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

Vishwa Nath Maurya¹, Madaki Umar Yusuf², Vijay Vir Singh², Babagana Modu²

¹Department of Mathematics and Statistics, School of Science & Technology, The University of Fiji, Lautoka, Fiji Islands
²Department 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:

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 Brunella [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, 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 (π₁) and BPD infants (π₂). These variables are;

For the Euclidean distance, we need the mean vectors and the covariance matrices of a sample of both normal infants (π₁) and BPD infants (π₂) [17].

For normal infants

\[ \bar{X}_1 = \begin{pmatrix} \bar{X}_{11} \\ \bar{X}_{12} \\ \bar{X}_{13} \end{pmatrix} \]

And

\[ S_1 = \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{pmatrix} \]

For BPD infants

\[ \bar{X}_2 = \begin{pmatrix} \bar{X}_{21} \\ \bar{X}_{22} \\ \bar{X}_{23} \end{pmatrix} \]

And

\[ S_2 = \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{pmatrix} \]

Hence, the Euclidean distance of the population of normal infants (π₁) is

\[ \hat{e}_1 = \hat{X}_1 S_1^{-1} (\bar{X}_1 - \bar{X}_2) \]

Similarly the Euclidean distance of the population of BPD infants (π₂) is

\[ \hat{e}_2 = \hat{X}_2 S_p^{-1} (\bar{X}_1 - \bar{X}_2) \]

Where, \( S_p \) is the pooled covariance matrix. In view of Usman [17] the mean Euclidean distance is given by
Discriminant analysis (DA) and classification are multivariate techniques concerned with separating distinct sets of objects (or observations) and allocating new objects. It is less exploratory in the sense that they lead to well-defined rules, which can be used for assigning new objects. Fisher’s linear discriminant function which is linear in the component of the observation vector.

### 3.2. Discriminant Analysis

Fisher’s linear discriminant function which is linear in the component of the observation vector.

\[ \hat{M} = \frac{1}{2} \left( \hat{\ell}_1 + \hat{\ell}_2 \right) \]

The discriminant function is therefore obtained as follows:

\[ \hat{y} = X^T \Sigma_p^{-1} \left( \bar{x}_1 - \bar{x}_2 \right) \]

Using some empirical data, the classification rule is that

\[ \hat{y} \geq \hat{M} \]

Classify as BPD infants (\( \pi_2 \)) if

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

### Table 1. Decision criteria for detecting health status

<table>
<thead>
<tr>
<th>S/No</th>
<th>Outcome</th>
<th>Criterion</th>
<th>Health Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \hat{y}_1 )</td>
<td>( \hat{y}_1 &gt; \hat{M} )</td>
<td>BPD infant</td>
</tr>
<tr>
<td>2</td>
<td>( \hat{y}_2 )</td>
<td>( \hat{y}_2 &lt; \hat{M} )</td>
<td>Normal infant</td>
</tr>
<tr>
<td>3</td>
<td>( \hat{y}_3 )</td>
<td>( \hat{y}_3 &lt; \hat{M} )</td>
<td>Normal infant</td>
</tr>
<tr>
<td>4</td>
<td>( \hat{y}_4 )</td>
<td>( \hat{y}_4 &gt; \hat{M} )</td>
<td>BPD infant</td>
</tr>
<tr>
<td>5</td>
<td>( \hat{y}_5 )</td>
<td>( \hat{y}_5 &lt; \hat{M} )</td>
<td>Normal infant</td>
</tr>
</tbody>
</table>

### 3.2.1. Theory of Discriminant Analysis

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

\[ N(\mu_1, \Sigma) \text{ and } N(\mu_2, \Sigma) \]

where, \( \mu_i \) (\( i = 1, 2 \)), \( \mu_1, \mu_2, \ldots, \mu_p \)' 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:

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

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

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

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

\[ -\frac{1}{2} \left( (X - \mu_1)^T \Sigma^{-1} (X - \mu_1) - (X - \mu_2)^T \Sigma^{-1} (X - \mu_2) \right) \geq \log k \]

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:

\[ X^T \Sigma^{-1} (\mu_1 - \mu_2) - \frac{1}{2} (\mu_1 + \mu_2) \Sigma^{-1} (\mu_1 - \mu_2) \geq \log k \]

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:

\[ \pi_1 \Rightarrow if \quad X^T \Sigma^{-1} (\mu_1 - \mu_2) - \frac{1}{2} (\mu_1 + \mu_2) \Sigma^{-1} (\mu_1 - \mu_2) \geq \log k \]

\[ \pi_2 \Rightarrow if \quad X^T \Sigma^{-1} (\mu_1 - \mu_2) - \frac{1}{2} (\mu_1 + \mu_2) \Sigma^{-1} (\mu_1 - \mu_2) < \log k \]

\[ k = \frac{q_C(1/2)}{q_C(2/1)} \]
Where; \( C(1/2) \) is the cost of misclassifying an observation into \( \pi_1 \) instead of \( \pi_2 \).

