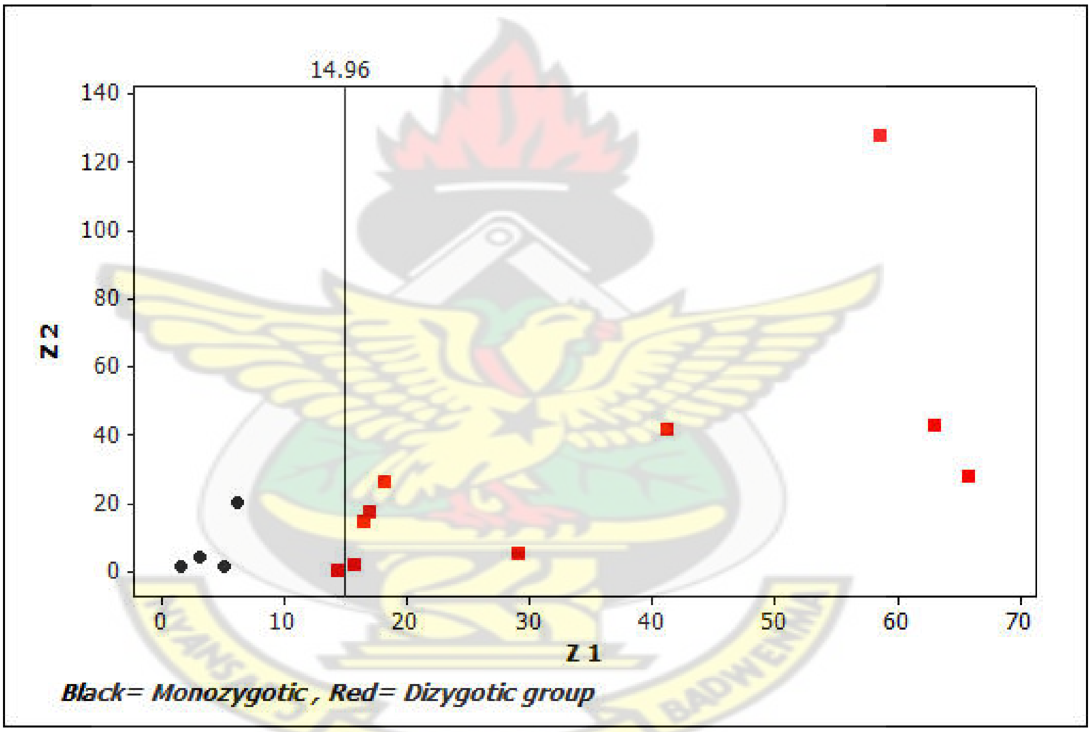

Figure 4.2: Scatter plot of observations from the sample selection, $n_1 = 5 : n_2 = 10$

From Figure 4.2 above, it was observed that, none of the five selected observations from the Monozygotic group was misclassified with only one observation being misclassified from the Dizygotic group. In obtaining a better and maximum separation the discriminant scores are being summarised in Table 4.14.

where D_M and D_D are the discriminant scores for the Monozygotic group and

Table 4.14: Discriminant scores for classifying observations as both twin groups

D_M	1.53	5.01	1.57	2.81	4.95					
D_D	15.79	28.74	16.52	50.15	60.19	**14.36**	63.99	38.40	15.58	15.51

the Dizygotic group.

The discriminant scores in Table 4.14 above observed only one misclassified observation from the Dizygotic twin group. Comparatively, the unequal prior probability provides maximum separations than that of equal prior probability.

The confusion matrix after the classification of the observation using the derived function are being summarised in Table 4.15 below. The table below gives the confusion matrix of the misclassified observations.

Table 4.15: Confusion Matrix for Bartlett and Please method with common ρ (5:10) samples

	Classified as		
True Pop.	π_1	π_2	*Total*
π_1	5.0	0.0	5.0
π_2	1.0	9.0	10.0
Total	6.0	9.0	15.0

From the table, the proportion of correct classification was observed to be 93 percent.

Taking the sample ratios $n_1 = 1 : n_2 = 3$

Five (5) samples from the Monozygotic twins and 15 from the Dizygotic twin groups were taken. The unequal prior probabilities from the Monozygotic and the Dizygotic twin groups were computed to be 0.25 and 0.75 respectively. Based on these and other estimates the cut-off points were also derived as shown in the scatter plot below. The plot describes the distribution of the Z_1 and Z_2 values. As observed from Figure 4.3, none of the twin pair observations from both twin groups were misclassified, giving a 100 percent correct classification of the observations. Based on the plot, a linear discriminant function was derived as shown below;

$$Z_1 - 0.066Z_2 \tag{4.6}$$

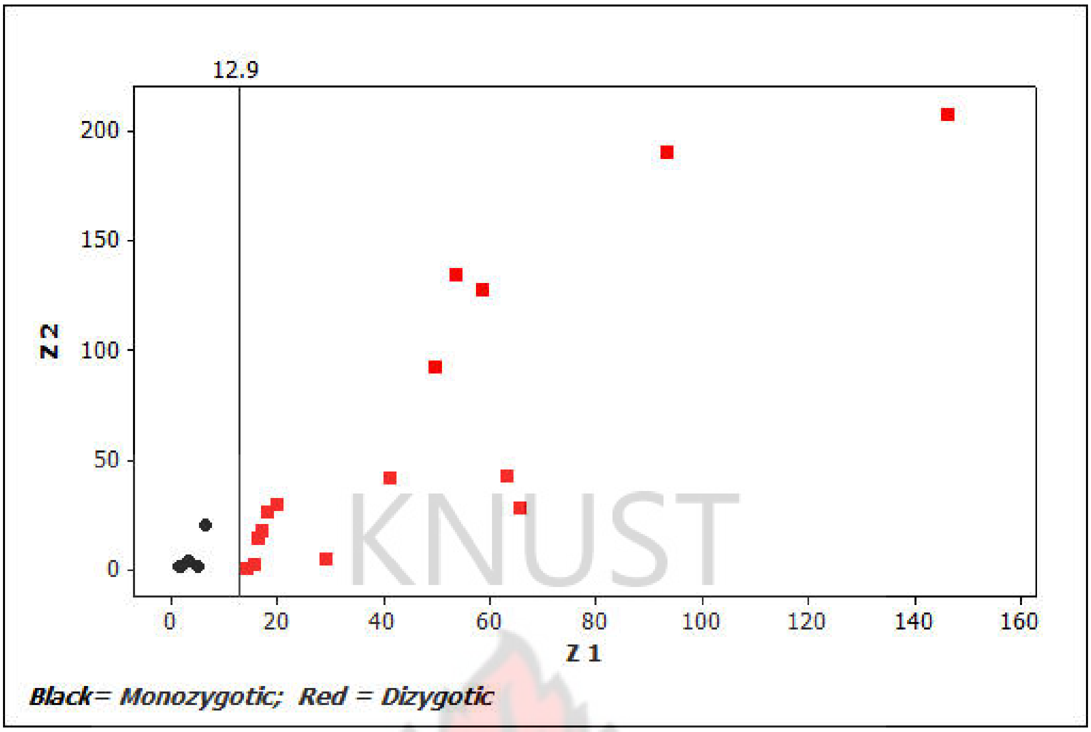

Figure 4.3: Scatter plot of observations from the sample ratio, 1 : 3

The linear function was used to compute the discriminant scores for the sample ratio, 1:3 as shown below. We also observed no misclassified observation from the scores. This results indicates a high significant effect of the unequal prior probabilities in this case on the classification rule. That is the classification rule was shifted much more backwards as compared to the equal prior probability case.

Table 4.16: Discriminant scores for classifying observations as both twin groups for ratio 1:3

D_M	1.53	5.01	1.57	2.81	4.94					
D_D	15.79	28.74	16.52	50.08	60.17	14.36	63.98	38.38	15.58	15.51
	132.50	44.85	80.76	43.42	18.04					

Taking the sample ratio, $n_1 = 1 : n_2 = 4$

Another sample ratio of 5 Monozygotic ($n_1 = 5$) observations and 20 Dizygotic ($n_2 = 20$) observations were taking. Their corresponding prior probabilities were recorded as ($p_1 = 0.20$) and ($p_2 = 0.80$). The values of the transformed observational vectors into Z_1 and Z_2 were plotted and a linear discriminant function was derived using the common population correlation coefficient (ρ) and the overall variance (σ^2). Below summarises the distribution of the observations in the form of Z_1 and Z_2 as used by Bartlett and Please (1963).

