American Journal of Theoretical and Applied Statistics
2015; 4(2-1): 44-51
Published online March 21, 2015 (http://www.sciencepublishinggroup.com/j/ajtas)
doi: 10.11648/j.ajtas.s.2015040201.16
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

Application of Discriminant Analysis on Broncho-Pulmonary Dysplasia among Infants: A Case Study of UMTH and UDUS Hospitals in Maidguri, Nigeria

Vishwa Nath Maurya[1], Madaki Umar Yusuf[2], Vijay Vir Singh[2], Babagana Modu[2]

[1]Department of Mathematics and Statistics, School of Science & Technology, The University of Fiji, Lautoka, Fiji Islands
[2]Department of Mathematics and Statistics, Yobe State University, P.M.B 1144, Damaturu, Yobe State, Nigeria

Email address:

prof.drvnmaurya@gmail.com (V. N. Maurya), prof_vnmaurya@yahoo.in (V. N. Maurya), bomayu84@gmail.com (M. U. Yusuf), singhvijayvir68@gmail.com (V. V. Singh), modubabagana70@gmail.com (B. Modu)

To cite this article:

Vishwa Nath Maurya, Madaki Umar Yusuf, Vijay Vir Singh, Babagana Modu. Application of Discriminant Analysis on Broncho-Pulmonary Dysplasia among Infants: A Case Study of UMTH and UDUS Hospitals in Maidguri, Nigeria. *American Journal of Theoretical and Applied Statistics*. Special Issue: Scope of Statistical Modeling and Optimization Techniques in Management Decision Making Process. Vol. 4, No. 2-1, 2015, pp. 44-51. doi: 10.11648/j.ajtas.s.2015040201.16

Abstract: Present research paper envisages investigation of the incidence and prevalence of Broncho-Pulmonary Dysplasia among infants in UMTH and UDUS Hospitals in Maidguri, Nigeria. The data used in this research paper were obtained from the University of Maiduguri Teaching Hospital (UMTH), Maiduguri and Usmanu Danfodiyo University Teaching Hospital Sokoto with sample size of Seventy (70) patients in 2014; Fifty (50) patients from Maiduguri and Twenty (20) from Sokoto respectively. Discriminant analysis model was employed for the analysis with the help of SPSS. The result of the analysis indicates that discriminant model has a perfect classification of new cases in Maiduguri while it has misclassified one of five new cases in Sokoto. This result suggests that the prediction of Broncho–Pulmonary Dysplasia (BPD) is better done with discriminant model in Maiduguri. The study recommends that Doctors and Clinics should adopt the use of the models built by this research to detect the prevalence of BPD among infants.

Keywords: Linear Discriminant Analysis Model, Broncho-Pulmonary Dysplasia (BPD), Infants, Logistic Regression, Dichotomous Factor, Wilks' Lambda test, Omnibus Chi-Square Test

1. Introduction

It is very important to start a research of this kind with an explanation of the purpose of statistical science. As proposed by Usman [17] that Statistics is basically defined as the collection, organization, analysis, presentation and interpretation of data with the aim of drawing a logical conclusion. Multivariate methods are relevant in virtually every branch of applied medicine, pharmacy and public health, for further details we refer Maurya et al. [8, 9,10,11]. Very recently, Maurya et al [12] employed some parametric tests and succeeded to explore and analyze on the rate of kidney (renal) failure. The statistical methods come into play either when we have a medical theory to test or when we have a relationship in mind that has some importance for medical decision or policy analysis in public health. Broncho-pulmonary dysplasia (BPD), a chronic type of lung disease prevalent among infants; using the discriminant model and logistic regression is of interest to this study. In most cases, the model is used to make predictions in either the testing of a medical theory or the study of a policy's impact in pharmacy and public health, [16] Broncho pulmonary dysplasia is a form of chronic lung disease that develops in preterm neonates treated with oxygen and positive-pressure ventilation. The pathogenesis of this condition remains complex and poorly understood; however various factors can not only injure small airways but also interfere with alveolarization (alveolar septation), leading to alveolar simplification with a reduction in the overall surface area for gas exchange. The developing pulmonary microvasculature can also be injured. Many infants born with Broncho pulmonary dysplasia exhibit signs and symptoms of respiratory distress syndrome, including the following: Tachypnea, Tachycardia, Increased respiratory effort (with retractions, nasal flaring, and grunting), frequent desaturations. Some statistical models in medical research

may contain dichotomous factor; in form of a person is male or female; a person does or does not have a disease in question, to mention but a few. In this study, we shall particularly build discriminant models with prior information for predicting the Broncho-pulmonary dysplasia (BPD) status of infants using gender and weights at two different time intervals of survival of the infant as predictor variables.

