Journal of Engineering and Applied Sciences 14 (13): 4485-4490, 2019
ISSN: 1816-949X

The Difference Between Hemolytic Anemia and the Deficiency of G6PD Among Children Using a Discriminant Analysis

Rasha Adil Saeed, Inaam Aboud Hussain and Suhail Najm Abbood
Department of Statistic, College of Administration and Economics, University of Baghdad, Baghdad, Iraq

Abstract: There are numerous and different reason for the causes, symptoms, consequences and treatments of anemia diseases. The correct diagnosis of each type of anemia is therefore, important. Those who are most vulnerable are the newborns and under-five children. Therefore, the main objective of this study is to identify the variables that affect the diagnosis of the two anemia diseases among children: hemolytic anemia and G6PD deficiency. This aim is intended to be achieved by using the discriminant analysis to differentiate between the G6PD deficiency and hemolytic anemia among the selected 150 children infected with the diseases. The result of the study showed that three factors such as jaundice, fever and age contribute to the significant differences between the group of children infected with hemolytic anemia and the group of children infected with G6PD deficiency. There is high capability of differentiating between these two groups with the discriminant function obtained from these factors. Therefore, this study recommends the use of discriminant function to help in diagnosing the cases of these diseases as carried out by the medical authorities concerned with anemia in children.

Key words: Discriminant analysis, anemia diseases, hemolytic anemia, G6PD deficiency, medical authorities, symptoms

INTRODUCTION

Generally, anemia is known to be a common disease among children. There are many cases of Anemia such as the G6PD deficiency and hemolytic anemia. In anemia, diagnosis is carried out based on the age of the patient, the gender of the patient and the symptoms of the disease

through clinical tests. Then in order to determine the appropriate treatment methods for each type of anemia, laboratory tests are conducted. Thus, to identify the factors affecting the diagnosis of these two types of anemia (follicular anemia and hemolytic anemia) among childrenis the aim of the study. The study explored differential analysis through simple random technique to select 150 children. A central education in the housing area of Baghdad was chosen the population with hemolytic anemia or follicular anemia in 2017. Due to the strong nature of the analysis of discriminatory statistical methods it can be used to analyze differences between two or more variables with similar qualities. This will allow the factors to be classified into the right groups with the least possible error using one or more discriminatory functions in accordance with the available information on the population of the study.

Literature review: From the concept of analysis, differential analysis is known as a statistical method for analyzing multivariate data which deals with two categories of techniques. First and foremost is the classification of the observations into their groups where they belong which is simply referred to as the Predictive Discriminant Analysis (PDA). It entails the prediction of observation return to their studied group. The second one is called the Descriptive Discriminant Analysis (DDA) (Erimafa *et al*., 2009). The following summarized the objectives of the analysis:

Maximizing the degree of excellence (Lambda) and minimizing the degree of overlapping between the studied groups to the lowest possible degree which represents the ratio of the sum of the squares of deviations between the groups to the sum of squares of deviations within the groups. Designing the linear combinations and discriminant function of the independent variables in the best way to differentiate between the classifications of the dependent variables.

Using the strength of the distinction between the groups selected in this study to determine the relative importance of the independent variables in the discriminatory function. Using the values of independent

Corresponding Author: Rasha Adil Saeed, Department of Statistic, College of Administration and Economics, University of Baghdad, Baghdad, Iraq

variables to make division of cases between the categories of the dependent variable. Percentage evaluation of the division accuracy. The following are the conditions and assumptions to be considered before applying discriminatory analysis:

- Although, there may be a degree of overlapping, the two or more studied groups should be statistically distinct and identifiable
- There should be two views at least in each group (Savic *et al*., 2008)
- The total number of observations should be greater than number of independent variables (Discriminating Variables) in the discriminant function (Savic *et al*., 2008)
- Absence of correlation problem between the variables used in the discriminatory function which leads to multiple linear duplications
- Unless special formulas are used, the variance matrices for groups are almost equal (Savic *et al*., 2008)
- In each of the studied group, a multiple natural distribution is followed by the differential variab les (Savic *et al*., 2008)
- From the group size, there should not be significant difference which leads to erroneous classification process where the larger group tends to have the views (Savic *et al*., 2008)

Practically it has been observed that the method of discriminatory analysis is reliable enough even if the conditions and above are not satisfied (Savic *et al*., 2008). Looking at the case of two groups is the distinguishing function of the analysis. The discriminant function is in accordance with the factors of the sample randomly chosen and placed in two different groups. The return to a group can be determined and the uniqueness tested in this function (Saleh, 2008).