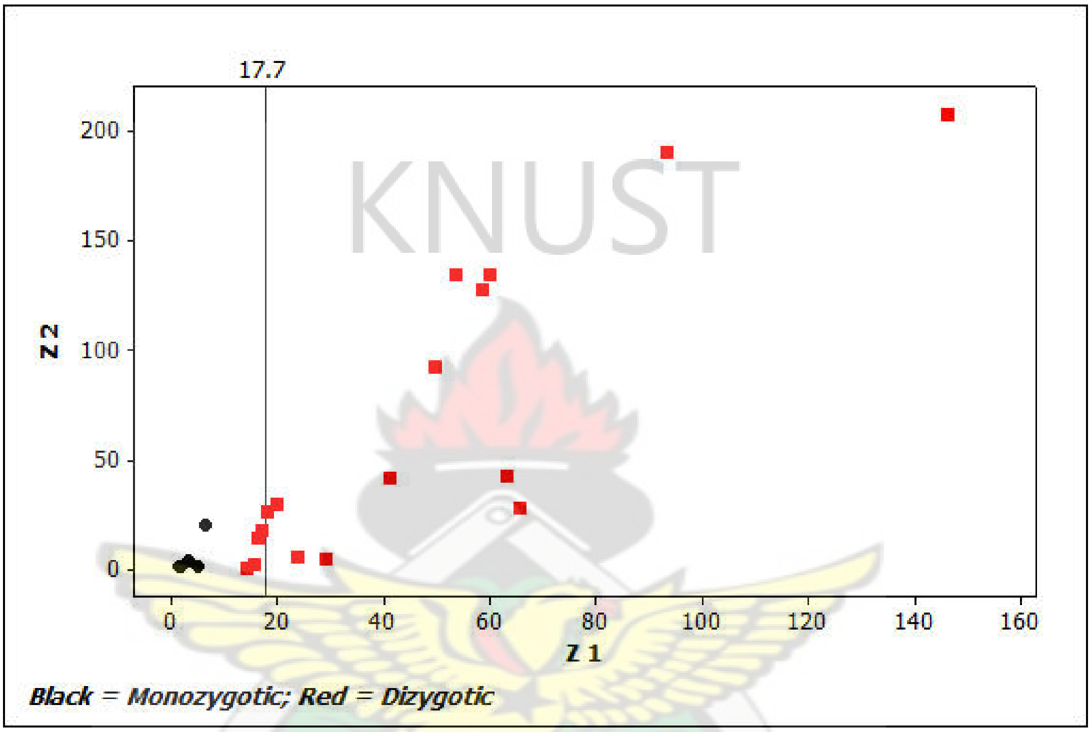

Figure 4.4: Scatter plot of observations from the sample ratio, 1 : 4

The linear discriminant function and the classification rule was obtained in this case as ; classify x_1 to π_1 if

$$Z_1 - 0.0709Z_2 \leq 17.7 \tag{4.7}$$

After substituting the observations (i.e the values of Z_1 and Z_2) into the linear equation above, the discriminant scores were also obtained as shown Table 4.17. From both the scores and the scatter plot, we observed five misclassified observations from the Dizygotic twin group with no misclassification from the Monozygotic group. The cut-off point in this case, was shifted forward, causing more observations in the Dizygotic group to be misclassified, hence there exist a significant effect of unequal prior probabilities on the classification rule.

Taking the sample ratios in the order $(n_2 : n_1)$***,i.e*** $(n_2 = 1 : n_1 = 2)$.

The sample size selection was alternated by keeping the first sample selected from the Dizygotic group ($n_2 = 5$) constant and then increasing the sample sizes of the Monozygotic group (n_1) in multiples of 5 to the 4th ratio.

Ten (10) observations of Monozygotic twins were sampled as well as five (5) samples from the Dizygotic twins.

Figure 4.5 below shows the plot of the values of both twin groups after the computation of their respective Z_1 and Z_2 values. From the scatter plot, it is clearly indicated that, only one observation belonging to the Monozygotic twin was misclassified after a little shift of the boundary of misclassification (cut-off) based on the unequal prior probabilities. However the linear function below was derived based on the scatter plot above and discriminant scores generated based on this function also misclassified the same number of observation (i.e only one observation from the Monozygotic group was misclassified).

Below gives the classification rule for classifying x_1 to π_1, otherwise to π_2.

$$Z_1 - 0.0702Z_2 \leq 15.90 \tag{4.8}$$

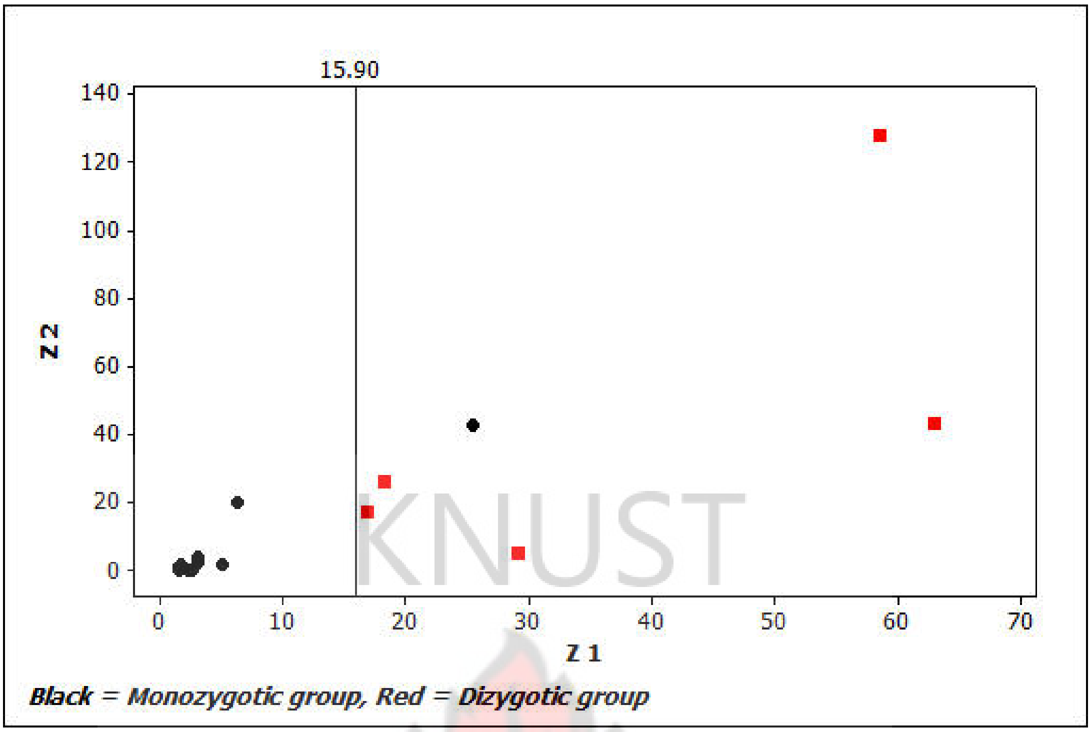

Figure 4.5: Scatter plot of observations from both Mono. and Dizy. groups for $n_2 = 1 : n_1 = 2$

Taking sample ratio $n_2 = 1 : n_1 = 3$

Five (5) and fifteen (15) samples were taken respectively from both the Dizygotic and the Monozygotic twin groups. The prior probability of the Monozygotic group was 0.75 and that of the Dizygotic to be 0.25. The distribution of the observations are summarised in the scatter plot below.

And the classification rule in this case was; classify x_1 to π_1 otherwise to π_2 if

$$Z_1 - 0.023Z_2 \leq 18.91 \tag{4.9}$$

Evidence from Figure 4.6 and the discriminant scores in the Appendix sections B shows two misclassified observations and one misclassified observation from the Dizygotic and the Monozygotic groups respectively. Hence the rate of error of misclassification based on these misclassified observations was obtained as 15 percent, indicating that, the above classification rule provided 75 percent correct classification of their observations into their true populations.

