

Classification Analysis of Gender Among Diabetic Patients in Nigeria Hospital

Ojo Timothy Ayodele^{1, *}, Alabi Olatayo Olusegun², Oyegoke Adiat Odunayo³,
Ayinde Liasu Adekunle³, Ogunwole Bolawa Adijat³

¹Lautech Teaching Hospital, Osogbo, Nigeria

²Department of Statistics, Federal University of Technology, Akure, Nigeria

³Department of statistics, Osun State Polytechnic, Iree, Nigeria

Email address:

timothyjo04@gmail.com (O. T. Ayodele)

*Corresponding author

To cite this article:

Ojo Timothy Ayodele, Alabi Olatayo Olusegun, Oyegoke Adiat Odunayo, Ayinde Liasu Adekunle, Ogunwole Bolawa Adijat. Classification Analysis of Gender Among Diabetic Patients in Nigeria Hospital. *International Journal on Data Science and Technology*. Vol. 6, No. 2, 2020, pp. 53-55. doi: 10.11648/j.ijdst.20200602.11

Received: April 25, 2020; Accepted: May 21, 2020; Published: August 10, 2020

Abstract: In medical research there is few record on scientific method of discriminating and classifying gender statistically into groups of study. The purpose of this study is to use discriminant analysis and classification analysis to classify diabetic patient into groups of gender; to estimate the proportion of observations in each of the prior group; and to estimate the probability of correct classification and misclassification respectively. To this effect, a sample of 152 cases (diabetic patients) was observed with the following measurements: Age (x_1), Urea (x_2), temperature (x_3), Fasting blood sugar (x_4), Body mass index (x_5), and marital status (x_6). The gender was classified into male and female. We observed that the Discriminant Function $Z=0.036x_1+0.008 x_2-0.897 x_3-0.021 x_4-0.017 x_5-2.872 x_6$. Also 64.5% of the original grouped cases were correctly classified. The percentage of misclassification is 34.5%. Conclusively the measure of the predictive ability which is the percentage of correct classification shows that discriminant analysis can be used to predict diabetic patients into two classes of gender and can also be used to predict group membership of any subject matter.

Keywords: Dicriminant, Classification, Multivariate, Misclassification, Diabetic Patient

1. Introduction

Diabetes mellitus is recognized as being a syndrome, a collection of disorders that have hyperglycemia and glucose intolerance as their hallmark, due either to insulin deficiency or to the impaired effectiveness of insulin's action, or to a combination of these. In order to understand diabetes it is necessary to understand the normal physiological process occurring during and after a meal. Food passes through the digestive system, where nutrients, including proteins, fat and carbohydrates are absorbed into the bloodstream. The presence of sugar, a carbohydrate, signals to the endocrine pancreas to secrete the hormone insulin. Insulin causes the uptake and storage of sugar by almost all tissue types in the body, especially the liver, musculature and fat tissues [1].

Diabetes is a lifelong condition that causes a person's blood

sugar level to become too high. Persons with diabetes have too much sugar in their blood. There is no cure for diabetes. People with diabetes need to manage their disease to stay healthy. it is a metabolic disorder of chronic hyperglycemia characterized by disturbances to carbohydrate, protein, and fat metabolism resulting from absolute or relative insulin deficiency with dysfunction in organ systems [14].

Populations previously unaffected or minimally affected by DM are now reporting soaring prevalence figures, which poses a real challenge to health financing by governments and nongovernmental organizations. The latest prevalence figure published by the International Diabetes Federation (IDF) is 425 million persons living with DM worldwide, with nearly 50% of these undiagnosed. The developing economies of Africa and Asia contribute a significant fraction of this figure. There is also a rising burden from the complications of DM alongside the ever-increasing prevalence of the

disease. We now see high rates of DM-related amputations, cerebra vascular disease, heart-related problems and in Nigeria, the current prevalence of DM among adults aged 20–69 years is reported to be 1.7% [4].