1.1. Objectives of the Study

The following are specific objectives:

1. To build a discriminant model that is capable of tracking Broncho-pulmonary dysplasia (BPD) infants based on their weight at birth, weight four weeks later and gender.
2. To predict the Broncho-pulmonary dysplasia (BPD) of some infants using the linear discriminant model.
3. To compare and contrast the predictive powers of the discriminant model and logistic regression for BPD.

1.2. Literature Review

It is pertinent to begin every research work, particularly in this kind of statistical modelling, by outlining how other relevant literatures were consulted. A review only those items relevant to the dissertation work has been made in this section, which has an immediate bearing to this work at hand. Broncho pulmonary dysplasia (BPD) continues to be a major cause of chronic morbidity among this population. In 2002, Denan [5] observed that there are large variations in the incidence and severity of this disease. According to the National Institutes of Health of USA (NICHD) consensus [7], the most recent report of the incidence of BPD in Latin America comes from the Neonatal Group study a very-low-birth-weight (VLBW) infants some Asian part of the world. Tapia [15] examined that the BPD is a chronic pulmonary disease which affects premature infants and contributes to their morbidity and mortality. Despite substantial changes in incidence, risk factors and severity after the introduction of new therapies and mechanical ventilation (MV) techniques, BPD remains common, for more details we refer Tapia [15]. Moreover, as per perception of Brunnella [2], the low MMP-2 level at birth is strongly associated with the development of BPD. Recently in 2011, Vimal Kumar et al [18] conducted a study to determine the prevalence risk factors of Nephropathy in type-2 diabetic patients. Here, it is remarked that Vimal Kumar et al [18] discovered that as the duration of type-2 diabetes increases, the incidence of Nephropathy also increases significantly. Hence, all the type-2 diabetic patients, especially those with increased duration should be screened for Nephropathy and be made aware of the complications. Carlos et al [3] conducted a research which involves the building of model for the prediction of Broncho-pulmonary dysplasia model for seven-day old infants and their aim was to develop a predictive model capable of identifying which premature infants have the greatest probability of presenting (BPD) based on assessment at the end of the first week of life. Carlos et al [3] concluded that at the end of the first week of life, the predictive model they developed was capable of identifying newborn infants at increased risk of developing BPD with high degree of sensitivity. Boule et al [1] proposed that adaptive control effects of exercise on glycemic control and body mass in type 2 diabetes mellitus is generally access by clinical trials.

2. Materials and Methodology

In this research design, having considered all the factors involved, the simple random sampling is the chosen sampling design.

Consider the three selected predictor variables which are capable of characterizing a BPD infant. From experience and records of medical practice, these variables are also believed to vary significantly between normal infants (π_1) and BPD infants (π_2). These variables are;

For the Euclidean distance, we need the mean vectors and the covariance matrices of a sample of both normal infants (π_1) and BPD infants (π_2) [17].

For normal infants

$$\bar{\mathbf{X}}_1 = \begin{pmatrix} \bar{X}_{11} \\ \bar{X}_{12} \\ \bar{X}_{13} \end{pmatrix}$$

And

$$\mathbf{S}_1 = \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{22} \end{pmatrix}$$

For BPD infants

$$\bar{\mathbf{X}}_2 = \begin{pmatrix} \bar{X}_{21} \\ \bar{X}_{22} \\ \bar{X}_{23} \end{pmatrix}$$

And

$$\mathbf{S}_2 = \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{22} \end{pmatrix}$$