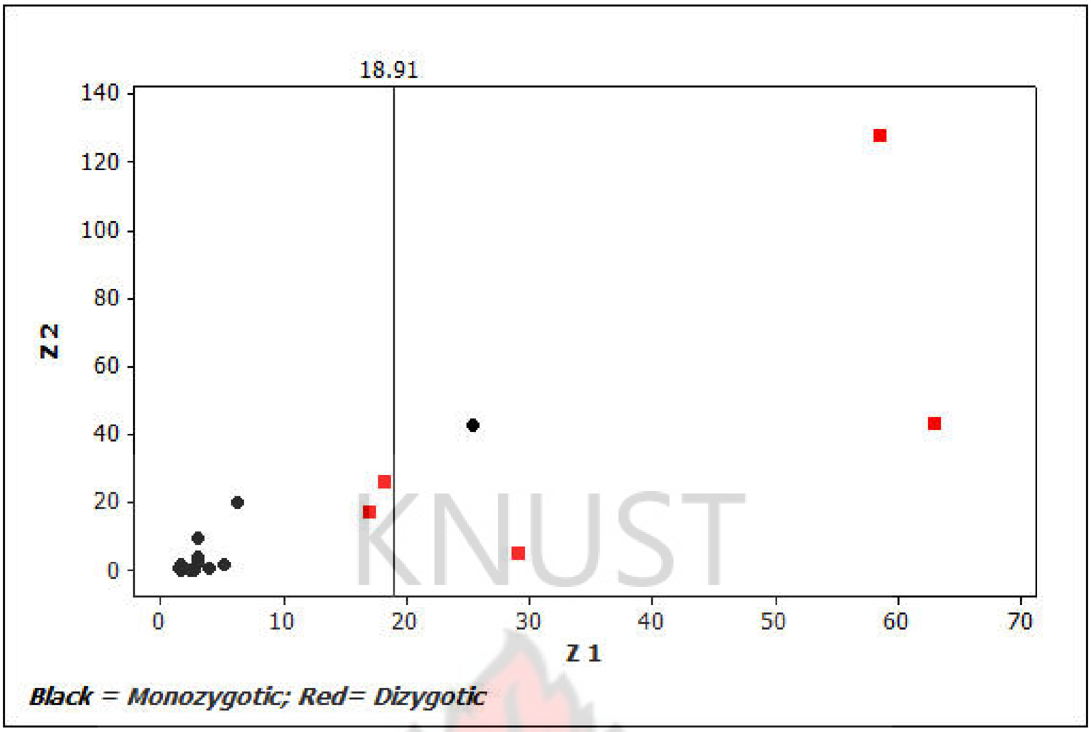

Figure 4.6: Scatter plot of observations from both Mono. and Dizy. groups for $n_2 = 1 : n_1 = 3$

Taking the sample ratio $n_2 = 1 : n_1 = 4$

Below gives the scatter plot of the separation of the twin observations into their respective groups.

The linear function and classification rule obtained for providing maximum separation between the two twin groups in the case of unequal prior probability was obtained as shown;

We classified the observations x_1 to π_1 when

$$Z_1 - 0.0158Z_2 \leq 17.99 \tag{4.10}$$

Two (2) observations from the Dizygotic twins were misclassified, whilst three (3) from the Monozygotic group were misclassified based on the above classification rule.

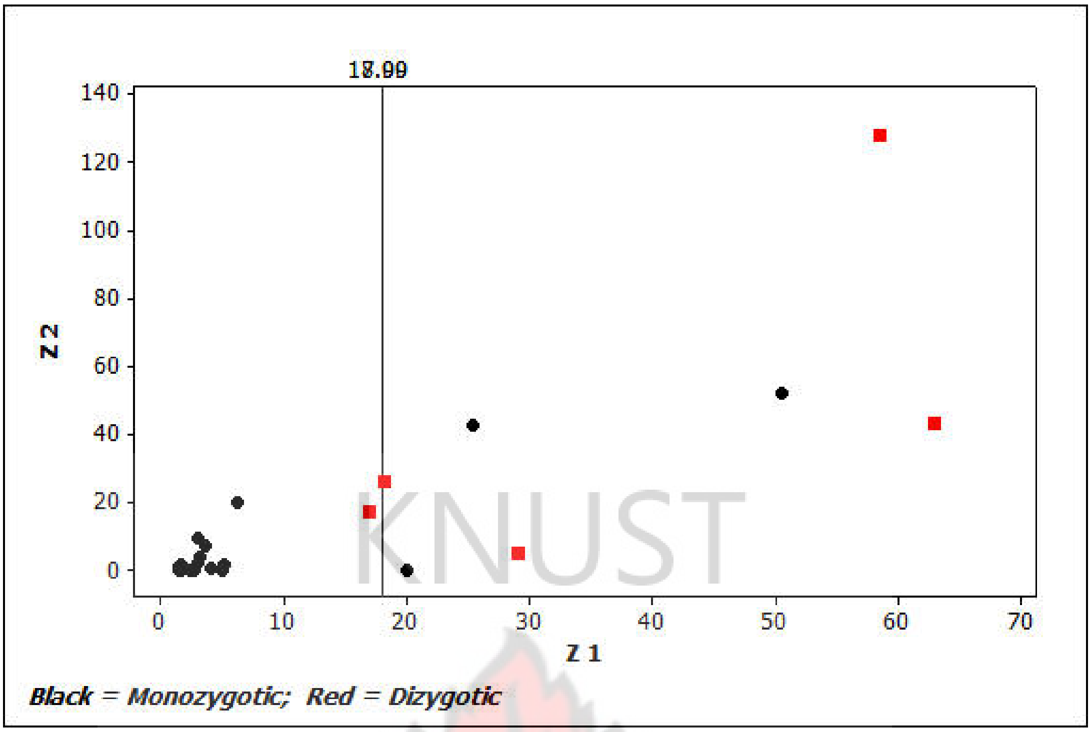

Figure 4.7: Scatter plot of observations from both Mono. and Dizy. groups for $n_2 = 1 : n_1 = 4$

Table 4.17: Summary table for discriminant scores for the three sampling ratios in the order $n_1 : n_2$

$n_1 : n_2$	**1: 2**		**1:3**		**1:4**	
	$D_M <$ 14.96	$D_D \geq 14.96$	$D_M < 12.90$	$D_D \geq 12.90$	$D_M < 17.7$	$D_D \geq 17.7$
	1.53	15.79	1.53	15.79	1.52	**15.69**
	5.01	28.74	5.01	28.74	5.01	28.72
	1.57	16.52	1.57	16.52	1.56	**16.38**
	2.81	50.15	2.81	50.08	2.79	49.46
	4.95	60.19	4.94	60.17	4.84	59.96
		14.36		14.36		14.36
		63.99		63.98		63.84
		38.40		38.38		38.18
		15.58		15.58		15.58
		15.51		15.51		15.43
				132.50		131.49
				44.85		44.19
				80.76		79.83
				43.42		42.97
				18.04		17.90
						20.76
						43.98
						19.67
						59.67
						34.88

Table 4.18: summary table for discriminant scores for the three sampling ratios in the order, $n_2 : n_1$

	1: 2		**1:3**		**1:4**	
	$D_M <$ 15.90	$D_D \geq 15.90$	$D_M <$ 18.91	$D_D \geq$ 18.91	$D_M <$ 17.99	$D_D \geq$ 17.99
	1.52	**15.70**	1.58	**16.53**	1.59	**16.66**
	5.01	28.72	5.09	28.97	5.10	29.01
	1.56	16.40	1.63	**17.63**	1.64	**17.83**
	2.79	49.55	2.99	55.59	3.02	56.51
	4.85	59.99	5.81	62.03	5.96	62.34
	1.60		1.60		1.61	
	2.42		2.42		2.42	
	2.88		2.99		3.02	
	22.39		24.42		24.74	
	2.70		2.70		2.771	
			4.00		4.00	
			2.82		2.89	
			1.65		1.66	
			2.51		2.51	
			1.48		1.49	
					1.57	
					3.98	
					6.87	
					2.31	
					4.89	

4.3.2 The Bayesian Posterior Probability Approach for $n_1 \neq n_2$

The posterior probability approach for classification using Bayes rule was applied when the prior probabilities were assumed to be unequal. The sample ratios in the order of $n_1 : n_2$ and $n_2 : n_1$ were used as already spelt out in the above section.