Linear Discriminant Analysis easily handles the case where the within-class frequencies are unequal and their performances has been examined on randomly generated test data. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability. [2] The use of discriminant analysis is related to group memberships. The use of this technique is for classifying individuals or objects into one the alternative groups. This is done on the basis of a set the existing predictor variables in research. The dependent variable that is there is discriminant analysis has to be categorical in nature and put on nominal scale. On the other hand the independent variable which is also called the predictor variable has interval or ratio scale tendency [3, 5-7].

2. Methodology

Multivariate Analysis consists of a collection of methods that can be used when several measurements are made on each individual or object in one or more samples. Measurements are referred to as variables and individuals or objects as units (research units, sampling units, or experimental units) or observations. In practice, Multivariate data sets are common, although they are not always analyzed as such. But the exclusive of univariate procedures with such data is no longer excusable, given the availability of multivariate techniques and inexpensive computing power to carry them out. Research in behavioural sciences mostly involves developing prediction and classification models. Discriminant function Analysis also called Discriminant Analysis is used to classify cases into the values of a categorical dependent, usually a dichotomy.

Discriminant analysis is a statistical technique that is used to classify the dependent variable between two or more categories. Discriminant Analysis also has a regression technique which is used for predicting the value of the dependent categorical variable. Discriminant analysis may be applied in a number of settings: in the armed forces, it is used in assigning new personnel to training programs. In the industry, it is used in assigning new employees to a particular job category. In health, it is used in classifying a patient into one of several diagnostic categories. In education, it is found useful as aids in both educational and vocational counseling. [5-7]

This concern for the classification ability of the linear discriminant function has obscured and even confused the fact that two very distinct purposes and procedures for conducting discriminant analysis exist. The first procedure, discriminant predictive analysis is used to optimize the predictive functions, that is the objective is to develop an equation that maximally discriminate the groups using P independent variables. The second procedure, discriminant classification analysis uses the predictive functions derived in the first procedure to either classify fresh sets of data of

known group membership, there by validating the predictive function; or if the function has previously been validated [8, 9, 12].

The goal of discriminant analysis include identifying the relative contribution of the P variables to separation of the groups and finding the optimal plane on which the points can be projected to best illustrate the configuration of the groups (see [2, 12]).

3. Material and Method

3.1. Model

$$D_i = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \dots + \beta_k X_k$$

k – number of variables taken into consideration during classification e.g.

discriminant variables

X_i – discriminant variables (explanatory variables), $i=1, 2, \dots, k$

β_i – coefficients of discriminant function, $i=1, 2, \dots, k$

w – maximum number of discriminant functions;
 $w = \min\{k, g - 1\}$

3.2. Classification Functions of Ronald Aylmer Fisher

$$F_1 = f_{01} + f_{11}X_1 + f_{21}X_2 + \dots + f_{j1}X_j$$

$$F_2 = f_{02} + f_{12}X_1 + f_{22}X_2 + \dots + f_{j2}X_j$$

⋮

$$F_G = f_{0G} + f_{1G}X_1 + f_{2G}X_2 + \dots + f_{jG}X_j$$

Rule: The observation unit belongs to the group for which the value F_g is maximal.

The data used for this research work comprises of 152 Diabetes patients with the following measurements on each (Age, fasting blood sugar, temperature, body mass index, marital and Urea).

4. Analysis

Table 1. Summary of Variables under Consideration for the Two Groups.

Sex		Mean	Standard error	N
Male	Fasting blood sugar	14.5357	9.91026	70
	Urea	11.9329	13.11921	
	Temperature	37.2314	.89435	
	Body mass index	22.9632	4.79146	
	Age	55.0000	17.44972	
	Marital	1.8429	.36656	
Female	Fasting blood sugar	17.0293	11.72939	82
	Urea	11.1622	8.89573	
	Temperature	36.7780	.82985	
	Body mass index	23.0270	4.77815	
	Age	56.8902	17.18671	
	Marital	1.9634	.18890	

Table 2. Classification Function Coefficients.

