INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

Statistical Analysis on Tackling Diabetic Patients among Adults in Damaturu, Nigeria

Umar Yusuf Madaki
Lecturer, Department of Mathematics and Statistics, Yobe State University, Damaturu, Nigeria
Abdulrahman Malik
Lecturer, Department of Statistics, Mai Idris Alooma Polytechnic Geidam, Yobe State, Nigeria
Maigana A. Bakawu
Lecturer, Department of Statistics, Mai Idris Alooma Polytechnic Geidam, Yobe State, Nigeria

Abstract:
This paper assesses the study that was carried out based on the broad objective of using the Box M and Wilks Lambda tests for tackling diabetic patients among adults in Damaturu, Yobe state. In this research work, 874 diabetic patients data was collected from the Yobe State Specialist Hospital, Damaturu for ten consecutive years (2008-2017).The data analysis were done using Box's M test and Wilks' Lambda test. The diabetic patient's responses were categorized into two groups which include: healthy ("0") and diabetic ("1"). The variables (predictors), that is; age, sex, weight, blood sugar and urine sugar are used based on discriminant analysis. By using Fisher's method of discriminant analysis, the classification table results were obtained while assumption of homogeneity of covariance matrices is tested with Box's M test. This test is very sensitive to meeting also the assumption of multivariate normality. For our data, the test is significant so we conclude the groups do differ in their covariance matrices, violating an assumption in discriminant analysis. Additionally, Wilks' Lambda of the function is 0.546 and has the value (Sig. 0.00), thus we can say that the model is a good fit for the data since p < .05. This multivariate test is a goodness of fit statistic, just like the F-test in the case of regression. 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 diabetics among adults so that adequate measures for prevention and control of diabetics can be taken early enough to alert the danger of the full manifestation of the disease.

Keywords: *Discriminant analysis, diabetes mellitus, box m and wilk's lambda test*

1. Introduction

Diabetes is an endocrinal and metabolic illness characterized by a partial or absolute deficit in the secretion of insulin, a hormone secreted by the beta cells of the pancreas. This deficit has multiple and diverse consequences in the organism, notable among which is the tendency to maintain inappropriately high levels of glucose in the blood (hyperglycemia). In order to avoid this occurring, diabetic patients have to inject themselves subcutaneously with insulin at regular intervals, in addition to exercising strict control over their diet and following the requirements of a complex program of treatment. For this reason, diabetes is a clear example of a metabolic disorder whose control depends on patients' behavior, so that, to a large extent, the course of their illness will be a function of their adherence to treatment. Diabetes mellitus is a group a metabolic disorders that present the phenotype of hyperglycemia. The etiologies of diabetes mellitus are a complex interaction of genetics, environmental factors and life-styles choices. The pathogenesis of` diabetes include reduce insulin secretion. Diabetes is classified into Type 1 diabetes (Insulin-dependent diabetes mellitus or IDDM) due to islet B-cell destruction, Type 2 diabetes (Non-insulin-dependent diabetes mellitus or NIDDM) with varying degree of resistance of insulin resistance and/or insulin secretary defects, other types of diabetes and gestational diabetes (where diabetes diagnosed for the first time in pregnancy). (Pickup and Williams, 2003).The clinical name of diabetes is D.M. Increased thirst and increased urination is the main symptom of this disease. The word mellitus proposed its name by Latin a word which means sweet. In this disease pancreas of the body does not use proper amount of insulin and this insulin changes the sugar and food into energy and this energy is very important for good health of body (Wells and Lawrence 2002).

Discriminant function analysis is a technique that has received lots of theoretical attention from statisticians as well as somewhat less attention from users. It is similar to multiple regressions and somewhat to Multivariate Analysis of Variance (MANOVA), and as we have seen, many of the other techniques use discriminant functions as part of the technique. Conceptionally, we can think of the discriminant function or equations defining the boundary between groups. The aim of discriminant analysis is to classify an observation, or several observations, into these known groups. 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". Multivariate methods are relevant in virtually every branch of applied medicine, pharmacy and public health.