Given that the sample area is W, it can then be divided into two parts: R returns to the first group while WR belongs to the second group. Any group of these two groups can receive the boundary between the two groups, i.e., $X_1, X_2, \ldots, X_p$. the aim of the equation is find the linear function. According to Saleh (2008), the mathematical formula will be as follows:

$$y = b_1x_1+b_2x_2+, \ldots, b_px_p \quad (1)$$

where, y stands for linear discriminant function which is the linear combination of independent variables (explanatory). It can be interpreted in the following form of matrices:

$$y = \underline{b}'\underline{x} \quad (2)$$

The best distinction between the two groups is achieved by the process of estimating the parameters ($b_1, b_2, \ldots, b_p$) that makes the function through the square ratio of the difference between the two groups to the common variation of the two groups as high as possible (Saleh, 2008) the greatest possible:

$$Q = \frac{[\bar{y}_1 - \bar{y}_2]^2}{\sum_{i=1}^{2}\sum_{j=1}^{ni}(y_{ij}-\bar{y}_i)^2} = \frac{\left[b'(\bar{x}_1-\bar{x}_2)\right]^2}{b'sb} \quad (3)$$

Where:

$$\bar{y}_i = b_1\bar{x}_{i1}+b_2\bar{x}_{i2}+,\cdots,b_p\bar{x}_{ip} \quad (3)$$

And:

$$\bar{x}_i = (\bar{x}_{i1}+\bar{x}_{i2},\ldots,\bar{x}_{ip}),\ i = 1,2,\ldots \quad (4)$$

Which represents the sample size E

$$s = \frac{1}{n_1+n_2-2}(x_1'x_1 - x_2'x_2) \quad (5)$$

The maximization of the ratio (Q) is achieved by partial derivation and maximization by zero. Thus, the following is obtained:

$$b = s^{-1}(\bar{x}_1-\bar{x}_2)' \quad (6)$$

After the extraction of the values ($b_1, b_2, \ldots, b_p$) and obtaining the special function, classification of the new vocabulary can be done into one of the two groups with this function by assigning a cut-off point or critical score that makes the probability of misclassification less of what it can be (AL-Hamdani, 2014).

The size of the groups under this study determines the process of assigning the point of separation. If the size of the second group is similar to the size of the first group then the following equation holds (Ramayah *et al*., 2010):

$$z_c = \frac{n_1\bar{y}_2+n_2\bar{y}_1}{n_1+n_2} \quad (7)$$

Where:

Z_c = Cut-off point
$\bar{y}_1$ = Centroid for group 1
$\bar{y}_2$ = Centroid for group 2
n_1 = First group size (number of views of the first group)
n_2 = Second group size (number of views of the second group)

When the size of the second group is not similar to the size of the first group, then the following equation holds (Ramayah *et al*., 2010):

$$z_c = \frac{\overline{y}_1+\overline{y}_2}{2} \tag{8}$$

$\hat{y}>z_c$ = The view is classified into the first group
$\hat{y}<z_c$ = The view is classified into the second group
$\hat{y} = z_c$ = Then the observation is randomly assigned to the first or second group

Whereas:

$$\hat{y} = (\overline{x}_1-\overline{x}_2)'s^{-1}x \tag{9}$$

The rejection of the null hypothesis that assumes the mean of the group determines the significance of the discriminant function. Hence:

$$H_0: \mu_1 = \mu_2 \text{ VS } H_1: \mu_1 \neq \mu_2 \tag{10}$$

The F test and its mathematical formula are one of the several measures to test this hypothesis as follows (Saleh, 2008):

$$F_{cal} = \frac{n_1+n_2-p-1}{(n_1+n_2-2)p} \times T^2 \tag{11}$$

The Hypothesis H_1 is accepted while H_0 is rejected with a significant level of $F\alpha$ if:

$$F_{cal} > F_{\alpha(p, n_1, n_2-p-1)} \tag{12}$$

This shows that there is no equality in the mean of the two groups. The scale of the Wilk's Lambda and its mathematical formula can be used as follows (Akkar, 2008):

$$\wedge = \frac{|W|}{|W+B|} = \frac{|W|}{|T|} \tag{13}$$

Where:

W = The matrix and common variance in the total values
B = The matrix of variance and contrast between the aggregates
T = The total range of the variance and contrast of the aggregates

If the value of Wilk's Lambda is close to or equal to one it ranges between zero and one. This shows that there is no discrimination as the value of the groups is equal; The strength of discrimination is indicated if the value is close to zero (Akkar, 2008).