Hence, the Euclidean distance of the population of normal infants (π_1) is

$$\hat{\ell}_1 = \bar{X}_1' S_p^{-1} \left(\bar{X}_1 - \bar{X}_2 \right)$$

Similarly the Euclidean distance of the population of BPD infants (π_2) is

$$\hat{\ell}_2 = \bar{X}_2' S_p^{-1} \left(\bar{X}_1 - \bar{X}_2 \right)$$

Where, S_p is the pooled covariance matrix. In view of Usman [17] the mean Euclidean distance is given by

$$\hat{M} = \tfrac{1}{2}\left(\hat{\ell}_1 + \hat{\ell}_2\right)$$

The discriminant function is therefore obtained as follows:

$$\hat{Y} = X' S_p^{-1}\left(\bar{X}_1 - \bar{X}_2\right)$$

Using some empirical data, the classification rule is that

$$\hat{Y} \geq \hat{M}$$

Classify as BPD infants (π_2) if

Otherwise the infant is normal. This is further exemplified in the following table

Table 1. *Decision criteria for detecting health status*

S/No	Outcome	Criterion	Health Status
1	$\hat{Y}_1$	$\hat{Y}_1 > \hat{M}$	BPD infant
2	$\hat{Y}_2$	$\hat{Y}_2 < \hat{M}$	Normal infant
3	$\hat{Y}_3$	$\hat{Y}_3 < \hat{M}$	Normal infant
4	$\hat{Y}_4$	$\hat{Y}_4 > \hat{M}$	BPD infant
5	$\hat{Y}_5$	$\hat{Y}_5 < \hat{M}$	Normal infant

3. Statistical Methods

3.1. Plans and Methods of Data Collection

Secondary means of data collection was employed through a well design clinical survey and the study was conducted in Maiduguri with data obtained from UMTH Teaching Hospital, Maiduguri and Sokoto with data obtained from UDUS Teaching Hospital, Sokoto. The birth weight (g), weight four weeks after birth (g) and sex were recorded, collected and tabulated for 50 and 20 infants respectively.

3.2. Discriminant Analysis

Discriminant analysis (DA) and classification are multivariate techniques concern with separating distinct sets of objects (or observations) and with allocating new objects (or observations) to previously defined groups. Discriminant analysis is rather exploratory in nature. As a classificatory procedure, it is often employed on a one-time basis in order to investigate observed differences when causal relationships are not well understood [17]. Classification procedures are less exploratory in the sense that they lead to well-defined rules, which can be used for assigning new objects. It is possible to have classifications into two or more multivariate normal populations, but in this case, we shall limit ourselves to classifications into two normal populations denoted by π_1 and π_2.

The methodologies of discriminant analysis were first developed by an English statistician – Fisher [6] who is arguably the most influential statistician of the twentieth century. He was educated at Cambridge University, where he studied Mathematics. During his time, Fisher [6] virtually invented the subject of experimental design and ANOVA, which motivated his derivation of the F-distribution. Moreover, the methodologies of discriminant analysis were later improved upon by a renowned Indian statistician called Chandra Prasanta Mahalanobis**,** in 1945. As an adviser to the Indian Government, Mahalanobis established in 1950, the first national survey of Indian population.

3.2.1. Theory of Discriminant Analysis

Suppose we have two multivariate normal populations with equal variance-covariance matrices,

$$N\left(\mu_1, \Sigma\right) and, N\left(\mu_2, \Sigma\right) where, \mu_i (i = 1,2), \left(\mu_1, \mu_2, ..., \mu_p\right)'$$

is the vector of means of the ith population and is the variance-covariance matrices of the two populations. The pdf of ith population is given below:

$$P_i(X) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left[-\tfrac{1}{2}(X-\mu_i)' \Sigma^{-1}(X-\mu_i)\right] \quad (1)$$

In the light of Usman [17] the ratio of the densities of two multivariate normal populations is given below;

$$\exp\left[-\tfrac{1}{2}\left\{(X-\mu_1)' \Sigma^{-1}(X-\mu_1) - (X-\mu_2)' \Sigma^{-1}(X-\mu_2)\right\}\right] \geq k$$

$$\frac{P_1(X)}{P_2(X)} = \frac{\exp\left[-\tfrac{1}{2}(X-\mu_1)' \Sigma^{-1}(X-\mu_1)\right]}{\exp\left[-\tfrac{1}{2}(X-\mu_2)' \Sigma^{-1}(X-\mu_2)\right]} \geq k$$

Taking the natural logarithms of the first inequality above; which is monotone increasing we have:

$$-\tfrac{1}{2}\left\{(X-\mu_1)' \Sigma^{-1}(X-\mu_1) - (X-\mu_2)' \Sigma^{-1}(X-\mu_2)\right\} \geq \log k \quad (2)$$

