Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.8, No.4, 2018 www.iiste.org

DISTRIBUTION OF ERRORS OF MISCLASSIFICATION FOR THE LINEAR DISCRIMINANT FUNCTION (A CASE OF EDGEWORTH SERIES NON NORMAL DISTRIBUTION)

[1]Awogbemi, C.A. and [2]Onyeagu, S.I.

[1, 2]Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria.

E-Mails: [1]awogbemiadeyeye@yahoo.com [2]Onyeagusidney@gmail.com,

Abstract

In this paper, the discrimination and classification problem associated with the persistent non normal distribution has been studied. Sampling from non normal distribution is assessed through the distribution of errors of misclassification in respect of Edgeworth Series Distribution (ESD) which is restricted to asymmetry. The effects of applying a normal classificatory rule (ND) when the distribution is ESD by empirical approach is examined by comparing the errors of misclassication for ESD with ND using small sample sizes at every level of skewness factor. The empirical results obtained show that the normal procedure is sturdy against departure from normality. This is evident from the total probabilities of misclassification that are not greatly affected by the skewness factor.

Keywords: Normal Distribution, Classificatory Rules, Apparent Probability of Misclassification, Skewness Factor and Optimum Probabilities of Classification.

(1.0) Introduction

Discriminant analysis is a widely employed multivariate technique with two closely related goals(Discrimination and Classification). Discrimination is focused on the description of group separation which elucidates the difference between two or more groups (Alvin, 2002; William and Mathew, 1984).

In classification, we are concerned with prediction or allocation of observations into groups. In this case, a sample of observations is given and the problem is to classify them into groups which shall be distinct as possible (Ogum, 2002). In essence, classification problem arises when a researcher makes a number of measurements on an individual and wishes to classify the individual into one of several groups on the basis of these measurements. The individual cannot be identified with a group directly without recourse to the measurements. Fisher (1936), illustrating this concept, classified iris flower from unknown group (specie) to any of the three known species (Iris setosa-red, Iris Versicolour – green, and Iris Virginica black) on the basis of their attributes (Sepal length in cm, Sepal width in cm, Petal length in cm and Petal width in cm).

Anderson (2003), described classification problem as the problem of statistical decision making and as such, a good classification procedure should result to few misclassifications.

The concepts of discrimination and classification have been carefully studied by Onyeagu (2003), Osuji and Onyeagu (2009), Olasunde and Soyinka (2013), Johnson and Wichern (2007), Dixon and Brereton (2009), De la Cruz (2008), Gupta and Nagari (2000) among others.

In constructing a classification procedure, there is a need to minimize on the average, the bad effects of misclassification (Ariyo and Adebanji (2010), Richard and Dean 1988).

Let $X^1 = \left(X_1, X_2, \ldots, X_r \right)$ denote the vector of measurements of an observation. To classify X^1 into π_1 or π_2, we consider an item as a point in a *r*-dimensional space and then partition the space Ω into two regions R_1 and R_2 which are mutually exclusive and exhaustive. If X^1 falls into R_1, we classify it as coming from π_1 and if it falls into R_2, we allocate it to population π_2.

In following the classification procedure, two kinds of errors in classification can be made when the sets of measured characteristics are not clearly distinct. If an item is actually from π_1, one may classify it as coming from π_2. Also, one may classify an item from π_2 as coming from π_1. In practice, such errors could prove costly and we need to know the relative undesirability of these errors in classification.

The performance criteria for any suggested classification rule are the probabilities of misclassification or error rates. The respective probabilities of misclassification are:

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

$$P\left(2|1\right) = P\left(x \in R_2 \,|\, \pi_1\right) = \int_{R_2 = \Omega - R_1} f_1(x)\,dx \tag{2}$$

The Total Probability of Misclassification (TPM) is expressed as

$$\begin{aligned} \text{TPM} &= P_1\, P\left(2|1\right) + P_2\left(1|2\right) \\ &= P_1 \int_{R_2} f_1(x)\,dx + P_2 \int_{R_1} f_2(x)\,dx \end{aligned} \tag{3}$$

The TPM is calculated easily when the population consists of multivariate normal densities with known parameters defined as:

$k = \ln\left[\frac{C(1|2)}{C(2|1)}\right]\left[\frac{P_2}{P_1}\right]$ is equal to $\Phi\left(-\frac{\Delta}{2}\right)$ exactly, where Φ denotes the cumulative density function of standard normal distribution and $\Delta^2 \equiv (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2)$ is the Mahalanobis squared distance between $\mu(\mu_1, \Sigma)$ and $\mu(\mu_1, \Sigma)$.

The problem of classifying an observation into one of two multivariate normal populations with a common covariance matrix implies the classical classification. Fisher's Linear Discriminant Function serves as a criterion when samples are used to estimate the parameters of the two distributions. However, when the population parameters are unknown, and there is a need for their estimation from the sample, the exact calculation of the TPM for Linear Discriminant Function (LDF) is quite complex or complicated due to the intractability of the nature of the forms of the probabilities.

Various techniques have since been developed and compared in trying to find the best approach for estimating the unknown parameters from the samples (Onyeagu and Adeboye, 1996; Fujikoshi, 2000; Gupta and Nagari, 2000; Batsidis, et al., 2006).

Linear Discriminant Function (LDF) is employed as an assignment rule when:

(a) The density of observation from $\pi_i, (i = 1, 2)$ are multivariate normal;
$\pi_i \sim N_\rho(\mu_i, \Sigma), \quad (i = 1, 2)$
(b) The variance – covariance matrix in π_1 is the same as π_2;
(c) The priori probabilities $p_i\,(i = 1, 2)$ of an observation coming from $\pi_i, (i = 1, 2)$ respectively are known;
(d) The parameters of the density functions in (a) are known.