Taking the sample ratio, $n_1 = 1 : n_2 = 2$

Ten twin pairs of observations were sampled from the Dizygotic twin group as well as five (5) from the Monozygotic. The main idea of these varying samples from each respective twin group was to find out the effect of unequal prior probabilities on the classification of the observations into their respective groups based on their scores. Table 4.19 below summarises the Bayes rule of posterior probabilities. By rule, an observation is classified into their respective groups based on the maximum posterior probability value they assumes.

Table 4.19: Likelihoods and Posterior probabilities for unequal group prior probability for $n_1 = 1 : n_2 = 2$

$P(j\|x_M)$	$P(j'\|x_M)$	$P(j'\|x_D)$	$P(j\|x_D)$
6.69E-01	3.31E-01	**3.39E-01**	**6.61E-01**
8.72E-01	1.28E-01	6.79E-01	3.21E-01
6.74E-01	3.26E-01	7.42E-01	2.58E-01
6.05E-01	3.95E-01	9.26E-01	7.40E-02
8.34E-01	1.66E-01	**3.19E-01**	**6.81E-01**
		6.05E-01	3.95E-01
		9.90E-01	9.95E-03
		3.73E-01	**6.27E-01**
		8.78E-01	1.22E-01
		6.11E-01	3.89E-01

From the table posterior probabilities above, we observed that none of the five (5) selected observations from the Monozygotic twin group was misclassified. Three (3) out of the ten selected observations from the Dizygotic group were misclassified. In all, 80 percent of correct classification rate was realised in this case, with the remaining 20 percent accounting for misclassification rate.

Taking the sample ratio $n_1 = 1 : n_2 = 3$

Bayes rule of posterior probability was applied when unequal sample size selections were taking from both groups. Five (5) observations were taken from the Monozygotic twins as against fifteen (15) observations from the Dizygotic twins. The table below presents the output of likelihoods and posterior probabilities based on the groups prior probabilities. From Table 4.20, observations made indicates that, none of the Monozygotic observations was misclassified. However six (6) observations were misclassified and hence the error rate of misclassifications becomes 30 percent.

Table 4.20: Posterior Probabilities for the sample ratio $n_1 = 1 : n_2 = 3$

$P(j\|x_M)$	$P(j'\|x_M)$	$P(j'\|x_D)$	$P(j\|x_D)$
5.76E-01	4.24E-01	9.08E-01	9.21E-02
9.10E-01	9.02E-02	7.26E-01	2.74E-01
7.36E-01	2.64E-01	6.80E-01	3.20E-01
6.23E-01	3.77E-01	9.86E-01	1.38E-02
7.90E-01	2.10E-01	**2.70E-01**	**7.30E-01**
		4.81E-01	**5.19E-01**
		9.11E-01	8.85E-02
		6.96E-01	3.04E-01
		5.52E-01	4.48E-01
		1.60E-01	**8.40E-01**
		2.52E-01	**7.48E-01**
		8.81E-01	1.19E-01
		7.31E-01	2.69E-01
		3.87E-01	**6.13E-01**
		2.94E-01	**7.06E-01**

Taking the sample ratio, $n_1 = 1 : n_2 = 4$

The sampling selection continued by keeping the 5 observations sampled from the Monozygotic group constant and taking an increased sample of 20 from the Dizygotic group. The likelihoods and the posterior probabilities for this particular case were obtained and summarised in Table 4.21. However the classification rule was obtained based on the observation with the highest posterior probabilities and likelihoods. The table below presents the likelihoods and posterior probabilities for the sample selection $n_1 = 1 : n_2 = 4$.

Table 4.21: Posterior Probabilities for the sample selection $n_1 = 1 : n_2 = 4$

$P(j\|x_M)$	$P(j'\|x_M)$	$P(j'\|x_D)$	$P(j\|x_D)$
6.86E-01	3.14E-01	9.18E-01	8.16E-02
9.17E-01	8.29E-02	6.73E-01	3.27E-01
7.65E-01	2.35E-01	6.47E-01	3.53E-01
6.44E-01	3.56E-01	9.85E-01	1.46E-02
8.57E-01	1.43E-01	**2.54E-01**	**7.46E-01**
		3.26E-01	**6.74E-01**
		8.75E-01	1.25E-01
		7.93E-01	2.07E-01
		5.09E-01	4.91E-01
		1.13E-01	**8.87E-01**
		2.08E-01	**7.92E-01**
		9.13E-01	8.72E-02
		7.27E-01	2.73E-01
		3.29E-01	**6.71E-01**
		2.68E-01	**7.32E-01**
		2.73E-01	**7.27E-01**
		2.78E-01	**7.22E-01**
		1.00E+00	1.51E-04
		9.68E-01	3.20E-02
		7.42E-01	2.58E-01

From Table 4.21, the posterior probabilities used as a classification rule was hugely affected by the unequal prior probabilities as the sample size of the Dizygotic group increased to 20. As a result of this, we observed eight (8) misclassified observations from the Dizygotic group. No observation was misclassified from the Monozygotic twin group. Hence the classification rule affects the population with the larger sample size.

Taking the sample ratios in the order of $n_2 = 1 : n_1 = 2$

We alternated the sample selections by keeping the first selected sample from the Dizygotic group (i.e $n_2 = 5$) constant and increasing the sample sizes of the Monozygotic group in multiples of five (5).

We first considered the sample ratio, $n_2 = 1 : n_1 = 2$, that is sampling 5 twin observations from the Dizygotic group and 10 twin observations from the Monozygotic twin group. The classification rule for classifying observations into either of the two populations for this particular method is based on likelihoods obtained from the respective density functions and their posterior probabilities.

Based on the evidences provided in Table 4.22, the posterior probability values detected two (2) misclassified observation out of the ten sampled, from the Monozygotic group and one (1) out of five (5) sampled from the Dizygotic twin group. The table below summarises the likelihoods and posterior probabilities of 5 observations from the Dizygotic group and 10 from the Monozygotic group.

Table 4.22: Likelihoods and Posterior probabilities for sample ratio, $n_2 = 1 : n_1 = 2$

$P(j\|x_M)$	$P(j'\|x_M)$	$P(j'\|x_D)$	$P(j\|x_D)$
7.07E-01	2.93E-01	**3.39E-01**	**6.61E-01**
8.19E-01	1.81E-01	6.79E-01	3.21E-01
5.58E-01	4.42E-01	7.42E-01	2.58E-01
4.89E-01	**5.11E-01**	9.26E-01	7.40E-02
6.78E-01	3.22E-01	**3.19E-01**	**6.81E-01**
4.44E-01	**5.56E-01**		
7.51E-01	2.49E-01		
8.45E-01	1.55E-01		
9.51E-01	4.89E-02		
8.88E-01	1.12E-01		

Taking the sample selection $n_2 = 1 : n_1 = 3$

The sample selection made up of 5 and 15 observations from the Dizygotic and Monozygotic populations were used to derive a classification rule under unequal prior probabilities using posterior probabilities of the individual twin observations. The table below presents the outcome of the posterior probabilities and likelihoods of the observations. From Table 4.23, one observation from the Dizygotic group was found to be misclassified.

Table 4.23: Likelihood and Posterior Probabilities for sample ratio, $n_2 = 1 : n_1 = 3$

$P(j\|x_M)$	$P(j'\|x_M)$	$P(j'\|x_D)$	$P(j\|x_D)$
8.80E-01	1.20E-01	9.97E-01	3.16E-03
9.87E-01	1.30E-02	7.18E-01	2.82E-01
8.53E-01	1.47E-01	5.71E-01	4.29E-01
8.75E-01	1.25E-01	9.96E-01	3.53E-03
9.63E-01	3.72E-02	**4.98E-02**	**9.50E-01**
8.77E-01	1.23E-01		
9.34E-01	6.57E-02		
9.60E-01	3.99E-02		
6.87E-01	3.13E-01		
9.23E-01	7.71E-02		
9.17E-01	8.28E-02		
8.40E-01	1.60E-01		
8.85E-01	1.15E-01		
9.66E-01	3.35E-02		
9.26E-01	7.37E-02		

Sample ratio $n_2 = 1 : n_1 = 4$

Table 4.24 summarises the likelihoods and posterior probabilities for this particular sampling ratio. That is 5 and 20 twin observations were sampled from both the Dizygotic and Monozygotic twin groups. From the table, one twin pair observation from each group was found to be misclassified. This results based on this sampling ratio provides better maximum separation than the already explained sampling ratios.