The second term of (2) above is the Mahalanobis square distance between and for k suitably chosen (which of course can be one and then log k will be zero), the LHS of (2) can be expanded and rearranged to obtain the following:

$$X' \Sigma^{-1}(\mu_1 - \mu_2) - \tfrac{1}{2}(\mu_1 + \mu_2)' \Sigma^{-1}(\mu_1 - \mu_2) \geq \log k \quad (3)$$

The first term of the inequality in (3) is the well-known Fisher's linear discriminant function which is linear in the component of the observation vector.

3.2.2. Criterion for Classification

The best regions of classification into and are given by:

$$\pi_1 \Rightarrow if \;\; X' \Sigma^{-1}(\mu_1 - \mu_2) - \tfrac{1}{2}(\mu_1 + \mu_2)' \Sigma^{-1}(\mu_1 - \mu_2) \geq \log k$$

$$\pi_2 \Rightarrow if \; X' \Sigma^{-1}(\mu_1 - \mu_2) - \tfrac{1}{2}(\mu_1 + \mu_2)' \Sigma^{-1}(\mu_1 - \mu_2) < \log k$$

$$k = \frac{q_1 C(1/2)}{q_2 C(2/1)} \quad (4)$$

Where; $C(1/2)$ is the cost of misclassifying an observation into π_1 instead of π_2

And $C(2/1)$ is the cost of misclassifying an observation into π_2 instead of π_1. But if the two populations are equally likely, and the costs of misclassifications being equal, k = 1 and log k = 0 which agrees with results of Singh R.K.[14]

Hence, the region of classification into and can further simplified as follows:

$$\pi_1 \Rightarrow if\ X^{'}\Sigma^{-1}(\mu_1-\mu_2) \geq \tfrac{1}{2}(\mu_1+\mu_2)^{'}\Sigma^{-1}(\mu_1-\mu_2)$$

$$\pi_2 \Rightarrow if\ X^{'}\Sigma^{-1}(\mu_1-\mu_2) < \tfrac{1}{2}(\mu_1+\mu_2)^{'}\Sigma^{-1}(\mu_1-\mu_2)$$

$$\pi_1 \Rightarrow if\ Y \geq \hat{M}$$

Or

$$\pi_2 \Rightarrow if\ Y < \hat{M}$$

Where,

$$Y = X^{'}\Sigma^{-1}(\mu_1-\mu_2)$$

In practice, all the population parameters can be estimated with their respective sample statistics;

μ_2 Can be estimated by $\bar{X}_1$

μ_1 Can be estimated by $\bar{X}_2$

Σ Can be estimated by the pooled variance, S_p (Usman, 2011), so that,

$$\mathbf{Y} = \mathbf{X}^{'}\mathbf{S_p^{-1}}(\mathbf{\bar{X}_1}-\mathbf{\bar{X}_2})$$

$$\mathbf{\hat{M}} = \tfrac{1}{2}(\mathbf{\bar{X}_1}+\mathbf{\bar{X}_2})^{'}\mathbf{S_p^{-1}}(\mathbf{\bar{X}_1}-\mathbf{\bar{X}_2})$$

$$\hat{M} = \tfrac{1}{2}(\mu_1+\mu_2)^{'}\Sigma^{-1}(\mu_1-\mu_2)$$

Where,

$$\mathbf{X}^{'} = (X_1 \quad X_2)$$

$$S_p = \frac{n_1S_1+n_2S_2}{n_1+n_2}$$

If then the $n_1 = n_2$

$$S_p = \frac{n_1S_1+n_2S_2}{n_1+n_2} \sin ce\ n_1 = n_2,$$

$$i.e\ S_p = \frac{n_1S_1+n_1S_2}{n_1+n_1},$$

$$= \frac{n_1}{n_1}\left(\frac{S_1+S_2}{1+1}\right) = \frac{S_1+S_2}{2}$$

Estimated by the pooled variance

Where S_1 and S_2 are the respective sample variance covariance matrices of the two populations.

$$S_p = \frac{S_1+S_2}{2}$$

3.3. Hypothesis

3.3.1. Wilk's Lambda Test for Canonical Correlation

(i) Hypothesis Canonical Correlation

H_0: There no linear relationship between the sets of variables

H_1: There linear relationship between the sets of variables

(ii) Test Statistic

$$\lambda = \frac{|W|}{|W+H|}$$

Where W is residual variance

H is variance due to linear relationship

W+H is the total variance

(iii) Decision Rule

Reject H_0 if $p < 0.05$ otherwise accept H_0 at the 5% level of significance.