Suppose the assumptions specified above are satisfied, then the Linear Discriminant Function (LDF) provides optimal assignment rule in that it cannot be improved upon and the errors of misclassification are minimized. However, when some or all the assumptions are violated it would be of interest to determine the effects of the violation on the procedures using Linear Discriminant Functions (LDF).

The purpose of the study is to discriminate a non-normal distribution using the persistent non transformable Edgeworth series distribution. We investigate the effects of applying a normal classificatory rule when it is Edge-worth Series Distribution by empirical methods. This is assessed through the distribution of the errors of misclassification.

(2.0) Methodology

(2.1) Discrimination of Non-normal Distributions by Edgeworth Series Distribution (ESD)

Suppose X_{ij}, $i=1,2,\ j=1,2,\ldots,n_i$ denote two independent random samples from populations $\pi_i,\ i=1,2$ respectively. Then

$$f(x)=\left(1-\frac{\lambda_3}{6}D^3\right)\ \phi\left(\frac{x-\mu_i}{\sigma}\right),\ -\infty<x<\infty, i=1,2 \tag{4}$$

$\lambda_3, \mu_i (i=1,2)$ and σ satisfy the conditions $-\infty<\lambda_3<\infty,\ -\infty<\mu_i<\infty$ and $\sigma>0$, where D denotes the operator $\frac{\partial}{\partial x}$ and $\phi\left(\frac{x-\mu_i}{\sigma}\right)$ is the density function

$$\frac{(2\pi)^{-\frac{1}{2}}}{\sigma}\exp\left[\frac{-(x-\mu_i)^2}{2\sigma^2}\right] \tag{5}$$

and λ_3 is the skewness factor.

(2.1.1) Optimum Probability of Misclassification of ESD

When all the parameters of the distributions in the populations are known, the probability of misclassification is optimal in the sense that we cannot improve upon it.

When an observation from π_1 is misclassified, the optimum probability of misclassification is given by

$$\begin{aligned}
\alpha_1[R,f] &= \Pr\left[x\geq\frac{(\mu_1+\mu_2)}{2}\right]\\
&=\int_\sigma^\infty\left[1-\frac{\lambda_3}{6}D^3\right]\phi\left(\frac{x-\mu_1}{\sigma}\right)dx\\
&=\int_\sigma^\infty\left[1+\frac{\lambda_3}{6\sigma^3}H_3\left(\frac{x-\mu_1}{\sigma}\right)\right]\phi\left(\frac{x-\mu_1}{\sigma}\right)dx\\
&=\int_\sigma^\infty\phi\left(\frac{x-\mu_1}{\sigma}\right)dx+\frac{\lambda_3}{6\sigma^3}\int_\sigma^\infty H_3\left(\frac{x-\mu_1}{\sigma}\right)\phi\left(\frac{x-\mu_1}{\sigma}\right)dx
\end{aligned} \tag{6}$$

where $\sigma=\left(\frac{\mu_1+\mu_2}{2}\right)$ and $H_r(x)$ is Chebyshev's - Hermite polynomial of degree r and defined by the identity:

$$H_r(x)\phi(x)=(-D)^r\phi(x) \tag{7}$$

Using the results of Kendall and Stuart (1958),

(8)

$$\int_{-\infty}^{t} H_r(x)\phi(x)dx = -H_{r-1}(t)\phi(t)$$

and setting $z = \frac{x-\mu_1}{\sigma}$ as in equation, we have

$$\alpha_1[R, f] = \int_{\frac{\sigma-\mu_1}{\sigma}}^{\infty} \phi(z)dz + \frac{\lambda_3}{6\sigma^2}\int_{\frac{\sigma-\mu_1}{\sigma}}^{\infty} H_3(z)\phi(z)dz$$

$$= 1-\phi\left(\frac{\sigma-\mu_1}{\sigma}\right) + \frac{\lambda_3}{6\sigma^2}H_2\left(\frac{\sigma-\mu_1}{\sigma}\right)\phi\left(\frac{\sigma-\mu_1}{\sigma}\right)$$

$$= 1-\phi\left(\frac{\mu_2-\mu_1}{2\sigma}\right) + \frac{\lambda_3}{6\sigma^2}\left[\left(\frac{\mu_2-\mu_1}{2\sigma}\right)-1\right]\phi\left(\frac{\mu_2-\mu_1}{2\sigma}\right)$$

If an observation from π_2 is misclassified, the optimum probability of misclassification is given by

$$\alpha_2[R, f] = \Pr\left[x < \frac{(\mu_1+\mu_2)}{2}\right]$$

$$= \int_{-\infty}^{\sigma}\left[1-\frac{\lambda_3}{6}D^3\right]\phi\left(\frac{x-\mu_2}{\sigma}\right)dx$$

$$= \int_{-\infty}^{\sigma}\phi\left(\frac{x-\mu_2}{\sigma}\right) + \frac{\lambda_3}{6\sigma^3}\int_{-\infty}^{\sigma}H_3\left(\frac{x-\mu_2}{\sigma}\right)\phi\left(\frac{x-\mu_2}{\sigma}\right)dx \tag{9}$$