And \( C(2/1) \) is the cost of misclassifying an observation into \( \pi_2 \) instead of \( \pi_1 \). 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:

\[
\pi_1 \Rightarrow if \quad X' \Sigma^{-1} (\mu_1 - \mu_2) \geq \frac{1}{2} (\mu_1 + \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2) \\
\pi_2 \Rightarrow if \quad X' \Sigma^{-1} (\mu_1 - \mu_2) < \frac{1}{2} (\mu_1 + \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2) \\
\pi_i \Rightarrow if \quad Y \geq \hat{M} \\
\text{Or} \\
\pi_2 \Rightarrow if \quad Y < \hat{M}
\]

Where,

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

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

\( \mu_1 \) Can be estimated by \( \bar{X}_1 \)

\( \mu_2 \) Can be estimated by \( \bar{X}_2 \)

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

\[
\hat{M} = \frac{1}{2} (\bar{X}_1 + \bar{X}_2)' S_p^{-1} (\bar{X}_1 - \bar{X}_2)
\]

\[
\hat{M} = \frac{1}{2} (\mu_1 + \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2)
\]

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

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 statistically significant

\( H_1 \): The model coefficients are statistically significant

Test statistic:

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

Or

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

(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

\[ M = \frac{S_1}{S_2} \]

Where \( s_1 \) is the larger variance and \( s_2 \) 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>
<thead>
<tr>
<th>Health status</th>
<th>Variables</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>Weight at birth (g)</td>
<td>1275.19</td>
<td>245.973</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Weight at four weeks (g)</td>
<td>1840.37</td>
<td>356.073</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>1.33</td>
<td>0.480</td>
<td>27</td>
</tr>
<tr>
<td>BPD</td>
<td>Weight at birth (g)</td>
<td>930.87</td>
<td>217.964</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Weight at four weeks (g)</td>
<td>1344.78</td>
<td>317.402</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>1.83</td>
<td>0.388</td>
<td>23</td>
</tr>
<tr>
<td>Total</td>
<td>Weight at birth (g)</td>
<td>1116.80</td>
<td>288.935</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Weight at four weeks (g)</td>
<td>1612.40</td>
<td>418.045</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>1.56</td>
<td>0.501</td>
<td>50</td>
</tr>
</tbody>
</table>

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

\[
\overline{X}_1 = \begin{pmatrix}
X_{11} \\
X_{12} \\
X_{13}
\end{pmatrix} = \begin{pmatrix}
1275.19 \\
1840.37 \\
1.33
\end{pmatrix}
\]

Where;

\[
\overline{X}_{1i} = \frac{1}{n} \sum_{j=1}^{n} X_{ij}
\]

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

\[
\overline{X}_2 = \begin{pmatrix}
X_{21} \\
X_{22} \\
X_{23}
\end{pmatrix} = \begin{pmatrix}
930.87 \\
1344.78 \\
1.83
\end{pmatrix}
\]

Where;

\[
\overline{X}_{2i} = \frac{1}{n} \sum_{j=1}^{n} X_{2ij}
\]

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

\[
S = \frac{n_1 S_1 + n_2 S_2}{n_1 + n_2} = \begin{pmatrix}
83483.429 & 120744.571 & -101.437 \\
120744.571 & 174761.469 & -147.086 \\
-101.437 & -147.086 & 0.251
\end{pmatrix}
\]

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>
<thead>
<tr>
<th>Test of Function</th>
<th>Wilks’ Lambda</th>
<th>Chi-square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.615</td>
<td>22.603</td>
<td>3</td>
<td>0.000</td>
</tr>
</tbody>
</table>

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 predefined significant level \( \alpha = 0.05 \), then the canonical correlation computed is very significant. Since P-value \( (0.00) < (0.05) \).