3.3.2. Omnibus Chi-Square Test

The omnibus Chi-square test is a log-likelihood ratio test for investigating the Discriminant model coefficients. The test procedures are as follows:

(i) Hypothesis for Omnibus Chi-Square Test

H_0: The model coefficients are not tatistically significant

H_1: The model coefficients are statistically significantTest statistic:

$$\chi^2 = 2\left[\sum_{i=1}^{r}\sum_{j=1}^{c} O_{ij}\ln\left(\frac{O_{ij}}{e_{ij}}\right)\right]$$

Or

$$\chi^2 = 2\left[\sum_{i=1}^{r}\sum_{j=1}^{c} O_{ij}\ln O_{ij} - \sum_{i=1}^{r} R_i \ln R_i - \sum_{j=1}^{c} C_j \ln C_j - n\ln n\right]$$

(ii) Decision Rule

Reject H_0 if p<0.05 otherwise accept H_0 at the 5% level of significance. Significance of the model coefficient in the Discriminant model. Hence, the Omnibus test is applied.

3.3.3. Box's M Test for the Equality of Covariance Matrices

(i) Hypothesis for Box's M Test

H_0: The two covariance matrices are equal.

H_1: The two covariance matrices are not equal.

(ii) Test Statistic

$$M = \frac{|S_L|}{|S_S|}$$

Where s_L is the larger variance and s_s is the smaller variance.

(iii) Decision Rule

Reject H_0 if $p<0.05$ otherwise accept H_0 at the 5% level of significance

3.4. Software Used for the Study

Statistical package for social science IBM SPSS Statistics 22 was employed for the data analysis.

4. Results and Analysis

4.1. Discriminant Model for Borno State

Table 2. *Group statistics for UMTH, Maiduguri*

Health status	Variables	Mean	Standard Deviation	N
Healthy	Weight at birth (g)	1275.19	245.973	27
	Weight at four weeks (g)	1840.37	356.073	27
	Sex	1.33	0.480	27
BPD	Weight at birth (g)	930.87	217.964	23
	Weight at four weeks (g)	1344.78	317.402	23
	Sex	1.83	0.388	23
Total	Weight at birth (g)	1116.80	288.935	50
	Weight at four weeks (g)	1612.40	418.045	50
	Sex	1.56	0.501	50

Using the SPSS output above, for Borno State, we have the mean vector and dispersion matrix for group 1 are as follows:

$$\bar{\mathbf{X}}_1 = \begin{pmatrix} \bar{X}_{11} \\ \bar{X}_{12} \\ \bar{X}_{13} \end{pmatrix} = \begin{pmatrix} 1275.19 \\ 1840.37 \\ 1.33 \end{pmatrix}$$

Where;

$$\bar{X}_{1i} = \frac{1}{n}\sum_{i=1}^{n} X_{1i}$$

And the mean vector and dispersion matrix for group 2 are as follows:

$$\bar{\mathbf{X}}_2 = \begin{pmatrix} \bar{X}_{21} \\ \bar{X}_{22} \\ \bar{X}_{23} \end{pmatrix} = \begin{pmatrix} 930.87 \\ 1344.78 \\ 1.83 \end{pmatrix}$$

Where;

$$\bar{X}_{2i} = \frac{1}{n}\sum_{i=1}^{n} X_{2i}$$

Table 3. *Pooled covariance matrices*

Covariance	Weight at birth (g)	Weight at 4 weeks (g)	Sex
Weight at birth (g)	83483.429	120744.571	-101.437
Weight at four weeks (g)	120744.571	174761.469	-147.086
Sex	-101.437	-147.086	0.251

From table 3 above, the pooled variance-covariance matrix is as follows:

$$\mathbf{S} = \frac{n_1\mathbf{S}_1 + n_2\mathbf{S}_2}{n_1 + n_2} = \begin{pmatrix} 83483.429 & 120744.571 & -101.437 \\ 120744.571 & 174761.469 & -147.086 \\ -101.437 & -147.086 & 0.251 \end{pmatrix}$$

Table 4. *Test results for UMTH, Maiduguri*

Box's M		4.524
F	Approx.	7.02
	df1	6
	df2	15524.259
	Sig.	0.012

Tests null hypothesis of equal population covariance matrices

The F- value of and the Box's M P- value of 0.012 has confirmed the equality of the Covariance matrices from the two groups.