Table 4.24: Likelihood and Posterior Probabilities for the sample ratio, $n_2 = 1 : n_1 = 4$

$P(j\|x_M)$	$P(j'\|x_M)$	$P(j'\|x_D)$	$P(j\|x_D)$
8.82E-01	1.18E-01	9.94E-01	6.48E-03
9.83E-01	1.65E-02	7.00E-01	3.00E-01
8.45E-01	1.55E-01	5.45E-01	4.55E-01
8.92E-01	1.08E-01	9.93E-01	7.08E-03
9.52E-01	4.83E-02	**7.49E-02**	**9.25E-01**
8.81E-01	1.19E-01		
9.32E-01	6.80E-02		
9.47E-01	5.27E-02		
5.60E-01	4.40E-01		
9.18E-01	8.16E-02		
9.06E-01	9.36E-02		
8.42E-01	1.58E-01		
8.84E-01	1.16E-01		
9.60E-01	4.05E-02		
9.27E-01	7.33E-02		
4.85E-02	**9.51E-01**		
9.86E-01	1.38E-02		
9.48E-01	5.19E-02		
7.12E-01	2.88E-01		
5.73E-01	4.27E-01		

4.3.3 The QDF Approach for $n_1 \neq n_2$

Based on the sample selections of 5 and 10 from both the Monozygotic and the Dizygotic twin groups, 0.33 and 0.67 prior probabilities for Monozygotic and Dizygotic group respectively were obtained. Based on these prior probabilities the QDF in this case was obtained as shown below and the equation was used to compute the discriminant scores in this case.

$$\left\{(x-\mu_1)^\tau \Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)^\tau \Sigma_2^{-1}(x-\mu_2)\right\} < \ln\left\{\frac{|\Sigma_1|}{|\Sigma_2|}\right\} + \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{P_2}{P_1}\right) \tag{4.11}$$

Hence assuming equal misclassification cost and unequal prior probabilities i.e 0.33 for Monozygotic and 0.67 for Dizygotic group, the above equation can be rewritten in this particular case as

$$\left\{(x-\mu_1)^\tau \Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)^\tau \Sigma_2^{-1}(x-\mu_2)\right\} < 1.658 \tag{4.12}$$

From the discriminant scores as shown in Table 4.25, we observed two (2) misclassified observations from the Dizygotic group and no observation was misclassified from the Monozygotic twin group. Hence the proportion for correct classification was computed as approximately 87 percent.

Taking sample ratio, $n_1 = 1 : n_2 = 3$

The sampling ratio of 5 Monozygotic and 15 Dizygotic twin observations were used to derive the quadratic discriminant function, based on their respective prior probabilities and equal misclassification cost assumption. The classification rule we derived was; classify an observation x_1 as π_1 if ;

$$\left\{(x-\mu_1)^{\tau}\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)^{\tau}\Sigma_2^{-1}(x-\mu_2)\right\} < 1.204 \tag{4.13}$$

The above function was used to calculate the discriminant scores of the twin observations for the provision of maximum separation. Five and one observations were misclassified from the Dizygotic and Monozygotic groups respectively. Hence the probability of correct classification was 0.70. (*see Table 4.25*).

Taking sample ratio, $n_1 = 1 : n_2 = 4$

The quadratic discriminant scores obtained based on the ratio of the determinants of the covariance matrices and Mahalanobis distances was obtained as; classify x_1 to π_1 otherwise to π_2 if

$$\left\{(x-\mu_1)^{\tau}\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)^{\tau}\Sigma_2^{-1}(x-\mu_2)\right\} < 1.585 \tag{4.14}$$

The proportion of correct classification based on four misclassified observations from the Dizygotic group was 0.84.

The table below summarises the discriminants scores using the QDF obtained from the three sampling ratios explained above. With reference from Table 4.25, the discriminant scores for the sample ratio 1 : 2 recorded two misclassified observations from the Dizygotic group. One (1) and five (5) observations from both the Monozygotic and Dizygotic twin groups were found to be misclassified from their groups respectively for the sample ratio of 1 : 3. For the sample ratio of 1 : 4, we observed 3 and 6 twin pair observations being misclassified from the Monozygotic and Dizygotic twin groups. This classification rules derived under the QDF method, discriminate similarly to that of the Bayesian Classifier, in the sense that, the number of misclassified observations increases, as the sample size selection for the Dizygotic group increases.

Table 4.25: Discriminant scores for the three sample ratios

Sample Ratios	**1 : 2**		**1 : 3**		**1 : 4**	
Disc. Scores	D_M <1.65	$D_D \geq 1.65$	DM <1.20	$D_D \geq 1.20$	D_M <1.58	$D_D \geq$ 1.58
	-0.786	-0.429	1.177	6.381	0.855	7.256
	-4.780	3.026	-2.828	3.743	-2.392	3.85
	-0.425	1.885	-0.257	3.301	0.054	3.622
	-1.636	7.692	0.783	10.338	1.226	10.837
	-1.744	0.026	-0.858	-0.194	-1.163	0.258
		2.531		1.644		0.965
		4.589		6.457		6.302
		6.765		3.453		5.102
		2.304		2.213		2.482
		-2.845		-1.515		-1.701
				-0.381		-0.261
				5.808		7.11
				3.789		4.371
				0.868		0.991
				0.035		0.402
						3.456
						4.756
						3.112
						1.876
						4.476

Taking the sample ratio in the order, $n_2 : n_1$

As already explained in the previous sections, we alternated the sample sizes of the two groups to assume unequal prior probabilities and study the behaviour of the resulting classification rules derived in each case based on effect unequal prior probabilities.

The QDF obtained under the sample ratio $n_2 = 1 : n_1 = 2$ was;

$$\left\{(x-\mu_1)^{\tau}\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)^{\tau}\Sigma_2^{-1}(x-\mu_2)\right\} < -2.422 \tag{4.15}$$

From Table 4.26, the proportion of correct classification was recorded as 0.60. For the sample ratio $n_2 = 1 : n_1 = 3$, the function was obtained as shown;

$$\left\{(x-\mu_1)^{\tau}\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)^{\tau}\Sigma_2^{-1}(x-\mu_2)\right\} < -2.574 \tag{4.16}$$

The proportion of correct classification for this sampling ratio with reference to Table 4.26 was observed to be 0.40. And that of the ratio, $n_2 = 1 : n_1 = 4$ was derived as;

$$\left\{(x-\mu_1)^{\tau}\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)^{\tau}\Sigma_2^{-1}(x-\mu_2)\right\} < -2.605 \tag{4.17}$$

The proportion of correct classification was 0.52, with the remaining representing the misclassification rate. The discriminant scores obtained based on the three QDF in this case are reported in Table 4.26.

Table 4.26: Discriminant scores for the sample ratios, $n_2 = 1 : n_1 = 2$

Sample Ratios	**1 : 2**		**1 : 3**		**1 : 4**	
Disc. Scores	DM <-2.42	$DD \geq -2.42$	DM < -2.57	$DD \geq$ -2.57	DM < -2.60	$DD \geq$ -2.60
	-0.536	14.666	-1.078	14.666	-1.175	12.921
	-5.851	4.948	-5.769	4.948	-5.316	4.552
	-0.691	4.761	-0.621	4.761	-0.537	3.218
	-0.726	15.191	-0.984	15.191	-1.364	12.742
	-3.811	-3.115	-3.602	-3.115	-3.104	-2.173
	-0.161		-1.033		-1.147	
	-2.596		-2.41		-2.379	
	-3.95		-3.467		-2.924	
	-0.524		1.327		2.374	
	-2.495		-2.064		-1.987	
			-1.914		-1.684	
			-0.402		-0.487	
			-1.18		-1.205	
			-3.829		-3.475	
			-2.163		-2.219	
					-2.111	
					-3456	
					-3.993	
					-4.980	
					-2.212	

Four (4) and one (1) observations from Table 4.26 were misclassified from their respective populations, i.e Monozygotic and Dizygotic twin groups respectively for the sample ratio, 1:2. The number of correct classifications and misclassifications for the other sample ratios are summarised in Table 4.26.