MATERIALS AND METHODS

In this research design, having considered all the factors involved, the simple random sampling is the chosen sampling design. The five selected predictor variables Considered which are are capable of characterizing a Diabetic patient. From experience and records of medical practice, these variables are also believed to vary significantly between normal healthy (π_1) and diabetics (π_2). These variables are;

X_1 = Sex, where x_1 is coded as 1 for male and 2 for female
X_2= Age
X_3 = Weight (kg)
X_4= Blood sugar (mill moles per litre)
X_5= Urine sugar (mill moles per litre)

For the Euclidean distance, we need the mean vectors and the covariance matrices of a sample of both normal healthy (π_1) and diabetics (π_2).

2. Literature Review

In this section review of relevant literature which has immediate bearing to this work has been outline. Several researchers like: (Cook, 2010) and (Sun & Ye, 2016)) have contributed a lot in the field of medical and health statistics in the literature. (Maurya, Yusuf, Singh, and Modu, 2015) conducted a study to determine the prevalence risk factors of Nephropathy in type-2 diabetic patients. Their aim was to build a binary logistic model for predicting Nephropathy status among type-2 diabetic patients using age, sex, socio-economic status, and duration of Nephropathy history as covariates. (Usman, 2012)]Proposed that adaptive control effects of exercise on glycemic control and body mass in type 2 diabetes mellitus is generally access by clinical trials. (Bhering, Mochdece, Moreira, Rocco, & Sant'Anna, 2007) 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 Bronchopulmonary Dysplasia. Similarly, (Boulé, Haddad, Kenny, Wells, & Sigal, 2001) noted that adaptive control effects of exercise on glycemic controland body mass in type 2 diabetes mellitus is generally access by clinical trials.

2.1. The Box M Test of Equality of Covariance Matrices

The Box's M test MANOVA'S assumption of homoscedasticity and its uses the F distribution. If P(M) <0.05, the covariance is significantly different. We want M not to be significant that is rejecting the null hypothesis that the covariances are not homogeneous. Box's M test extremely sensitive to violations of the assumption of normality, this makes the Box's M test less attractive than might otherwise appear. For this reason, some researchers test at P=0.001 level, especially when sample sizes are unequal.

2.2. The Wilks' Lambda Test

This is the most common traditional test where there are more than two groups formed by the independent variables. Wilks' lambda is a multivariate F test. It is a measure of the difference between groups of the centroids (vector) of means on the independent variables, the smaller the lambda the greater the differences. The Bartlett's V transformation of lambda is then used to compute the significance of lambda. Wilks' lambda is used in conjunction with Bartlett's V as a multivariate significance test of means difference in MANOVA, for the case of multiple interval dependents and multiple (>2) groups formed by the independent(s). The T-test, Hotelling's T^2, and the F test are special cases of Wilks' lambda. According to (Tabachnick and Fidell, 2001), it is also used in a second context to assess the improvement in classification when using sequential discriminant analysis.

2.3. Wilks' Lambda Test for Significance of Canonical Correlation

2.3.1. Hypothesis Canonical Correlation

- H_0: There no linear relationship between the sets of variables
- H_1: There linear relationship between the sets of variables

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

2.3.2. Decision Rule
Reject H_0 if $p < 0.05$ otherwise accept H_0 at the 5% level of significance.
Box M Test for the Equality of Covariance Matrices
Hypothesis for Box's M Test:
H_0: The two covariance matrices are equal.
H_1: The two covariance matrices are not equal.
Test statistic:

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

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

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

3. Results and Discussions

3.1. Box's M Test of Equality of Covariance Matrices

Diabetic	Rank	Log Determinant
0	4	15.094
1	4	14.997
Pooled within-groups	4	14.123

Table 1: Log Determinants

3.2. Log Determinants and Box's M Tables (Table 1 and 2)

In ANOVA, an assumption is that the variances were equivalent for each group but in DA the basic assumption is that the variance-co-variance matrices are equivalent. Box's M tests the null hypothesis that the covariance matrices do not differ between groups formed by the dependent. The researchers want this test not to be significant so that the null hypothesis that the groups do not differ can be retained. For this assumption to hold, the log determinants should be equal. When tested by Box's M, we are looking for a non-significant M to show similarity and lack of significant differences. In this case the log determinants appear similar and Box's M is 96.658 with F = 8.705 which is significant at p < .05 (Tables 1 and 2). However, with large samples, a significant result is not regarded as too important. Where three or more groups exist, and M is significant, groups with very small log determinants should be deleted from the analysis.