Probability of classification error: To measure the performance of a discriminatory function, the probability of a rating error is an ideal criterion. Thus, the less the efficiency of the discriminant function, the lower the probability of the misclassification. Re-substitution method is one of the several methods for estimating the probability of error of classification where the probability of error is estimated by applying the following mathematical formula (Al-Hamdani, 2014):

$$\hat{P}_{ki} = \frac{n_{ki}}{n_k} \tag{14}$$

Where:

$\hat{P}_{ki}$ = The probability of classifying the individual into group i which originally belongs to group k
n_{ki} = The number of views in group (k) that is categorized as belonging to group (i)
n_k = The number of group views (k)
K = The number of groups studied (k = 1, 2)

MATERIALS AND METHODS

Hemolytic anaemia: Is a disorder in the blood system as a result of red blood cell with higher breakdown than the rate of production in the spleen; Although, this is just one of the causes of anaemia. Hemolytic anaemia is caused due to the following factors varying from one pattern to another:

- Internal hemolytic anemia such as sickle cell anemia and thalassemia
- Exogenous hemolytic anemia such as infection and pharmacological interactions of certain pharmacological drugs

The signs of hemolytic anemia are shown based on the causes and severity of the symptoms which differs from one condition to another. In a simple case, no symptom may appear on the patient. Generally, all types of anaemia contain the hemolytic anemia. The symptoms of the diseases are the breakage of red blood cells and the releases of hemoglobin in the blood are: yellowing of the eye and skin, difficulty in breathing, stress, paleness, chest pain, dizziness, heart failure, dark yellow or brown urine, arrhythmias, cold limbs, headache and lack of concentration.

The cause and severity of the condition of the diseases, the development of the condition, the history of the disease and the age of the patient determine the treatment of the hemolytic anemia. In some cases where the cause can be controlled, the acquired hemolytic anemia can be fully recovered. The disease persists with the patient throughout his life in the case of hereditary anemia. This, therefore, necessitates continuous medical care. Generally, the aim of the treatment is to:

- Stop or reduce the breakdown of the red blood cells
- Raise the appropriate number of the red blood cells to some extent
- If possible, to treat the cause of the condition

The treatments for the disease include:

- Blood transfusion
- The use of drugs such as cortisone and folic acid
- Splenectomises

G6PD deficiency: Also known as voile, folic acid and bean anemia is a hereditary anemia as a result of decomposition of red blood cells due to genetic disorder that causes a deficiency of the enzyme G6PD. Therefore, the disease is medically referred to as the G6PD deficiency or Favism. To strengthen the defence of this biochemical means on the surface of the cell within a series of interactions against oxidizing substances whose attack may break down cell walls and then decompose them is the function of this enzyme.

The disease is genetically and sexually transmitted. It affects the male child more through their mother. The disease may appear in some cases among the female children. The male child affected with this disease can transmit it to their daughters but not to other children.

When the patient is infected with the disease, the symptoms of folic anemia begins to show. After the infection of the virus, the lentils or any type of legumes especially hepatitis appears. Then, the symptoms may disappear if some drugs are taken by the person with the disease or without taking any kind of oxidized substances like legumes. Similarly, some of the symptoms of jaundice are discoloration of skin, mucous membranes (conjunctivitis), nausea, palpitations, vomiting, dark urine, back pain, abdominal pain and light fever. These symptoms are as a result of severe destruction of erythrocytes through increased production of bilirubin. However, the new-borns may be vulnerable to this substance as they can be very harmful as they cause serious problems by getting deposited inside the brain. The folicacid fracture has no cure and its prevention from passing on from one generation to another generation is impossible. To treat the disease, the patient has no avoid blood fracture by refraining from oxidized substance to minimize the viral and bacterial diseases.

Variables in the composition of the linear discriminant function: This study relies on set of variables collected from each of the items of the research sample in order to form a linear discriminative function: these variables are:

The dependent variable (y): This is the diagnosis hemolytic anemia disease of a child and is identified by the number of the child's diagnosis of follicular anemia.