4.3.4 Evaluating the performance of the Classification Methods for $n_1 \neq n_2$

The table below summarises the error rates of the classification methods and their respective sample ratios under each of the three methods which where applicable to deriving a classification rule under unequal prior probability situation.

Table 4.27: Error rates for the classification method under unequal prior probability case.

Classification methods	Error Rates CV	BER	Mean
Bartlett and Please			
$n_1 : n_2$			
1 : 2	0.133	0.050	0.092
1 : 3	0.100	0.000	0.050
1 : 4	0.120	0.150	0.135
$n_2 : n_1$			
1: 2	0.200	0.250	0.225
1: 3	0.200	0.200	0.200
1: 4	0.160	0.175	0.168
The Bayesian Rule			
$n_1 : n_2$			
1: 2	0.266	0.150	0.208
1: 3	0.300	0.333	0.316
1: 4	0.340	0.234	0.287
$n_2 : n_1$			
1: 2	0.133	0.040	0.086
1: 3	0.350	0.433	0.391
1: 4	0.300	0.398	0.349
The QDF Approach			
$n_1 : n_2$			
1: 2	0.440	0.250	0.345
1: 3	0.250	0.167	0.209
1: 4	0.300	0.342	0.321
$n_2 : n_1$			
1: 2	0.333	0.350	0.342
1: 3	0.150	0.300	0.225
1: 4	0.320	0.425	0.373

The table shows that generally unequal prior probabilities hugely influence the classification rules of the three methods namely Bartlett and Please, QDF and Bayesian Posterior Probability approach. The AEDC method was not included in this particular case since their functions do not include the involvement and computation of prior probabilities. From Table 4.27, it was observed that, the error estimates increases appreciably as the size of one group increases relative to another. Comparatively, all the three methods recorded almost similar error estimates in both the sample selection ratios and their corresponding alternated sampling ratio. Bartlett and Please classification method recorded the least mean error estimates as compared to the QDF and the Bayesian Classifier. The QDF recorded the highest mean error rates. For unequal prior probability and equal misclassification situation, Bartlett and Please method was observed to perform much better than the other methods for the provision of maximum separation between the two populations.

4.4 Case 3: Classification under Non-normality assumption

We introduced outliers into the working data, for the normality assumption to be violated and applied the Bayes rule of posterior probability method, the AEDC, the QDF approach and the Bartlett and Please method. The table below gives the error estimates of the classification rules for the four methods under equal and unequal prior probabilities. The main reason for contaminating the data was to study the effect of the contaminated twin observations on the classification rules. Outliers were introduced into the first five observations in each of the twin groups, one at a time and a classification rule was obtained in each case. The errors incurred in the classification of the twin pair observations are presented in Table 4.28. Generally the performances of the all the classification methods with equal and unequal prior probabilities deteriorated after the introduction of outliers into the twin data. However the mean error estimates AEDC method performed slightly better than the Bayesian Posterior Probability approach with a mean error rate of 0.381 under equal prior probabilities and 0.375 and 0.249 for the unequal prior probabilities based on the predetermined choice of sampling ratios. In other words, the AEDC method recorded the least error rate of 0.339 and hence provides maximum separation than the remaining methods under non normality. It was also observed that, The Bartlett and Please approach performed poorly under non normality assumption with an error estimates for both equal and unequal prior probabilities ranging from 0.466 to 0.667. The QDF performed appreciably better for the equal prior probability case, but the performance of the method under unequal prior probabilities was abysmal, that is with a recorded mean error rate for the two sample ratios as 0.523 and 0.625 for the sample ratios of 1:2 and 1:3 respectively. This results/findings shows some conformity with the research/study by Lachenbruch (1975), where after contaminating the twin data, he discovered that, the performance of the QDF was very poor, but the absolute linear discriminant function performed reasonably well.

Table 4.28: Evaluation of the classification methods under Non- normality

Classification methods	**Error Rates**		**Mean**
	CV	**BER**	
Bartlett and Please			
$n_1 = n_2$	0.466	0.466	0.466
$n_1 : n_2$			
1 : 2	0.800	0.850	0.825
1 : 3	0.756	0.757	0.762
1 : 4	0.699	0.861	0.780
$n_2 : n_1$			
1: 3	0.600	0.733	0.667
1: 2	0.614	0.801	0.708
1: 4	0.703	0.788	0.746
The Bayesian Rule			
$n_1 = n_2$	0.352	0.410	0.381
$n_1 : n_2$			
1: 2	0.400	0.350	0.375
1: 3	0.423	0.422	0.423
1: 4	0.478	0.317	0.398
$n_2 : n_1$			
1: 3	0.333	0.166	0.249
1: 2	0.301	0.123	0.212
1: 4	0.314	0.107	0.211
The QDF Approach			
$n_1 = n_2$	0.300	0.300	0.300
$n_1 : n_2$			
1: 2	0.446	0.600	0.523
1: 3	0.545	0.555	0.550
1: 4	0.500	0.689	0.595
$n_2 : n_1$			
1: 3	0.750	0.500	0.625
1: 2	0.800	0.607	0.704
1: 4	0.713	0.578	0.646
The AEDC			
$n_1 = n_2$	0.333	0.345	0.339

Chapter 5

Summary of Results, Conclusion and Recommendation

This Chapter outlines the summary results gathered in the entire analyses, conclusions drawn based on the study objectives and recommendations which will aid in further studies.

5.1 Findings and Conclusions

This study was specifically aimed at evaluating the performance of four classification methods namely; Bartlett and Please method, The Bayesian Posterior Probability approach, the QDF approach and the AEDC method.
The following objectives were set to be achieved based on the anlaysed data.

1. To obtain discriminant functions for the equal mean case based on the Bartlett and Please approach, the Bayesian approach, the QDF and that of the AEDC approach.

2. To derive discriminant functions as well as their classification rules under unequal prior probabilities and under non-normality(contaminated data).

3. To compare the performance of all the four discriminant classification rules by using the various estimated error rates.

The data analysis comprised of both the preliminary analysis and detailed analysis.

The researcher used the preliminary anlaysis to test the statistical significance between the two mean vectors of Monozygotic and Dizygotic twin groups using Hotelling T^2. However from the test, we observed no significant difference between the mean vectors between the two groups. Hence the equal mean vectors assumption based on the hypothesis tested was not violated.

The detailed data analyses was based on three case situation; classification with equal prior probabilities, classification with unequal prior probabilities and classification with non-normality assumption.

1. The following results were obtained based on the case of classification with equal prior probabilities and equal misclassification cost.

    - The discriminant function derived using the Bartlett and Please classification method is shown below.

    $$Z_1 - 0.0606Z_2$$

    - Classification rule for the use of Bayesian posterior probability approach was derived as: Assign x_M to j if $P(j|x_M) > P(j'|x_M)$ and X_D to j' if $P(j'|x_D) > P(j|x_D)$.

    - The Quadratic discriminant Function was derived as

    $$\left\{(x-\mu_1)^{\tau}\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)^{\tau}\Sigma_2^{-1}(x-\mu_2)\right\} < -0.1664$$

    - The discriminant function as well as the classification rule for the AEDC method was derived as: Assign $|X|$ to π_1 if

$$y_1\left(\sqrt{\sigma_{11}^{(1)}}-\sqrt{\sigma_{11}^{(2)}}\right)+y_2\left(\sqrt{\sigma_{22}^{(1)}}-\sqrt{\sigma_{22}^{(2)}}\right)+y_3\left(\sqrt{\sigma_{33}^{(1)}}-\sqrt{\sigma_{33}^{(2)}}\right)$$
$$+\ldots$$
$$+y_{10}\left(\sqrt{\sigma_{1010}^{(1)}}-\sqrt{\sigma_{1010}^{(2)}}\right)\geq -1.2368$$

Otherwise to π_2.