Setting $\sigma = \frac{(\mu_1+\mu_2)}{2}$ *and* $z = \frac{x-\mu_1}{\sigma}$, we have

$$\alpha_2[R, f] = \Pr\int_{-\infty}^{\frac{(\sigma-\mu_2)}{\sigma}}\phi(z)dz + \frac{\lambda_3}{6\sigma^2}\int_{-\infty}^{\frac{(\sigma-\mu_2)}{\sigma}}H_3(z)\phi(z)dx$$

$$= \phi\left[\frac{\sigma-\mu_2}{\sigma}\right] - \frac{\lambda_3}{6\sigma^2}H_2\left[\frac{\sigma-\mu_2}{\sigma}\right]\phi\left[\frac{\sigma-\mu_2}{\sigma}\right]$$

$$= \phi\left[\frac{\mu_1-\mu_2}{2\sigma}\right] - \frac{\lambda_3}{6\sigma^2}\left[\left(\frac{\mu_1-\mu_2}{2\sigma}\right)^2-1\right]\phi\left(\frac{\mu_1-\mu_2}{2\sigma}\right) \tag{10}$$

The optimum probability of misclassification is of interest in this study as it would be used subsequently for comparison purposes.

(2.1.2) Classificatory Rules for Estimating Errors of Misclassification

For the ESD with $\mu_1 < \mu_2$, the classificatory rule is

$$X \in \pi_2 \; if \; \frac{P \exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma}\right)^2\right]}{Q \exp\left[-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma}\right)^2\right]} < 1 \tag{11}$$

and

$$classify \; X \in \pi_1 \; if \; \frac{P \exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma}\right)^2\right]}{Q \exp\left[-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma}\right)^2\right]} \geq 1 \tag{12}$$

$$where \; P = \left[1 - \left\{\frac{\lambda_3}{2\sigma^3}\right\}\left(\frac{x-\mu_1}{\sigma}\right) + \left\{\frac{\lambda_3}{6\sigma^3}\right\}\left(\frac{x-\mu_1}{\sigma}\right)^3 \phi\left(\frac{x-\mu_1}{\sigma}\right)\right]$$

and

$$Q = \left[1 - \left\{\frac{\lambda_3}{2\sigma^3}\right\}\left(\frac{x-\mu_2}{\sigma}\right) + \left\{\frac{\lambda_3}{6\sigma^3}\right\}\left(\frac{x-\mu_2}{\sigma}\right)^3 \phi\left(\frac{x-\mu_2}{\sigma}\right)\right]$$

The normal classificatory rule for $\mu_1 < \mu_2$ is

$$Classify \; X \in \pi_1 \; if \; X < \left(\frac{\mu_1 + \mu_2}{2}\right) \tag{13}$$

and

$$classify \; X \in \pi_2 \; if \; otherwise$$

(2.1.3) Estimating the Probabilities of Misclassification

$Let \; X_{ij}, i = 1,2; \; j = 1,2,\ldots n_i$ *be independent samples of sizes* n_1, n_2 *from populations* π_1, π_2 To estimate the apparent probabilities of misclassification, we define

$$E_{12E} = \sum_{j=1}^{n_1} \frac{\gamma_j}{n_1} \quad and \; E_{21E} = \sum_{j=1}^{n_1} \frac{\delta_j}{n_1} \tag{14}$$

where $\gamma_j = 1$ *if* X_{1j} *is classified as belonging to* π_2 *and* $\gamma_j = 0,$ *if* X_{1j} *is classified as belonging to* π_1
$\delta_j = 1$ *if* X_{2j} *is classified as belonging to* π_1 *and* $\delta_j = 0,$ *if* X_{2j} *is classified as belonging to* π_2
E_{12E} and E_{21E} represent the apparent probabilities of misclassification when observations from π_1 *and* π_2 are misclassified respectively by Edge worth Series Distribution rule.

$$E_{12N} = \sum_{j=1}^{n_1} \frac{\gamma_j}{n_1} \ and \ E_{21N} = \sum_{j=1}^{n_2} \frac{\sigma_j}{n_2} \tag{15}$$

E_{12N} and E_{21N} represent the apparent probabilities of misclassification when observations from populations 1 and 2 are misclassified respectively by Normal Distribution rule.

(3.0) Simulation Experiments and Results

The optimum probabilities of misclassification for the Edgeworh Series Distribution (ESD) are computed with $\mu_1 = 0, \mu_2 = 1 \, and \, \sigma = 1$ with the skewness factor λ_3 in the range (6.25 x 10^{-3}, 0.4) .See (Barton and Dennis ,1952).

The apparent probabilities of misclassification for the (ESD) and ND are also examined when $\mu_1, and \ \mu_2$ are known and when the parameters are estimated from the samples. Two independent samples of simulation size of 200 each are configured at each value of the skewness factor (λ_3) from $\pi_1 \ and \ \pi_2$ whose distributions are of ESD with the respective parameters:
$\mu_1 = 0, \sigma^2 = 1 \, and \ \mu_2 = 1, \sigma^2 = 1$.