Table 5. *Wilk's lambda test result*

Test of Function	Wilks' Lambda	Chi-square	df	Sig.
1	0.615	22.603	3	0.000

To justify the significant of the canonical correlation Wilk's Lambda statistic give 0.615 with p-value of 0.00.Comparing the p- value of Wilk's Lambda of 0.000 with the predefine significant level α =0.05, then the canonical correlation computed is very significant. Since P-value (0.00) < (0.05).

Table 6. *Fisher's linear discriminant function (Maiduguri) Fisher's classification function coefficients*

	Health status	
	Healthy	BPD patients
Weight at birth (g)	0.039	-0.015
Weight at four weeks (g)	0.005	0.039
Sex	21.229	22.252
(Constant)	-44.529	-40.177

Fisher's linear discriminant functions

The Fishers linear discriminant model for each group is

computed as follows:

Normal infants π_1

$$\mathbf{Y_1 = X' S^{-1}\left(\bar{X}_2 - \bar{X}_1\right)}$$

$$\mathbf{Y_1 = -44.529 + 0.039X_{1i} + 0.005X_{2i} + 21.229X_{3i}}$$

Infant BPD π_2

$$\mathbf{Y_2 = X' S^{-1}\left(\bar{X}_2 - \bar{X}_1\right)}$$

$$\mathbf{Y_2 = -189.104 + 0.518X_{1i} + 2.268X_{2i} + 0.657X_{3i}}$$

The classification rule is to substitute into the Fishers linear discriminant model for each group and evaluate; then classify into the group whose model produced the higher discriminant score. This criterion is absolutely equivalent to the unstandardized linear discriminant model.

4.2. Unstandardized Discriminant Function

Table 7. *Unstandardized classification function coefficients*

	Function 1
Weight at birth (g)	0.035
Weight at four weeks (g)	-0.022
Sex	-0.658
(Constant)	-2.860

Table 8. *Functions at group centroids*

Health status	Function 1
Healthy	0.715
BPD patients	-0.840

Unstandardized canonical discriminant functions evaluated at group means

The Cutoff point ($\hat{\mathbf{M}}$) is computed as follows:

$$\therefore \hat{\mathbf{M}} = \tfrac{1}{2}\left(\hat{\mathbf{I}}_1 + \hat{\mathbf{I}}_2\right) = \tfrac{1}{2}\left(\mathbf{0.715 - 0.840}\right) = \mathbf{-0.063}$$

The Unstandardized discriminant model is computed as follows:

$$\mathbf{Y = -2.860 + 0.035X_{1i} - 0.022X_{2i} - 0.658X_{3i}}$$

$$\mathbf{Y_1 = -2.860 + 0.035(1100) - 0.022(1400) - 0.658(2)}$$
$$\mathbf{= 3.52 > -0.063}$$

$$\mathbf{Y_2 = -2.860 + 0.035(1550) - 0.022(2240) - 0.658(2)}$$
$$\mathbf{= 1.452 > -0.063}$$

$$\mathbf{Y_3 = -2.860 + 0.035(790) - 0.022(1130) - 0.658(2)}$$
$$\mathbf{= -1.386 < -0.063}$$

$$\mathbf{Y_4 = -2.860 + 0.035(1480) - 0.022(2140) - 0.658(1)}$$
$$\mathbf{= 1.202 > -0.063}$$

$$\mathbf{Y_5 = -2.860 + 0.035(980) - 0.022(1240) - 0.658(2)}$$
$$\mathbf{= -1.116 > -0.063}$$

4.3. Classifying BPD Status Using the Discriminant Model

It is pertinent to use the discriminant model to classify BPD status of infants using the discriminant model:

$$\mathbf{Y = -2.860 + 0.035X_{1i} - 0.022X_{2i} - 0.658X_{3i}}$$

$$\mathbf{Y = -2.860 + 0.035X_{1i} - 0.022X_{2i} - 0.658X_{3i}}$$

The classification rule that is as follows:

Classify as Group 1 (Healthy) if $\mathbf{Y < -0.063}$

Classify as Group 2 (BPD) if $\mathbf{Y \geq -0.063}$

To compute estimates or forecasts, consider the discriminant model as given below:

$$\mathbf{Y = -2.860 + 0.035X_{1i} - 0.022X_{2i} - 0.658X_{3i}}$$

That will be used to predict the BPD status of infants. Using the discriminant model and the following table containing the data of additional five infants whose BPD status is already known, the model is hereby tested for goodness of fit and classificatory power:

$$\mathbf{Y = -2.860 + 0.035X_{1i} - 0.022X_{2i} - 0.658X_{3i}}$$

Table 9. *New set of observations for UMTH, Maiduguri*

SNO	**Weight 1 (g) (X_1)**	**Weight 2 (g) (X_2)**	**Sex (X_3)**	**Health Status**
1	1000	1400	F	Healthy
2	1550	2440	F	Healthy
3	790	1130	F	BPD
4	1480	2140	M	Healthy
5	980	1420	F	BPD

$$\mathbf{Y = -2.860 + 0.035X_{1i} - 0.022X_{2i} - 0.658X_{3i}}$$

Table 10. *Classification of new observations for UMTH Maiduguri Classified as group 1 (Healthy)*

SNO	**Weight 1 (g) (X_1)**	**Weight 2 (g) (X_2)**	**Sex (X_3)**	**Health Status**	**Predicted Status**
1	1000	1400	F	Healthy	Healthy
2	1550	2440	F	Healthy	Healthy
3	790	1130	F	BPD	BPD
4	1480	2140	M	Healthy	Healthy
5	980	1420	F	BPD	BPD

Classified as group 1 (Healthy)

Classified as group 2 (BPD patient)

Classified as group 1 (Healthy)

Classified as group 2 (BPD patient)

Comparative Table for BPD Status

For the new observations, the probabilities of

misclassifications are obtained as follows:

$$\mathbf{p(group\ 1\ /\ group\ 2)} = \frac{0}{5} = 0$$

$$\mathbf{p(group\ 2\ /\ group\ 1)} = \frac{0}{5} = 0$$

The model is also tested for goodness of fit and classificatory power for new observations. The discriminant model has no misclassified case for UMTH, Maiduguri which proves to be very good.

5. Discussion of Findings

The discriminant model has a perfect classification of five new cases in UMTH, Maiduguri, while it has misclassified one of five new cases in UDUTH, Sokoto. Hence, from this analysis, the power of a classificatory model depends on the situation, location or even the data at hand. However, from the angle of statistical inference, if normality assumption fails, the logistic models are more preferred to the discriminant models. But if the normality assumption holds, the principle of parsimony prevails which model fits better for a particular situation must be determined using the goodness of fit results. In this case, the prediction of BPD is better done with discriminant model in Maiduguri but is moderately good in Sokoto.

6. Summary and Conclusions

In this study, linear discriminant model was applied to Broncho-Pulmonary Dysplasia (BPD) data. The result of present shows that the prediction of Broncho pulmonary Dysplasia (BPD) is better done with discriminant model in Maiduguri. However, we remark here that the present study has several limitations also related to cross-sectional and single-center studies, although it contextualizes in recent years in the assessed unit, which has been in operation for six years. Because of its characteristics, the present study is not a cause and effect study but one of association. Reducing the incidence of prematurity is the most effective way to alleviate BPD.

7. Recommendations

On the basis of our findings and conclusions, we recommend as followings:

1. In the light of the above it is recommended that Doctors and Clinics should adopt the use of the models built by this research to detect prevalence of BPD among infants so that adequate measures for prevention and control of BPD can be taken early enough to alert the danger of the full manifestation of the disease.
2. It is also recommended that the Discriminant model built should be used for cases UMTH, Maiduguri and Logistic regression model built should be used for cases in UDUTH,

References

[1] Boulé N.G., Haddad E., Kenny G.P., Wells G.A. and Sigal R.J., Effects of exercise on glycemic control and body mass in type 2 diabetes mellitus: a meta-analysis of controlled clinical trials. The Journal of the American Medical Association, Vol. 286(10), pp.1218-1227, 2001