<table>
<thead>
<tr>
<th>Test of Function</th>
<th>Wilks’ Lambda</th>
<th>Chi-square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
</table>
| Fisher's linear discriminant function (Maiduguri) Fisher's classification function coefficients

<table>
<thead>
<tr>
<th>Health status</th>
<th>Weight at birth (g)</th>
<th>Weight at four weeks (g)</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>0.039</td>
<td>-0.015</td>
<td>0.039</td>
</tr>
<tr>
<td>BPD patients</td>
<td>0.005</td>
<td>0.039</td>
<td>0.039</td>
</tr>
<tr>
<td>(Constant)</td>
<td>-44.529</td>
<td>-40.177</td>
<td>-40.177</td>
</tr>
</tbody>
</table>

Fisher's linear discriminant functions

The Fisher's linear discriminant model for each group is...
computed as follows:

Normal infants $\pi_1$

\[ Y_1 = X S^{-1} (\bar{X}_2 - \bar{X}_1) \]

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

Infant BPD $\pi_2$

\[ Y_2 = X S^{-1} (\bar{X}_2 - \bar{X}_1) \]

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

The classification rule is to substitute into the Fisher 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

<table>
<thead>
<tr>
<th>Function 1</th>
<th>Weight at birth (g)</th>
<th>Weight at four weeks (g)</th>
<th>Sex</th>
<th>(Constant)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.035</td>
<td>-0.022</td>
<td>-0.658</td>
<td>-2.860</td>
</tr>
</tbody>
</table>

Table 8. Functions at group centroids

<table>
<thead>
<tr>
<th>Health status</th>
<th>Function 1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>0.715</td>
<td></td>
</tr>
<tr>
<td>BPD patients</td>
<td>-0.840</td>
<td></td>
</tr>
</tbody>
</table>

Unstandardized canonical discriminant functions evaluated at group means.

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

\[ \hat{M} = \frac{1}{2} (\hat{I}_1 + \hat{I}_2) = \frac{1}{2} (0.715 - 0.840) = -0.063 \]

The Unstandardized discriminant model is computed as follows:

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

\[ Y_1 = -2.860 + 0.035(1100) - 0.022(1400) - 0.658(1) \]

\[ = 3.52 > -0.063 \]

\[ Y_2 = -2.860 + 0.035(1550) - 0.022(2240) - 0.658(2) \]

\[ = 1.452 > -0.063 \]

\[ Y_3 = -2.860 + 0.035(790) - 0.022(1130) - 0.658(2) \]

\[ = -1.386 < -0.063 \]

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:

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

\[ Y = -2.860 + 0.035(980) - 0.022(1240) - 0.658(2) \]

\[ = -1.116 > -0.063 \]

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

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

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:

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

Table 9. New set of observations for UMTH, Maiduguri

<table>
<thead>
<tr>
<th>SNO</th>
<th>Weight 1 (g)</th>
<th>Weight 2 (g)</th>
<th>Sex (X)</th>
<th>Health Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1000</td>
<td>1400</td>
<td>F</td>
<td>Healthy</td>
</tr>
<tr>
<td>2</td>
<td>1550</td>
<td>2440</td>
<td>F</td>
<td>Healthy</td>
</tr>
<tr>
<td>3</td>
<td>790</td>
<td>1130</td>
<td>F</td>
<td>BPD</td>
</tr>
<tr>
<td>4</td>
<td>1480</td>
<td>2140</td>
<td>M</td>
<td>Healthy</td>
</tr>
<tr>
<td>5</td>
<td>980</td>
<td>1420</td>
<td>F</td>
<td>BPD</td>
</tr>
</tbody>
</table>

Table 10. Classification of new observations for UMTH Maiduguri

<table>
<thead>
<tr>
<th>SNO</th>
<th>Weight 1 (g)</th>
<th>Weight 2 (g)</th>
<th>Sex (X)</th>
<th>Health Status</th>
<th>Predicted Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1000</td>
<td>1400</td>
<td>F</td>
<td>Healthy</td>
<td>Healthy</td>
</tr>
<tr>
<td>2</td>
<td>1550</td>
<td>2440</td>
<td>F</td>
<td>Healthy</td>
<td>Healthy</td>
</tr>
<tr>
<td>3</td>
<td>790</td>
<td>1130</td>
<td>F</td>
<td>BPD</td>
<td>BPD</td>
</tr>
<tr>
<td>4</td>
<td>1480</td>
<td>2140</td>
<td>M</td>
<td>Healthy</td>
<td>Healthy</td>
</tr>
<tr>
<td>5</td>
<td>980</td>
<td>1420</td>
<td>F</td>
<td>BPD</td>
<td>BPD</td>
</tr>
</tbody>
</table>

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:

\[
p(\text{group 1} / \text{group 2}) = \frac{p_1}{p_2} = 0
\]

\[
p(\text{group 2} / \text{group 1}) = \frac{p_2}{p_1} = 0
\]

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 Bronchopulmonary 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