Using the ESD and Normal Distribution (ND) classification rules, the proportion misclassified in $\pi_1 \ and \ \pi_2$ are obtained and repeated for small samples (n = 5, 10, 15, 20, 25). The random numbers are generated using R program and simulation results are obtained and displayed in Tables 3.1- 3.7

Table 3.1: Optimum Probabilities of Misclassification at Different Values of Skewness for ESD (all Parameters Known)

	Optimum Probability of Misclassification		
Skewness Factor(λ_3)	**E_{12E}**	**E_{21E}**	**Total**
0.00625	0.3082	0.3088	0.6170
0.0125	0.3079	0.3091	0.6170
0.025	0.3074	0.3096	0.6170
0.05		0.3107	0.6170
0.10	0.3041	0.3129	0.6170
0.15	0.3019	0.3151	0.6170
0.20	0.2997	0.3173	0.6170
0.25	0.2975	0.3195	0.6170
0.30	0.2953	0.3217	0.6170
0.35	0.2931	0.3239	0.6170
0.40	0.2909	0.3261	0.6170

Table 3.2: Comparison of Errors of Misclassification for All known Parameters of ESD with ND Averaged over 5 Samples

Skewness Factor(λ_3)	E_{12E}	E_{21E}	Total	E_{12N}	E_{21N}	Total
0.00625	0.329	0.303	0.632	0.327	0.303	0.630
0.0125	0.321	0.321	0.642	0.318	0.322	0.640
0.025	0.282	0.326	0.608	0.278	0.330	0.608
0.05	0.312	0.289	0.601	0.308	0.296	0.604
0.10	0.328	0.296	0.624	0.315	0.307	0.622
0.15	0.355	0.298	0.653	0.330	0.324	0.654
0.20	0.328	0.305	0.633	0.287	0.331	0.618
0.25	0.339	0.262	0.601	0.298	0.296	0.594
0.30	0.377	0.248	0.625	0.320	0.304	0.624
0.35	0.387	0.255	0.642	0.328	0.310	0.638
0.40	0.396	0.247	0.643	0.297	0.301	0.598

Table 3.3: Comparison of Errors of Misclassification of ESD with ND for Means unknown and Estimated by Averaged Values over 5 Samples

Skewness Factor(λ_3)	E_{12E}	E_{21E}	Total	E_{12N}	E_{21N}	Total
0.00625	0.140	0.40	0.540	0.14	0.400	0.540
0.0125	0.220	0.41	0.630	0.22	0.410	0.630
0.025	0.225	0.46	0.690	0.22	0.475	0.695
0.05	0.210	0.39	0.605	0.20	0.400	0.605
0.10	0.205	0.47	0.680	0.17	0.495	0.670
0.15	0.260	0.28	0.545	0.23	0.320	0.550
0.20	0.305	0.36	0.670	0.29	0.395	0.690
0.25	0.455	0.18	0.640	0.42	0.230	0.650
0.30	0.195	0.46	0.660	0.11	0.545	0.660
0.35	0.225	0.46	0.660	0.12	0.520	0.645
0.40	0.440	0.18	0.610	0.36	0.250	0.610

Table 3.4: Comparison of Errors of Misclassification of ESD with ND for Means unknown and Estimated by Averaged Values over 10 Samples

Skewness Factor(λ_3)	E_{12E}	E_{21E}	Total	E_{12N}	E_{21N}	Total
0.00625	0.252	0.249	0.501	0.252	0.315	0.567
0.0125	0.236	0.236	0.472	0.236	0.236	0.472
0.025	0.266	0.219	0.485	0.231	0.295	0.526
0.05	0.224	0.282	0.506	0.216	0.314	0.530
0.10	0.290	0.278	0.568	0.208	0.336	0.544
0.15	0.387	0.203	0.590	0.215	0.220	0.435
0.20	0.277	0.320	0.597	0.270	0.337	0.607
0.25	0.255	0.245	0.500	0.230	0.292	0.522
0.30	0.248	0.334	0.582	0.182	0.394	0.576
0.35	0.216	0.339	0.555	0.175	0.354	0.529
0.40	0.253	0.209	0.462	0.170	0.196	0.366

Table 3.5: Comparison of Errors of Misclassification of ESD with ND for Means unknown and Estimated by Averaged Values over 15 Samples

Skewness Factor(λ_3)	E_{12E}	E_{21E}	Total	E_{12N}	E_{21N}	Total
0.00625	0.345	0.145	0.490	0.345	0.150	0.495
0.0125	0.310	0.310	0.620	0.310	0.310	0.620
0.025	0.405	0.280	0.685	0.400	0.285	0.685
0.05	0.230	0.390	0.620	0.225	0.395	0.620
0.10	0.375	0.305	0.680	0.350	0.315	0.665
0.15	0.405	0.180	0.585	0.360	0.225	0.585
0.20	0.355	0.325	0.680	0.320	0.355	0.675
0.25	0.295	0.340	0.635	0.235	0.395	0.630
0.30	0.320	0.350	0.670	0.230	0.385	0.615
0.35	0.260	0.345	0.605	0.200	0.430	0.630
0.40	0.315	0.375	0.690	0.145	0.415	0.560

Table 3.6: Comparison of Errors of Misclassification of ESD with ND for Means unknown and Estimated by Averaged Values over 20 Samples

Skewness Factor (λ_3)	E_{12E}	E_{21E}	Total	E_{12N}	E_{21N}	Total
0.00625	0.220	0.206	0.426	0.220	0.206	0.426
0.0125	0.280	0.280	0.560	0.192	0.295	0.487
0.025	0.330	0.210	0.540	0.290	0.230	0.520
0.05	0.345	0.205	0.550	0.295	0.250	0.545
0.10	0.265	0.300	0.565	0.230	0.390	0.620
0.15	0.340	0.350	0.690	0.330	0.375	0.705
0.20	0.350	0.240	0.590	0.320	0.255	0.575
0.25	0.295	0.270	0.565	0.270	0.295	0.565
0.30	0.300	0.195	0.495	0.265	0.200	0.465
0.35	0.310	0.350	0.660	0.270	0.360	0.630
0.40	0.405	0.285	0.690	0.380	0.400	0.780