[2] Brunnella Alcantara Chagas de freitas[i,II]; Mirene peloso[i]; Guilherme Lobo da silveira[iii]; Giana Zarbato Longo, Journal of Clinical research, Prevalence and factors associated with bronchopulmonary dysplasia in a referral hospital in Minas Gerais, Brazil, 2012

[3] Carlos A., Bhering Christieny C., Mochdece M., Moreira E.L., José R. Rocco, Guilherme M. Sant'Anna, Bronchopulmonary dysplasia prediction model for 7-day-old infants Journal of Pediatria. (Rio J.), Vol. 83(2), Porto Alegre, March/April 2007

[4] Croux C. and Dehon C., Influence of observations on the misclassification probability in quadratic discriminant analysis. Journal of Multivariate Analysis, Vol. 96, pp. 384-403, 2005

[5] Danan C., Gelatinase activities in the airways of premature infants and development of bronchopulmonary dysplasia. American Journal of Physiol. Lung Cell Mol Physiol., Vol. 283: L1086- L1093, 2002

[6] Fisher R.A., The statistical utilization of multiple measurements. Annals of Eugenics, Vol. 8, pp.376–386, 1938

[7] Jobe H. and Bancalari E., Bronchopulmonary dysplasia, American Journal of Respiratory and Critical Care Medicine, Vol. 163(7), pp. 1723–1729, 2001

[8] Maurya V.N., Arora D.K. and Maurya A.K., Elements of Advanced Probability Theory and Statistical Techniques, Scholar's Press Publishing Co., Saarbrucken, Germany, 2014, ISBN 978-3-639-51849-8

[9] Maurya V.N., Arora Diwinder Kaur, Maurya Avadhesh Kumar and Gautam R.A., Exact modeling of annual maximum rainfall with Gumbel and Frechet distributions using parameter estimation techniques, *World of Sciences Journal*, Engineers Press Publishing, Vienna, Austria, Vol. 1(2), pp.11-26, 2013, ISSN: 2307-3071

[10] Maurya V.N., Maurya A.K. and Kaur D., A survey report on nonparametric hypothesis testing including Kruskal-Wallis ANOVA and Kolmogorov–Smirnov goodness-fit-test, *International Journal of Information Technology & Operations Management,* Academic and Scientific Publishing, New York, USA, Vol.1(2), pp. 29-40, 2013, ISSN: 2328-8531

[11] Maurya V.N., Numerical simulation for nutrients propagation and microbial growth using finite difference approximation technique, *International Journal of Mathematical Modeling and Applied Computing,* Academic & Scientific Publishing, New York, USA, Vol. 1(7), pp. 64-76, November 2013, ISSN: 2332-3744

[12] Maurya V.N., Singh V.V. and Yusuf Madaki Umar, Statistical analysis on the rate of kidney (renal) failure, *American Journal of Applied Mathematics and Statistics*, Special Issue: Application and Future Scope of Fundamental Mathematical and Computational Sciences in Engineering & Technology, Science & Education Publishing, USA, Vol. 2, No. 6A, pp. 6-12, 2014

[13] Rencher Alvin C., Methods of multivariate analysis / Alvin C. Rencher.—2nd ed. Page: 270-272, 2002, The Canadian Journal of Statistics, Vol.29, pp.473–492, 2001.

[14] Singh R.K., Multivariate Statistical Analysis. IAENG International Journal of Applied Mathematics, Vol. 39(1), pp.1-06, 2011

[15] Tapia J.L., Agost D., Alegria A. et al., Bronchopulmonary dysplasia: incidence, risk factors and resource utilization in a population of South American very low birth weight infants, Journal de Pediatria, Vol. 82(1), pp. 15–20, 2006

[16] Umejiaku C.A, Discriminant analysis of protein-energy malnutrition in children. Journal of Science, Vol.12-16, pp. 27-34, 2006.

[17] Usman A. Statistical methods for biometric and medical research, Millennium Printing ad Publishing Company Limited, Kaduna, Nigeria: 486-495.

[18] Vimal Kumar, Moses A. and Padmanaban, Prevalence and risk factors of nephropathy in type 2 Diabetic patients, Journal of Collaborative Research on Internal Medicine and Public Health, 2011.