	Sex	
	Male	Female
Fasting blood sugar	-.367	-.338
Urea	-.036	-.030
Temperature	50.676	49.955
Body mass index	.112	.095
Age	.075	.062
Marital	3.981	6.291
(Constant)	-948.211	-925.300

$$Z = 28.465 + 0.036_{\text{FBS}} + 0.008_{\text{Urea}} - 0.897_{\text{Temp}} - 0.021_{\text{BMI}} - 0.017_{\text{Age}} - 2.872_{\text{Marital Status}}$$

Table 3. Predicted Group Membership.

Grouping	Male	Female	Total
Male	29	41	70
Female	13	69	82
% Male	41.4	58.6	100.0
%Female	15.9	84.1	100.0

5. Result

The Probability of misclassification $(13+41)/152=0.355$.

Percentage of Misclassification 35.5%.

The probability of correct classification $(29+69)/152=0.6447$.

The percentage of correct classification 64.5%.

This implies that 64.5% of original grouped cases correctly classified.

6. Conclusion

Discriminant score was gotten and it was used in classifying the variables into two groups. 64.5% of the original grouped cases were correctly classified while 35.5% were misclassified.

The probability of misclassification is 0.355. That is to say that the level of accuracy is high and significant.

In Summary, the measure of the predictive ability which is the percentage of correct classification shows that discriminant analysis can be used to predict diabetes into two classes of gender and can also be used to predict group membership of any subject matter.

References

- [1] Adimora, G. N, Nigerian Journal of Clinical Practices. Vol 7.2004, pg 33- 36 Official Publication of the Medical and Dental Consultants Association of Nigeria.
- [2] Deswal B. S., Singh J. V., Kumar D.- A Study of Risk Factors for Low Birth Weight, Indian J. Community Med. 24, 2008: 127-131.
- [3] Philip, J. Disala, *Clinical Gynaecologicon. cology*. (4th Edition). Mosby year book Inc. 1995.
- [4] Geoffrey. V. P Chamberline, *Gynaecology by Ten Teachers*. (16th Edition). ELBS. 1988.
- [5] Daniel B. Rowe, *Multivariate Bayesian Statistics*. Chapman and Hall/CRC, 2003.
- [6] Anderson, T. W. *An Introduction to Multivariate Statistical Analysis*. New York: Wiley. 1984.
- [7] Barnett, V. (ed.), *Interpreting Multivariate Data*. Wiley. 1981.
- [8] David, W. Stockburger, *Multivariate Statistics: Concepts, Models and Applications*. 1998.
- [9] Nwobi and Nduka. *Statistical Notes and Tables for Research*, Second Edition. Alphabet Nigeria Publishers. 2003.
- [10] Nwachukwu V. O, *Principles of Statistical Inference*. Second Edition. Zelon Enterprises. 2006.
- [11] Everitt, B. S & Dunn, G, *Applied Multivariate Statistical Analysis*. (2nd Edition). Arnold. 2001.
- [12] Alvin B. Rowe, "Methods of Multivariate Analysis, second edition" Wiley –Interscience, 2003.
- [13] Adejumo, A. O and Onyenekwe, C. E (2014) A Study of Classification and Discriminant Analysis of Infants at Birth in Nigeria. *IOSR Journal of Mathematics (IOSR-JM) e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 9, Issue 4 (Jan. 2014), PP 22-26 www.iosrjournals*.
- [14] Andrew E. Uloko, Baba M. Musa, Mansur A. Ramalan, Ibrahim D. Gezawa, Fabian H. Puepet, Ayekame T. Uloko, Musa M. Borodo, and Kabiru B. Sad (2018) Prevalence and Risk Factors for Diabetes Mellitus in Nigeria: A Systematic Review and Meta-Analysis *BMJ Open Diabetes Res Care.*; 2 (4) 127-130.