Table 3.7: Comparison of Errors of Misclassification of ESD with ND for Means unknown and Estimated by Averaged Values over 25 Samples

Skewness Factor(λ3)	E_{12E}	E_{21E}	Total	E_{12N}	E_{21N}	Total
0.00625	0.27	0.22	0.490	0.270	0.22	0.490
0.0125	0.29	0.33	0.620	0.290	0.23	0.525
0.025	0.39	0.29	0.685	0.375	0.31	0.685
0.05	0.34	0.27	0.610	0.335	0.28	0.615
0.10	0.37	0.30	0.680	0.360	0.31	0.675
0.15	0.36	0.23	0.590	0.345	0.24	0.590
0.20	0.27	0.43	0.705	0.225	0.48	0.705
0.25	0.37	0.25	0.630	0.320	0.29	0.610
0.30	0.39	0.24	0.630	0.300	0.33	0.630
0.35	0.29	0.30	0.590	0.240	0.34	0.585
0.40	0.40	0.22	0.630	0.305	0.29	0.595

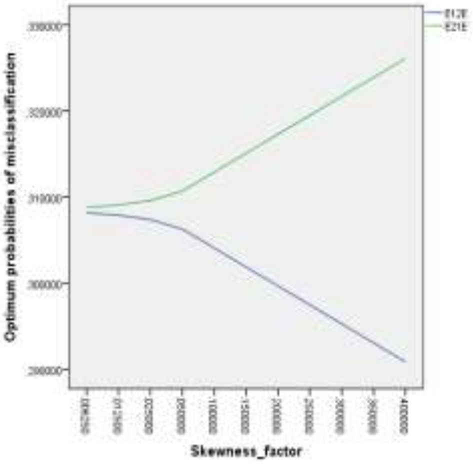

fig.3.1: graph showing optimum probabilities of misclassification for ESD.

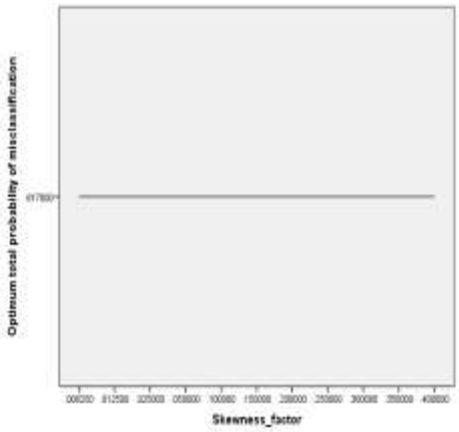

Fig. 3.2: Graph showing Optimum Total Probabilities of Misclassification for ESD .

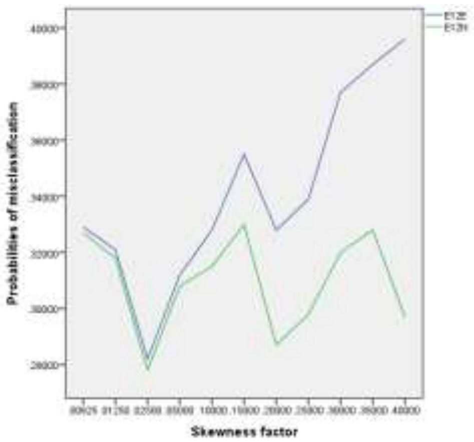

Fig.3.3: Graph showing probabilities of misclassification for all known parameters averaged over 5 samples (E_{12E} and E_{12N})

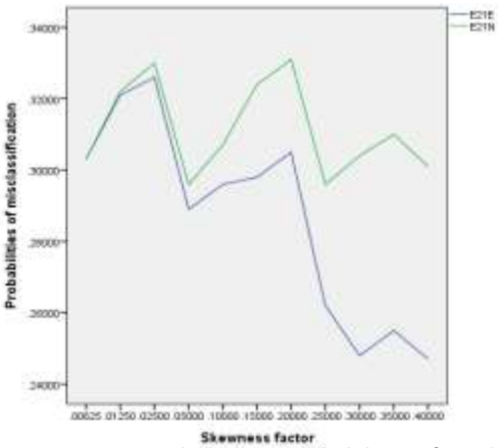

Fig.3.4: Graph showing probabilities of misclassification for all known parameters averaged over 5 samples (E_{21E} and E_{21N})

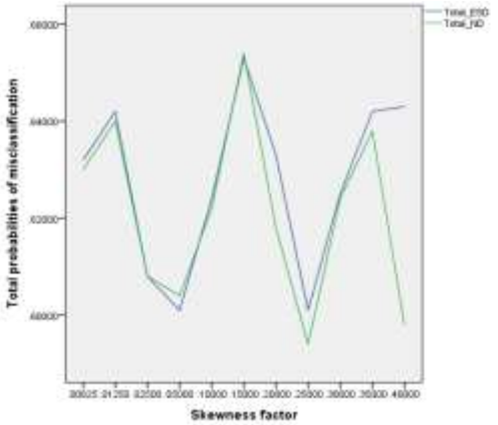

Fig.3.5: Graph showing total probabilities of misclassification(ESD and ND) for all known parameters averaged over 5 samples

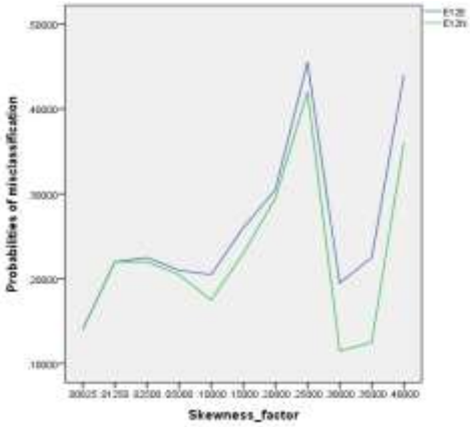

Fig.3.6: Graph showing probabilities of misclassification for unknown parameters averaged over 5 samples (E_{12E} and E_{12N})