- The classification rule obtained with the application of Bartlett and Please method was able to correctly classify approximately 87 percent of the twin observations.
- The proportion of correct classification using the Bayesian Posterior Probability approach was approximately 0.767.
- The QDF correctly classified 73 percent of the 30 twin observations sampled.
- The AEDC classification rule misclassified four twin observations from both groups representing approximately 0.867 proportion of correct classification.
- The performance of the four methods were evaluated using their mean error rate and the Bayesian classifier was observed to be performing better than the three other methods with a recorded mean error rate of 0.149.
- The mean error rate of the AEDC method of 0.166 was quiet observed to be closed to the BPM.
- The QDF performed poorly with a mean error rate of 0.283.
- The Bayesian Classifier provided better maximum separation between the two groups as compared to the AEDC, QDF and BPM.

2. The following observations were made under the case of classifying observations with unequal prior probabilities.

- The Bartlett and Please discriminant function/classification rules derived under the various sampling ratios are shown in the table below.

Table 5.1: Classification rules under the various sampling ratios

Sample Ratios	Classification Rules	Cut -off
$n_1 : n_2$		
1:2	$Z_1 - 0.0655Z_2 \leq 14.96$	14.96
1:3	$Z_1 - 0.0660Z_2 \leq 12.90$	12.90
1:4	$Z_1 - 0.0709Z_2 \leq 17.70$	17.70
$n_2 : n_1$		
1:2	$Z_1 - 0.0702Z_2 \leq 15.90$	15.90
1:3	$Z_1 - 0.0230Z_2 \leq 18.91$	18.91
1:4	$Z_1 - 0.0158Z_2 \leq 17.99$	17.99

- The Quadratic Discriminant functions derived under the various sampling ratios 1 : 2 ,1 : 3 and 1 : 4 were:

$$\left\{(x-\mu_1)^{\tau}\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)^{\tau}\Sigma_2^{-1}(x-\mu_2)\right\} < 1.658$$

$$\left\{(x-\mu_1)^{\tau}\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)^{\tau}\Sigma_2^{-1}(x-\mu_2)\right\} < 1.204$$

and

$$\left\{(x-\mu_1)^{\tau}\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)^{\tau}\Sigma_2^{-1}(x-\mu_2)\right\} < 1.585$$

- The classification rule for the Bayesian Posterior Probability approach was obtained under this case as follows: Assign x_M to j if $P(j|x_M) > P(j'|x_M)$ and X_D to j' if $P(j'|x_D) > P(j|x_D)$.
- The proportion of correct classification using BPM under the sample ratio $n_1 = 1 : n_2 = 2$ was 0.93.
- The proportion of correct classification using BPM under the sample ratio $n_1 = 1 : n_2 = 3$ was 1 or 100 percent.
- The proportion of correct classification using BPM under the sample ratio $n_1 = 1 : n_2 = 4$ was 0.80.

- The proportion of correct classification using the BPM under the sample ratio $n_2 = 1 : n_1 = 2$ was approximately 0.93.
- The proportion of correct classification using BPM under the sample ratio $n_2 = 1 : n_1 = 3$ was 0.75
- The proportion of correct classification using BPM under the sample ratio $n_2 = 1 : n_1 = 4$ was 0.80.
- The proportion of correct classification using Bayesian Posterior probability under the sample ratio $n_1 : n_2$ and $n_2 : n_1$ ranged from 0.65-0.95.
- The proportion of correct classification using QDF under the sample ratio $n_1 : n_2$ and $n_2 : n_1$ ranged from 0.52-0.84.
- Bartlett and Please method (BPM) outperformed the QDF and the Bayesian classifier approach. With the QDF performing poorly.

3. The following results were gathered when the data was contaminated to be non-normal.

   - Generally the performance of all the classification methods with equal and unequal prior probabilities deteriorated after the introduction of the outliers into the data.
   - The mean error estimate of AEDC method performed slightly better than the Bayesian Posterior Probability approach with mean error rate of 0.381.
   - The BPM was found to be very sensitive to outliers since it performed poorly under non-normality.
   - The QDF performed appreciably better under equal prior probability case and its performance is abysmal under unequal prior probability situation.

5.2 Recommendation

Based on the findings from our study, the following recommendations are made for equal mean discriminant analysis when the prior probabilities are equal, unequal and with contaminated data.

- It is recommended that Equal mean discrimination for more than two populations be researched into in order to come out with a classification rule that will discriminate effectively between more than two groups under unequal prior probabilities and equal misclassification cost situation.
- One should consider using the QDF when the working data is contaminated to be non-normal under equal prior probabilities.
- The Bayesian Posterior Probability approach as well as the AEDC classification methods should be employed in classifying equal mean vector observations under equal prior probability situations.
- One should consider using Bartlett and Please method in classifying observation with zero mean difference populations under unequal prior probabilities situation.
- One should also consider more sample selections in order to study the effect of the classification rule under unequal prior probabilities when the sample sizes are increased.
- For further research, one should consider the equal mean discrimination problem when the prior probabilities are equal and the misclassification cost are assumed to be unequal. Also other research studies should also go into establishing the actual distribution of a given data under non-normality and establish a classification rule when both the prior probabilities and the misclassification costs are unequal.

REFERENCES

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

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

Ganeslingam, S., Nanthakumar, A., and Ganesh, S. (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 Desu, M. M. (1968). Predictive zero-mean uniform discrimination. *Biometrika*, 55, No. 3:519–524.

Ḧardel, W. K. and Simar, L. (2007). *Applied Multivariate Statistical Analysis*. Springer-Verlag Berlin Heidelberg, second edition.

Hosseini, J. C. and Armacost, R. L. (1992). The two-group discriminant problem with equal group mean vectors: An experimental evaluation of six linear/nonlinear programming formulations. *European Journal of Operational Research*, 77:241–252.

Huberty, C. J. and Olejnik, S. (2006). *Applied MANOVA and Discriminant Analysis*. A John Wiley and Sons, Inc., Publication, second edition.

Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. NJ: Pearson Education, Inc.

Krzanowski, W. J. and Hand (1997). Assessing error rate estimators: The leave-one-out reconsidered. *Australian J. Statist*, 39(1):35–46.

Lachenbruch, P. A. (1975). Zero-mean difference discrimination and the absolute linear discriminant function. *Biometrika*, 62(2):397–401.

Lachenbruch, P. A., Sneeringer, C., and Revo, L. T. (1973). Robustness of the linear and quadratic discriminant functions to certain types of non-normality. *Computational Statistics*, 1:39–56.

Lachenbruch, P. A., W, C., B, B., and L, L. (1977). The effect of non-normality on quadratic discriminant function. *Technometrics*, 21:129–132.

Lavine, M. and West, M. (1992). A bayesian method for classification and discrimination. *The Canadian Journal of Statistics*, 20(4):451–461.

Lee, J. C. (1975). A note on equal-mean discrimination. *Communication Statististics-Theory Methods*, A4:251–254.

Marco, V. R., Young, D. M., and Tubne, D. W. (1987a). Asymptotic expansions and estimation of the expected error rate for equal-mean discrimination with uniform covariance structure. *Biometrika*, 29(1):103–111.

Marco, V. R., Young, D. M., and Turner, D. W. (1987b). The euclidean distance classifier: an alternative to the linear discriminant function. *Computational Statistics and Simulation*, 16:485–505.

McFarland, H. R. and Richards, D. S. P. (2001). Exact misclassification probabilities for plug-in normal quadratic discriminant functions:the equal-means case. *Journal of Multivariate Analysis*, 77:21–53.

McLachlan, G. J. (2004). *Discriminant Analysis and Statistical Pattern Recognition*. A John Wiley and Sons, Inc., Publication.

Noh, Y.-K., Byoung-Tak, Z., and Lee, D. D. (2010). Fluid dynamics models for low rank discriminant analysis. In *13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 9, Chia Laguna Resort, Sardinia, Italy.

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

Rausch, J. R. and Kelley, K. (2009). A comparison of linear and mixture models for discriminant analysis under nonnormality. *Behavior Research Methods*, 41(1):85–98.

Rencher, A. C. (2002). *Methods of Multivariate Analysis.* A John Wiley and Sons, Inc., Publication, second edition.

Rigby, R. A. (1997). Bayesian discrimination between two multivariate normal populations with equal covariance matrices. *Journal of the American Statistical Association*, 92(439):1151–1154.