	Box's M	96.658
F	Approx.	8.705
	df1	10
	df2	1234.002
	Sig.	.000

Table 2: Box's M Test Results

Box's M test tests the assumption of homogeneity of covariance matrices. This test is very sensitive to meeting also the assumption of multivariate normality. For our data, the test is significant so we conclude the groups do differ in their covariance matrices, violating an assumption in DA. However, discriminant functional analysis is robust even when the homogeneity of variances is not met, provided the data do not contain important outliers.

4. Summary of Canonical Discriminant Functions

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1.299[a]	100.0	100.0	.876

Table 3: Eigen Values

Canonical discriminant functions were used in the analysis. In (Table 3) the Eigenvalue shows how much of the variance in the independent, which is the diabetic response, is accounted for by each of the functions. It can be clearly seen that the variant accounts for 100%. We are only using two groups here, namely 'Diabetic' and 'Healthy', so only one function is displayed. The canonical correlation is the multiple correlations between the predictors and the discriminant function. In (Table 3) a canonical correlation of 0.876 suggests the model explains the variation in the grouping variable, i.e. whether a respondent is Diabetic or not.

4.1. Table of Eigen values

We are only using two groups here, namely 'Diabetic' and 'Healthy', so only one function is displayed. The canonical correlation is the multiple correlations between the predictors and the discriminant function. In (Table 3) a canonical correlation of .764suggests the model explains the variation in the grouping variable, i.e. whether a respondent is Diabetic or not.

Test of Function(s)	Wilks' Lambda	Chi-square	Df	Sig.
1	.546	486.723	4	.000

Table 4: Wilks' Lambda Test

4.2. Wilks' Lambda Test

Wilks' lambda indicates the significance of the discriminant function. (Table 4) indicates a highly significant function (p <0.05) and provides the proportion of total variability not explained. That is, it is the converse of the squared canonical correlation. Here the multivariate test Wilks' Lambda acts just like in MANOVA. Since $p < .05$, we can say that the model is a good fit for the data. This multivariate test is a goodness of fit statistic, just like the *F*-test in the case of regression. The associated significance value indicates whether the difference is significant. Here, the lambda of the function is 0.546 and has the value (Sig. 0.00). However, smaller the Wilks's Lambda value, the more important the independent variable to the discriminant function.

4.3. Analysis 2

4.3.1. Box's Test of Equality of Covariance Matrices

Diabetic	Rank	Log Determinant
0	4	15.345
1	4	14.456
Pooled within-groups	4	15.098

Table 5: Log Determinants

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

	Box's M	69.876
F	Approx.	7.345
	df1	10
	df2	109876.237
	Sig.	.000

Table 6: Box's M Test Results

Box's M test tests the assumption of homogeneity of covariance matrices. This test is very sensitive to meeting also the assumption of multivariate normality. For our data, the test is significant so we conclude the groups do differ in their covariance matrices, violating an assumption in DA. However, Discriminant Function Analysis is robust even when the homogeneity of variances is not met, provided the data do not contain important outliers.

Function	Eigen value	% of Variance	Cumulative %	Canonical Correlation
1	1.432[a]	100.0	100.0	.864

Table 7: Eigen Values

Canonical discriminant functions were used in the analysis and the results were shown in (Table 7). The Eigenvalues shows how much of the variance in the independent, diabetic responses, is accounted for by each of the functions. The variate accounts for 100% of Variance is an indication of how much each Discriminant Function Contributes to the analysis. The canonical relation is correlation between discriminant variables and the levels of the dependent variable. A high correlation indicates a function that discriminates well.