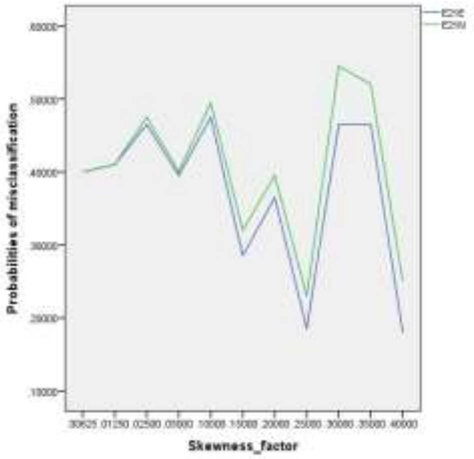

Fig.3.7: Graph showing probabilities of misclassification for unknown parameters averaged over 5 samples (E_{21E} and E_{21N})

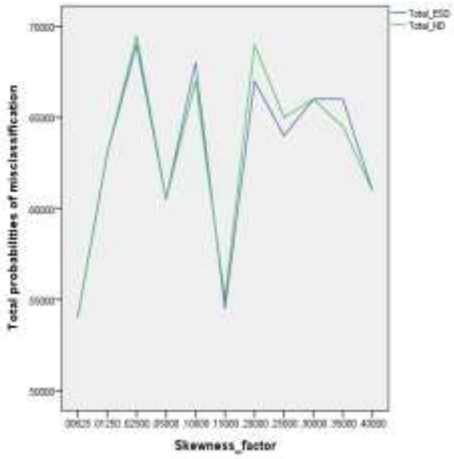

Fig.3.8: Graph showing total probabilities of misclassification(ESD and ND) for unknown parameters averaged over 5 samples

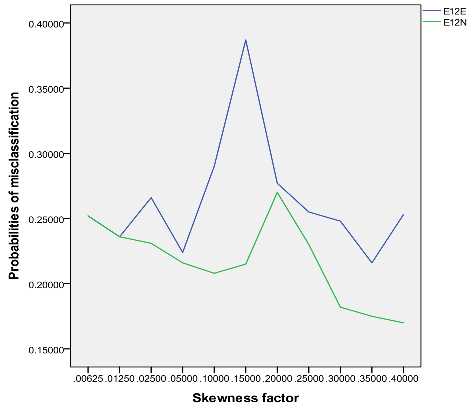

Fig.3.9: Graph showing probabilities of misclassification (E_{12E} and E_{12N}) for unknown parameters averaged over 10 samples

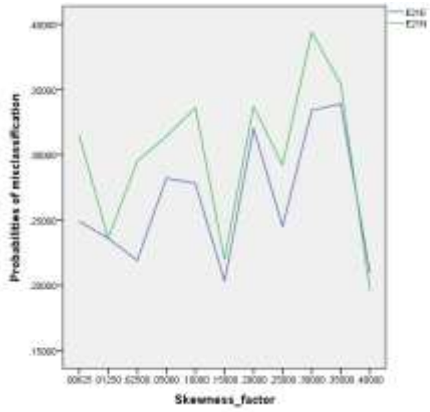

Fig.3.10: Graph showing probabilities of misclassification (E_{21E} and E_{21N}) for unknown parameters averaged over 10 samples

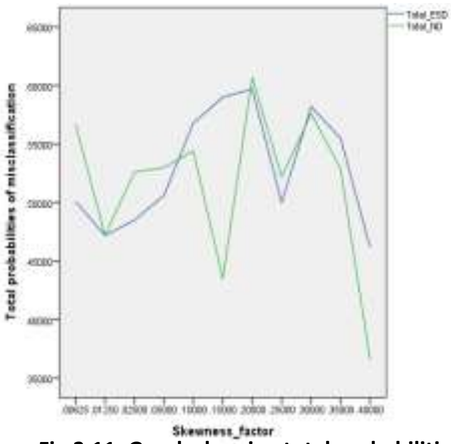

Fig.3.11: Graph showing total probabilities of misclassification (ESD and ND) for unknown parameter averaged over 10 samples

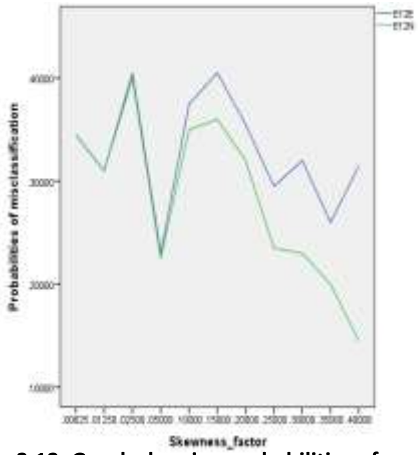

Fig.3.12: Graph showing probabilities of misclassification (E_{12E} and E_{12N}) for unknown parameters averaged over 15 samples