Sajobi, T. T., Lix, L. M., Li, L., and Laverty, W. (2010). Discriminant analysis for repeated measures data: Effects of mean and covariance misspecification on bias and error in discriminant function coefficients.

Schwemer, G. T. and Mickey, M. R. (1980). A note on the linear discriminant function when group means are equal. *Comunication Statististics and Simulation*, B9:633–638.

Sharma, S. (1996). *Applied Multivariate Techniques.* John Wiley and Sons. Inc. New York.

Stocks, P. (1933). A biometric investigation of twins. part ii. *Annals of Eugenics*, 5:1–55.

Young, D. M., Turner, D. W., and Marco, V. R. (1988). On the robustness of the equal-mean discrimination rule with uniform covariance structure against serially correlated training data.

Appendix A

5.3 Table A1

Table 5.2: Likelihoods for equal group prior probability

$f_1(x_M\|j)$	$f_2(x_M\|j')$	$f_2(x_D\|j')$	$f_1(x_D\|j)$
4.08e-05	1.690e-05	5.620e-06	1.140e-05
1.510e-05	3.338e-06	7.892e-07	4.085e-07
3.604e-05	2.850e-05	3.012e-05	5.243e-06
1.312e-05	1.370e-05	1.881e-07	1.133e-08
1.060e-05	5.041e-06	1.019e-08	1.833e-08
3.555e-05	4.444e-05	4.011e-06	6.101e-06
3.866e-05	1.279e-05	1.993e-09	2.227e-10
3.467e-05	6.344e-06	4.797e-07	1.587e-07
1.425e-09	7.334e-11	1.137e-05	4.404e-06
3.181e-05	4.024e-06	1.585e-05	1.124e-05
1.701e-05	1.074e-05	4.903e-12	6.430e-15
1.962e-05	9.258e-06	2.156e-06	2.384e-07
4.063e-05	1.116e-05	4.370e-08	2.766e-10
3.924e-05	2.867e-05	3.689e-08	4.071e-08
4.310e-05	4.340e-05	2.502e-05	6.330e-06

5.4 Table A2: Likelihoods for, $n_1 = 1 : n_2 = 2$

Table 5.3: Likelihoods for unequal group prior probability for $n_1 = 1 : n_2 = 2$

$f_1(x_M\|j)$	$f_2(x_M\|j')$	$f_2(x_D\|j')$	$f_1(x_D\|j)$
1.29E-05	6.39E-06	4.77E-06	9.31E-06
6.42E-06	9.45E-07	5.35E-07	2.53E-07
1.52E-05	7.35E-06	2.84E-05	9.88E-06
5.49E-06	3.58E-06	2.07E-07	1.65E-08
6.26E-06	1.25E-06	1.76E-08	3.76E-08
		4.09E-06	2.67E-06
		9.47E-09	9.52E-11
		3.52E-07	5.92E-07
		2.73E-05	3.78E-06
		2.26E-05	1.44E-05

5.5 Table A3: Likelihoods for sample ratio, $n_1 = 1 : n_2 = 3$

Table 5.4: Likelihoods for the sample ratio $n_1 = 1 : n_2 = 3$

$f_1(x_M\|j)$	$f_2(x_M\|j')$	$f_2(x_D\|j')$	$f(x_D\|j)$
9.78E-06	7.19E-06	4.13E-11	4.19E-12
4.86E-06	4.82E-07	6.24E-06	2.35E-06
1.15E-05	4.13E-06	4.27E-06	2.01E-06
4.16E-06	2.51E-06	5.13E-09	7.18E-11
4.73E-06	1.26E-06	3.77E-06	1.02E-05
		4.11E-07	4.43E-07
		2.13E-06	2.07E-07
		7.62E-09	3.33E-09
		3.50E-06	2.84E-06
		5.37E-09	2.82E-08
		5.33E-07	1.58E-06
		2.26E-07	3.04E-08
		3.21E-06	1.18E-06
		4.39E-06	6.97E-06
		1.94E-05	4.66E-05

..

5.6 Table A4:Likelihoods for sample ratio, $n_1 = 1 : n_2 = 4$

Table 5.5: Likelihoods for the sample selection $n_1 = 1 : n_2 = 4$

$f(x_M\|j)$	$f(x_M\|j')$	$f(x_D\|j')$	$f(x_D\|j)$
7.82E-06	3.59E-06	5.03E-11	4.46E-12
3.89E-06	3.52E-07	5.17E-06	2.51E-06
9.20E-06	2.83E-06	3.91E-06	2.14E-06
3.33E-06	1.84E-06	5.17E-09	7.66E-11
3.79E-06	6.33E-07	3.69E-06	1.09E-05
		2.29E-07	4.73E-07
		1.54E-06	2.20E-07
		1.36E-08	3.55E-09
		3.13E-06	3.03E-06
		3.84E-09	3.01E-08
		4.43E-07	1.69E-06
		3.40E-07	3.25E-08
		3.35E-06	1.26E-06
		3.65E-06	7.44E-06
		1.82E-05	4.97E-05
		4.22E-06	1.12E-05
		6.18E-07	1.61E-06
		4.14E-11	6.27E-15
		3.34E-07	1.11E-08
		6.68E-07	2.33E-07

..

5.7 Table A5:Likelihoods for sample ratio, $n_2 = 1 : n_1 = 2$

Table 5.6: Likelihoods for sample ratio, $n_2 = 1 : n_1 = 2$

$f_1(x_M\|j)$	$f_2(x_M\|j')$	$f_2(x_D\|j')$	$f_1(x_D\|j)$
4.08E-05	1.69E-05	4.77E-06	9.31E-06
1.51E-05	3.34E-06	5.35E-07	2.53E-07
3.60E-05	2.85E-05	2.84E-05	9.88E-06
1.31E-05	1.37E-05	2.07E-07	1.65E-08
1.06E-05	5.04E-06	1.76E-08	3.76E-08
3.56E-05	4.44E-05		
3.87E-05	1.28E-05		
3.47E-05	6.34E-06		
1.43E-09	7.33E-11		
3.18E-05	4.02E-06		

5.8 Table A6:Likelihood for sample ratio, $n_2 = 1 : n_1 = 3$

Table 5.7: Likelihood for sample ratio, $n_2 = 1 : n_1 = 3$

$f(x_M\|j)$	$f(x_M\|j')$	$f(x_D\|j')$	$f(x_D\|j)$
3.04E-05	4.15E-06	2.31E-09	7.32E-12
1.13E-05	1.49E-07	2.64E-06	1.04E-06
2.69E-05	4.63E-06	2.05E-06	1.54E-06
9.79E-06	1.40E-06	1.60E-08	5.68E-11
7.79E-06	3.01E-07	8.04E-08	1.53E-06
2.66E-05	3.72E-06		
2.89E-05	2.03E-06		
2.59E-05	1.08E-06		
1.07E-09	4.87E-10		
2.38E-05	1.99E-06		
1.28E-05	1.15E-06		
1.45E-05	2.75E-06		
3.03E-05	3.92E-06		
2.93E-05	1.02E-06		
3.22E-05	2.56E-06		

5.9 Likelihood for the sample ratio, $n_2 = 1 : n_1 = 4$

Table 5.8: Likelihood for the sample ratio, $n_2 = 1 : n_1 = 4$

$f(x_M\|j)$	$f(x_M\|j')$	$f(x_D\|j')$	$f(x_D\|j)$
3.32E-05	4.43E-06	1.82E-09	1.18E-11
9.38E-06	1.58E-07	2.11E-06	9.03E-07
2.69E-05	4.93E-06	1.64E-06	1.36E-06
1.23E-05	1.49E-06	1.30E-08	9.26E-11
6.32E-06	3.21E-07	6.47E-08	7.99E-07
2.93E-05	3.96E-06		
2.96E-05	2.16E-06		
2.05E-05	1.14E-06		
6.60E-10	5.19E-10		
2.38E-05	2.12E-06		
1.19E-05	1.23E-06		
1.57E-05	2.95E-06		
3.19E-05	4.19E-06		
2.56E-05	1.08E-06		
3.45E-05	2.72E-06		
5.18E-15	1.02E-13		
1.09E-05	1.52E-07		
4.31E-05	2.36E-06		
7.38E-06	2.99E-06		
1.60E-05	1.20E-05		