Test of Function(s)	Wilks' Lambda	Chi-square	Df	Sig.
1	.576	348.659	4	.000

Table 8: Wilks' Lambda

It is apparent from (Table 8) that p< .05, therefore, we can say that the model is fit for the data good. It is also obtained that the Wilks's lambda function is 0.576 with has the value (Sig. 0.00). That is the proportion of the total variance in the discriminant scores not explained by differences among groups. A lambda of 1.00 occurs when observed group means are equal (all the variance is explained by factors other than difference between those means). The associated significance value indicates whether the difference is significant.

5. Conclusion

This paper assesses the study that was carried out based on the broad objective of using the Box M and Wilks Lambda tests for tackling diabetic patients among adults in Yobe state, specialist hospital Damaturu. The Diabetic patient's responses are categorized into two groups which include: healthy ("0") and diabetic ("1"). Various variables (predictors) i.e. age, sex, weight, blood sugar and urine sugar are used based on discriminant analysis. By using Fisher's method of discriminant analysis, the classification table results were obtained while assumption of homogeneity of covariance matrices is tested with Box's M test. This test is very sensitive to meeting also the assumption of multivariate normality. For our data, the test is significant so we conclude the groups do differ in their covariance matrices, violating an assumption in discriminant analysis. Additionally, Wilks' Lambda of the function is 0.546 and has the value (Sig. 0.00), thus we can say that the model is a good fit for the data since $p < .05$. This multivariate test is a goodness of fit statistic, just like the *F*-test in the case of regression.

In conclusion, this study shows that discriminant analysis provides results that are both more interpretable and statistically sound, in addition to being a statistically correct procedure for prediction purpose than traditional measures and Discriminant analysis classification is found to be vital and useful for diagnoses in any hospital.

6. Recommendation

In the light of the findings, it is recommended that doctors and clinics should adopt the use of the models built by this research to detect prevalence of Diabetics among adults so that adequate measures for prevention and control of diabetics can be taken early enough to alert the danger of the full manifestation of the disease. It is also recommended that the Discriminant model built should be used for tackling diabetes mellitus cases in Yobe state, specialist hospital Damaturu.

7. References

i. Pickup, J. C. (2004). Inflammation and activated innate immunity in the pathogenesis of type 2 diabetes. *Diabetes care*, *27*(3), 813-823.

ii. Tabachnick, B. G., & Fidell, L. S. (1996). Using multivariate statistics. Northridge. *Cal.: Harper Collins*.

iii. Wells, H.G and Lawrence (2002). Diabetes U.K retrived from www.diabetes.org.uk/diabetes

iv. Maurya, V.N., Yusuf, M. U.,Singh,V. V. and Modu B. (2015). Application of Discriminant Analysis on Broncho-Pulmonary Dysplasia among Infants: A Case Study of UMTH and UDUS Hospitals in Maidguri, Nigeria. 4(2-1) 44-51

v. Usman, A. (2012). Statistical Methods for Biometric & Medical Research. Millennium Printing ad Publishing Company Limited, Kaduna, Nigeria: 486-495.

vi. Bhering, Carlos A., Christieny C. Mochdece, Maria EL Moreira, José R. Rocco, and Guilherme M. Sant'Anna. "Bronchopulmonary dysplasia prediction modelfor 7-day-old infants." *Jornal de pediatria* 83, no. 2 (2007): 163-170.

vii. Boulé, N. G., Haddad, E., Kenny, G. P., Wells, G. A., & Sigal, R. J. (2001). Effects of exercise on glycemic control and body mass in type 2 diabetes mellitus: a meta-analysis of controlled clinical trials. *Jama*, *286*(10), 1218–1227

viii. Cook, P. (2010), Discriminant Function Analysis in SPSS. University of Colorado Denver, Center for Nursing Research

ix. Sun, L., Ji, S., & Ye, J. (2016). *Multi-label dimensionality reduction*. Chapman and Hall/CRC.