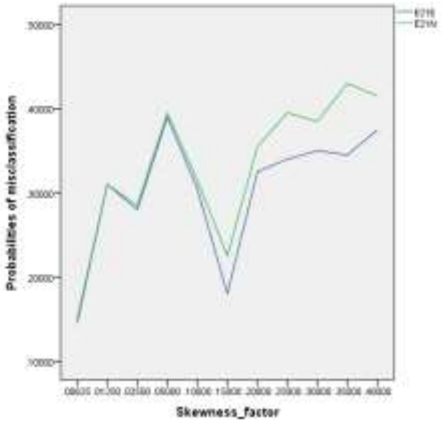

Fig.3.13: Graph showing probabilities of misclassification (E_{21E} and E_{21N}) for unknown parameters averaged over 15 samples

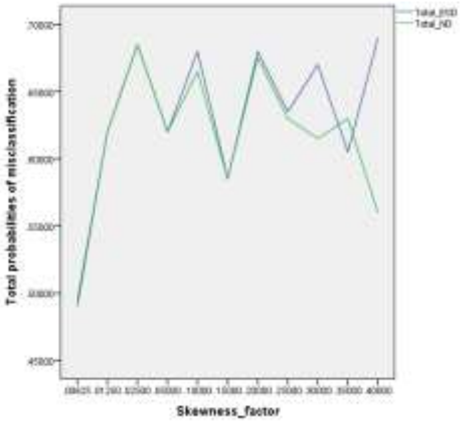

Fig.3.14: Graph showing total probabilities of misclassification (ESD and ND) for unknown parameter averaged over 15 samples

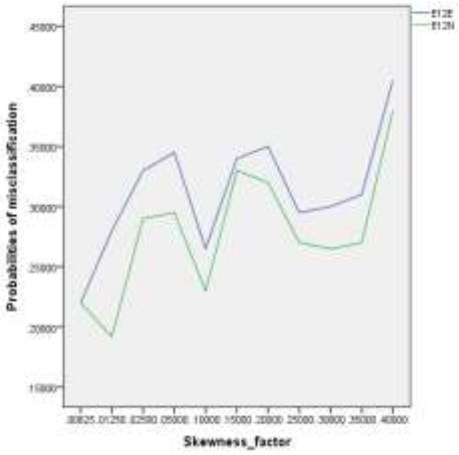

Fig.3.15: Graph showing probabilities of misclassification (E_{12E} and E_{12N}) for unknown parameters averaged over 20 samples

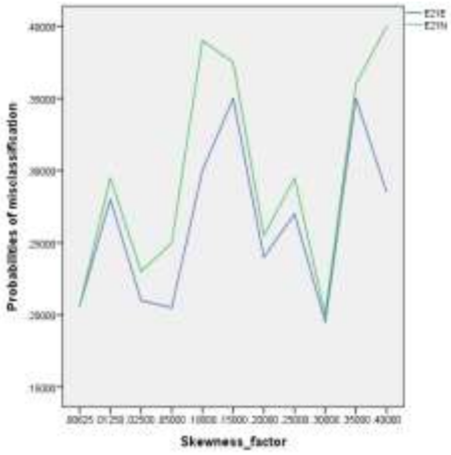

Fig.3.16: Graph showing probabilities of misclassification (E_{21E} and E_{21N}) for unknown parameters averaged over 20 samples

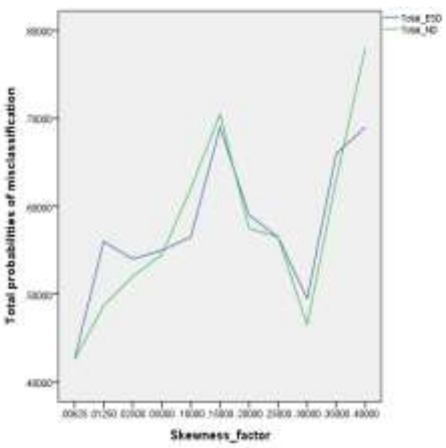

Fig.3.17: Graph showing total probabilities of misclassification (ESD and ND) for unknown parameter averaged over 20 samples

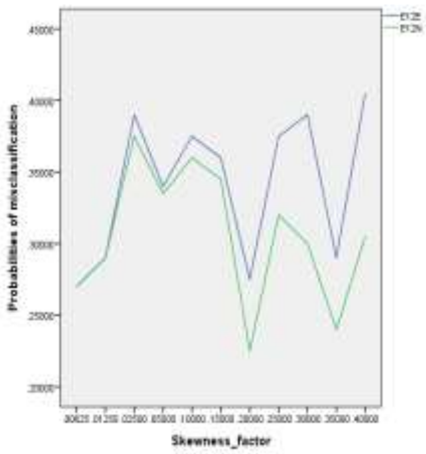

Fig.3.18: Graph showing probabilities of misclassification (E_{12E} and E_{12N}) for unknown parameters averaged over 25 samples

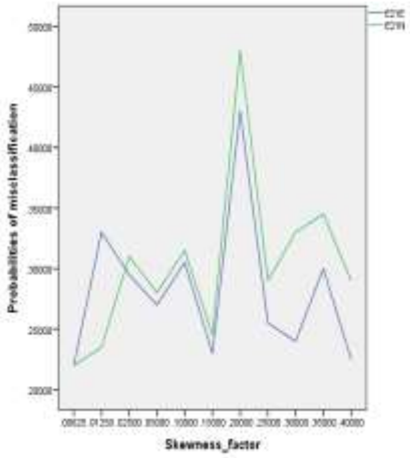

Fig.3.19: Graph showing probabilities of misclassification (E_{21E} and E_{21N}) for unknown parameters averaged over 25 samples