The independent variables (X_1, X_2, ..., X_p): In order to find the linear discrimination function, the symptoms of hemolytic anemia, follicular anemia and other common characteristics were used. The classifications are made into two groups: A group of children with hemolytic and another group with follicle anemia. These variables are: These variables and their representations are: X_1: Sex, X_2: Age, X_3: Nausea, X_4: Fever, X_5: Jaundice, X_6: Paleness of the Face, X_7: Darkness. The analysis technique employed for the data and to obtain results is SPSS Version 21.

RESULTS AND DISCUSSION

The results of the discriminatory analysis: The result from the Table 1 and 2 shows that there is no correlation between the independent variables. The method of stepwise was employed to evaluate the independent variables with the moral effect in the differential analysis

Table 1: The computational circles and standard deviations of the independent variables of the two groups in separation and in joint

Group statistics

			Valid N (listwise)	
Diagnosis	Mean	SD	Unweighted	Weighted
Hemolytic anemia				
Gender	1.431818	0.4981680	88	88.000
Age	3.557765	2.3149949	88	88.000
Nausea	1.034091	0.1825026	88	88.000
Fever	0.522727	0.5023456	88	88.000
Jaundice	0.636364	0.4838024	88	88.000
Paleness of the face	0.522727	0.5023456	88	88.000
Dark urine color	0.545455	0.5007831	88	88.000
Gender	1.451613	0.5017157	62	62.000
Age	8.825806	3.5671054	62	62.000
Nausea	1.096774	0.2980636	62	62.000
Fever	0.467742	0.5030315	62	62.000
Jaundice	0.193548	0.3983042	62	62.000
Paleness of the face	0.467742	0.5030315	62	62.000
Dark urine color	0.532258	0.5030315	62	62.000
Total				
Gender	1.440000	0.4980499	150	150.000
Age	5.735222	3.8875739	150	150.000
Nausea	1.060000	0.2382824	150	150.000
Fever	0.500000	0.5016750	150	150.000
Jaundice	0.453333	0.4994852	150	150.000
Paleness of the face	0.500000	0.5016750	150	150.000
Dark urine color	0.540000	0.5000671	150	150.000

Table 2: The correlation matrix within the independent variables

Pooled within-groups matrices

Variables	Gender	Age	Nausea	Fever	Jaundice	Paleness of the face	Dark urine color
Correlation							
Gender	1.000	-0.031	-0.057	-0.107	0.001	0.001	0.064
Age	-0.031	1.000	0.001	0.049	-0.099	-0.099	-0.023
Nausea	-0.057	0.001	1.000	-0.078	-0.021	-0.021	-0.104
Fever	-0.107	0.049	-0.078	1.000	0.010	0.010	0.039
Jaundice	0.072	0.385	-0.004	-0.205	-0.056	-0.056	-0.028
Paleness of the face	0.001	-0.099	-0.021	0.010	1.000	1.000	-0.389
Dark urine color	0.064	-0.023	-0.104	0.039	-0.389	-0.389	1.000

Table 3: Significant independent variables

Variables entered/removed[a-d]

		Min. F				
Steps	Entered	Statistic	df1	df2	Sig.	Between groups
1	Age	120.246	1	148.000	7.494E-21	Hemolytic anemia and the deficiency of G6PD
2	Jaundice	119.786	2	147.000	1.369E-31	Hemolytic anemia and the deficiency of G6PD
3	Fever	84.569	3	146.000	9.118E-32	Hemolytic anemia and the deficiency of G6PD

At each step, the variable that maximizes the smallest F ratio between pairs of groups is entered; [a]Maximum number of steps is 14; [b]Minimum partial F to enter is 3.84; [c]Maximum partial F to remove is 2.71 and [d]F level, tolerance or VIN insufficient for further computation

Table 4: The detailed of Wilk's Lambda

Wilk's Lambda

						Exact F			
Steps	Number of variables	Lambda	df1	df2	df3	Statistic	df1	df2	Sig.
1	1	0.552	1	1	148	120.246	1	148.000	0.000
2	2	0.380	2	1	148	119.786	2	147.000	0.000
3	3	0.365	3	1	148	84.569	3	146.000	0.000

while the method of inserting or removing the variables in the gradual differential analysis is controlled by the smallest F ratio. In this analysis, the minimum value of F for each variable must be at least 3.84. The required upper limit of the partial F value is 2.71 to bring out any variable of the analysis. Table 3 reveals that X_2, age is the first variable included in the analysis due to its largest value of F (120.246). The second variable with the largest value of F in this analysis is the X_5 Jaundice with 119.786. The third variable, X_4 fever has F value of 84.569.

The result obtained as presented according to Table 4 indicates that Lambda's value (0.552) is required first before the introduction of the first variable X_2, age in the analysis. The value of Lambda is 0.380 when the second variable X_5, Jaundice was introduced in the analysis. Then, the value of Lambda reduced to 0.365 later in the analysis. This shows that the value of Lambda decreases as effective variable is added to the analysis. The greater the difference between the two groups, the lower the value of the Lambda. The tabular value of F is lesser than the value of F at each step. This indicates that the level of the statistical significance at each step is zero. Table 5 shows the inherent root value of the discriminant function to be 1.738. This is an indication of high ability of the discriminant function to differentiate between the two groups. The canonical correlation is valued as 0.797 showing the linear discrimination quality.

The value of the square test X_2 is 147.544 as shown in Table 6 and 7 which is larger than the tabulated value.

Table 5: The underlying roots

Eigenvalues

Function	Eigenvalue	Variance (%)	Cumulative (%)	Canonical correlation
1	1.738[a]	100.0	100.0	0.797

[a]First 1 canonical discriminant functions were used in the analysis

Table 6: The Wilk's Lambda

Wilk's Lambda

Test of function (s)	Wilk's lambda	Chi-square	df	Sig.
1	0.365	147.544	3	0.000

Table 7: Non-standard discriminatory transactions

Canonical discriminant function coefficients

	Function
Variables	1
Age	-0.348
Fever	0.512
Jaundice	1.800
(Constant)	0.925

Unstandardized coefficients

Table 8: Result classification

Classification results[a]

VariablesDiagnosis	Predicted group membership: Hemolytic anemia	Foley anemia	Total
Original			
Count			
Hemolytic anemia	83	5	88
Foley anemia	2	60	62
Percetage			
Hemolytic anemia	94.3	5.7	100.0
Foley anemia	3.2	96.8	100.0

[a]95.3% of original grouped cases correctly classified

Due to the three variables, age, jaundice and fever it is possible to conclude that there are statistically significant differences between the two groups. The vale indicates that the moral level of significance is equal to zero. Therefore, the linear discriminatory function is mathematically depicted as follows:

$$Y = 0.925-0.348X_2+0.512X_4+1.800X_5$$

Table 8 above presents the accurate final results of the classification. There are 83 cases of children with hemolytic anemia in the first group with 94.3% correct classification while wrong classification was done to 5 cases with 5.7% for the first group. There are 60 cases of children with follicular anemia in the second group with 96.8% correct classification while wrong classification was done to 2 cases with 3.2% for the first group. It shows that 95.3% have been correctly categorized.

CONCLUSION

From the seven variables, only three have the greatest effect on the diagnosis of hemolytic anemia and folic anemia in children. They are: jaundice, fever and age in order of importance. Due to the three variables (jaundice, fever and age), there are significant differences between the group of children with follicular anemia and the group of children with hemolytic anemia. The discriminatory function from these variable relationships is suitable to make distinction between the groups.

For the group of children with follicular anemia, the correct classification is 96.8% while the rate of correct classification for the group of children with acute hemolytic anemia is 94.3%. The total rate of correct classification for the entire selected sample is 95.3%.

RECOMMENDATIONS

The discriminant function is recommended to be used to assist the medical authorities dealing with children diagnosed with hemolytic anemia and folic anemia. The method of discriminatory analysis and its reliability in making distinction between all types of anemia in children is generally recommended.

REFERENCES

Akkar, A.A.A., 2008. Classification of certain types of dates using distinctive analysis. J. Administration Econ., 68: 96-109.

Al-Hamdani, B.R., 2014. Distinguishing the medical staff according to their knowledge of the International Classification (ICD-10) using the distinctive function. MBA Thesis, University of Baghdad, Baghdad, Iraq.

Erimafa, J.T., A. Iduseri and I.W. Edokpa, 2009. Application of discriminant analysis to predict the class of degree for graduating students in a university system. Int. J. Phys. Sci., 4: 16-21.

Ramayah, T., N.H. Ahmad, H.A. Halim, S.R.M. Zainal and M.C. Lo, 2010. Discriminant analysis: An illustrated example. Afr. J. Bus. Manage., 4: 1654-1667.

Saleh, A.H., 2008. Using distinct analysis to diagnose some ophthalmology. J. Administration Econ., 67: 264-287.

Savic, M., D. Brcanov and S. Dakic, 2008. Discriminant analysis-applications and software support. Manage. Inf. Syst., 3: 29-33.