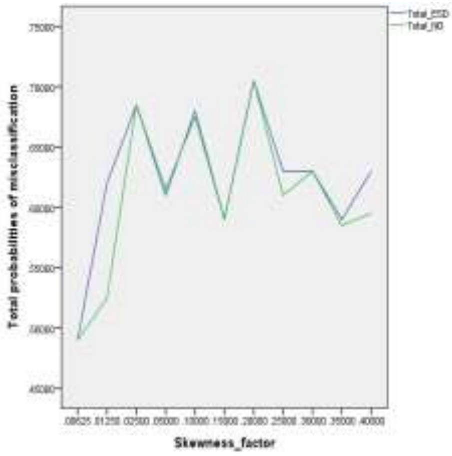

Fig.3.20: Graph showing total probabilities of misclassification (ESD and ND) for unknown parameter averaged over 20 samples

The results in Table 3.1 shows that E_{12} decreases as λ_3 increases and E_{21} increases as λ_3 increases. The Total Probability of Misclassification is also stable (constant) as λ_3 increases.

From Table 3.2 , E_{12E} is always higher than E_{12N} at every level of λ_3 and E_{21E} is also higher than E_{21N} at every level of $\lambda_{3.}$

The total probabilites of misclassification for the ESD and ND classification rules indicate no major difference between them at each value of $\lambda_{3.}$ It is also evident from Tables 3.2 – 3.7 that the total probability of misclassification at every value of λ_3 is either under or overestimated when small samples are employed to estimate $\mu_1, and\ \mu_2$.

The skewness factor (λ_3) has a very little effect on the total probability of misclassification, which implies that it is not affected by the departures from normality.

For the individual probabilities of misclassification: E_{12E} and E_{12N} at every level of λ_3 , their behaviours show that for small sample sizes $E_{12E} \geq E_{12N}$ and $E_{21E} \leq E_{21N}$. The observed equality occurs when λ_3 is very small with an increasing parity as λ_3 increases.

(3.0) Conclusion and Recommendation

The results obtained in Tables (3.2 – 3.7) asserts that the normal procedure is strong against departures from normality as shown by the asymmetry factor of ESD. The apparent probabilities of misclassification E_{12E} , E_{21E} and their totals are close to the corresponding errors prompted by the normal classification rule when λ_3 is small.

Estimation of the errors when small sample sizes are used to estimate the means is an indication that the optimum probability of misclassification is underestimated or overestimated. This is anchored on the data generated and strictly limited to this work.

It is recommended that further work should be done on the effects of using small sample sizes on the parameters. Efforts should also be made to derive the algebraic justification for equality of probabilities when λ_3 is very small.

(4.0) References

Alvin, C.R. (2002). *Methods of Multivariate Analysis*, 2nd ed. John Wiley & Sons Publications.

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Methods.* New York: John Wiley.

Barton, D.E. and Denis, K.E. (1952). *The Condition under which Gram-Charlier and Edgeworth Curves are Positive Define and Unimodal.*Biometrika, (39), pp. 425- 427.

Batsidis, A., Zografos, K. and Loukas, S. (2006). *Errors in Discrimination with Monotone Missing Data from Multivariate Normal Distribution.* Computational Statistics and Data Analysis, Vol. 50, No. 10, pp. 2600–2634.

De la Cruz (2008). *Bayesian Non-Linear Regression Models with Skew-Elliptical Errors: Applications to the Classification of Longitudinal Profiles.* Computational Statistics and Data Analysis (53), pp. 436–449.

Dixon, S.J. and Brereton (2009). *Comparison of Performance of Five Common Classifiers Represented as Boundary Methods:* Euclidean Distance to Centroids, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Learning Vector Quantization and Support Vector Machines as Dependent on Data Structure, Chemometrics, Intelligent Laboratory System, 95, pp. 1–17.

Gupta, A.K. and Nageri, D.K. (2000). *Matrix Variate Distribution.* Chapman and Hall, London.

Fisher, F.A. (1936). *The Use of Multiple Measurements in Taxonomic Problems.* Annals of Eugenics, 7, pp. 179–188.

Fujikoshi, Y. (2000). *Error Bound for Asymptotic Approximations of the Linear Discriminant Function when the Sample Size and the Dimension are Large.* Journal of Multivariate Analysis, 73, pp. 1-17.

Johnson, R.A. and Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis,* Pearson Education Inc. Newy Jersey, NJ. pp. 575–585.

Kendall, M.G. and Stuart, A. (1958). *The Advanced Theory of Statistics,* Vol. 3, Hafner: New York.

Ogum, G.E.O. (2002). *Introduction to Methods of Multivariate Analysis*, Afri-Towers Ltd., Aba, Nigeria.

Olosunde, A. and Soyinka, A. (2013). *Discrimination and Classification of Poultry Feed Data*. International Journal of Mathematical Research, 2, pp. 37–44.

Onyeagu, S.I. (2003). *Derivation of an Optimal Classification Rule for Discrete Variables.* Journal of Nigerian Statistical Association. Vol. 4, pp. 79–80.

Onyeagu, S.I. and Adegboye, O.S. (1996). *Some Methods of Estimating Probability of Misclassification in Discriminant analysis.*ABACUS, MAN, Vol. 24, No. 2, pp. 104–112.

Osuji, G.A. and Onyeagu, S.I. (2009). *A Review of Classification Models Arising from full Mathematical Rule.* Journal of the Nigerian Statistical Association, 21, pp. 84–89.

Richard, A.J. and Dean, W.W. (1998). *Applied Multivariate Statistical analysis.*4th Ed. Prentice Hall, Inc. New Jersey.

William, R.D. and Matthew, G. (1984). *Multivariate Analysis, Methods and Applications.*John Wiley & Sons, Inc.