Open Access Scientific Reports

Research Article Open Access

Keywords: GDM; Odds ratio; Logistic regression; Dichotomous; Newton-raphson

Introduction

e constant evolution of medicine over the last two decades has meant that statistics has had to develop methods to solve the new problems that have appeared, and has come to play a central part in methods of diagnosis of diseases [1]. A diagnostic method consists of the application of a test with a group of patients, in order to obtain a provisional diagnosis regarding the presence or the absence of a particular disease [2]. In this work, logistic regression has been proposed for the purpose of estimating the e ects of various predictors on some binary outcome of interest. Here, logistic regression regresses a dichotomous dependent variable on a set of independent variables, as a way of knowing the e ects of these independent variables [3,4].

We, therefore here, propose to develop a matrix approach for solving a system of nonlinear equations, with P+1 unknown parameters. ese methods will be applied in estimating the e ects of risk factors on the occurrence of gestational diabetic mellitus (GDM) [5-7]. e proposed method will be illustrated using data on gestational diabetic mellitus (GDM), and have been shown to compare favorably with other existing methods in terms of e ciency.

e Proposed Method

e fundamental model for any multiple regression analysis assumes that the outcome variable is a linear combination of a set of predictors, and this is represented as:

$$L(\beta|Y) = \prod_{i=1}^{N} \left(\frac{1}{1+e^{-\beta' x_i}}\right)^{y_i} \left(1 - \frac{1}{1+e^{-\beta' x_i}}\right)^{1-y_i} \quad (9)$$

Recall from Equation 6 that

$$\frac{\pi_i}{1-\pi_i} \quad \frac{\frac{1}{1+e^{-\sum_{k=0}^{P} \beta_k x_{ik}}}}{1 - \frac{1}{1+e^{-\sum_{k=0}^{P} \beta_k x_{ik}}}} \quad \frac{\frac{1}{1+e^{-\sum_{k=0}^{P} \beta_k x_{ik}}}}{1 - \frac{1}{1+e^{-\sum_{k=0}^{P} \beta_k x_{ik}}}} \quad \frac{1}{1+e^{-\sum_{k=0}^{P} \beta_k x_{ik}}} \quad \frac{e^{\sum_{k=0}^{P} \beta_k x_{ik}}}{1+e^{\sum_{k=0}^{P} \beta_k x_{ik}}}$$

$$\frac{1}{1+e^{-\sum_{k=0}^{P} \beta_k x_{ik}}} \quad \frac{e^{\sum_{k=0}^{P} \beta_k x_{ik}}}{1+e^{\sum_{k=0}^{P} \beta_k x_{ik}}} \quad e^{\sum_{k=0}^{P} \beta_k x_{ik}}$$

4 -4.32 Tx0

maximum of the log likelihood function, and vice versa. us, taking

$$\left(X^{T}WX\right)^{-1}X^{T}WZ \text{ where } Z = X\beta^{old} + W^{-1}(Y - P) \quad (25)$$

Where $Z = X\beta^{old} + W^{-1}(Y - P)$ is a vector and W is the diagonal weight vector, with entries $_i(1-\ _i)$.

e last equation is called the weighted least square regression, which nds the best least-squares solution to the equation. e equation is called recursive weighted least squares, because at each step, the weight vector W keeps changing (since β are changing). Now, Equation 25 can be written:

$$\beta^{(1)} = \beta^{(0)} + \left(X^{T}WX\right)^{-1}X^{T}(Y - P) \quad (26)$$

Continue applying Equation 26 until there is essentially no change between the elements of from one iteration to the next. At that point, the maximum likelihood estimates are said to have converged, and Equation 19 will hold the variance-covariance matrix of the estimates. Because the estimation algorithm for the parameter of the logistic regression model is iterative, parameter estimates based on small samples way fail to converge, or converge to local rather than global, stationary points. is informed the application of large sample in this study. is iterative procedure is handled by SAS so ware in this work.

Illustrative Example

In estimating the e ects of risk factors on GDM, 1000 subjects (pregnant women at risk for GDM) were sampled from the ve randomly selected hospitals from January 2010 to December 2011 in Ebonyi State through a retrospective study, out of which 490 (49%) were those less than 28 weeks of their gestational age, and 510 (51%) were those at least 28 weeks of their gestational age. In the total sampled subjects, 530 (53%) were gestational diabetic and 470 (47%) were non-gestational diabetic. Since GDM is a dichotomous variable, it is coded as 0 or 1, and the independent factors considered in this work are Age, Category of pregnant women, Obesity, Income group, Life-style and exercise, F.H of diabetes, Hypertension, and Diet habit are also categorical and coded between 0 and 3. ese are presented in table 1.

Results of Analysis

e results are shown in the following tables: Tables 2 and 3

e table 3 shows that three risk factors: Obesity, F.H and Exercise, were signi cant because for all the above variables p-value was less than 0.05. Since the hospitals where these data were collected are mainly located in the urban areas, it means that by the results

No	Variables	Code number	Coding	Frequency
1	Age	0 if age <30, and 1 for at least 30	0 1	247 753
2	Category of pregnant women	0 if <28 wks of gestational age, and 1 if at least 28 wks	0 1	490 510
3	Obesity	0=non-obsessed and 1 for obsessed	0 1	415 585
4	Income	1=High, 2=Middle, 3=Low	1 2 3	140 390 470
5	Family history	0=Absent,1=Present	0 1	551 449
6	Exercise	0=Sedentary, 1=Light, 2=Moderate	0 1 2	391 472 137
7	Hypertension	0=Non-hypertension,1=Hypertension	0 1	636 364
8	Diet Habit (DH)	0=if absent,1=if present	0 1	652 348

Table 1: Code sheet of concerned independent variables.

Variable	χ^2	Df	P-value	Result	Phi or Creamer's V value
Age	1.350	1	0.245	N.S	-0.037/0.037
Categories of women	0.451	1	0.502	N.S	0.021
Obesity	74.34	1	0.000	S	0.273
Systolic hypertension	1.166	2	0.558	N.S	0.034
Family history	58.357	1	0.000	S	0.242

Table 2: & K L V T X D U H D Q D O \ V L V R I F R Y D U L D W H V V K R Z L Q with p and phi-value for the sample.

Variable	$\hat{\beta}$	SE($\hat{\beta}$)	Wald	Df	P-value	Odds ratio	LCL	UCL
Obesity	1.104	0.142	60.597	1	0.000	3.017	2.285	3.984
FH	0.912	0.139	43.170	1	0.000	2.489	1.896	3.267
Constant	-0.709	0.147	23.145	1	0.000	0.492		

Table 3: 5 H V X O W V R I ¿ W W L Q J W K H 0 X O W L S O H / R J L V W L F and 95% C.I, by using stepwise logistic procedure for the sample.

obtained, it implies that lifestyle of urban area, taking high calories food, less physical activity, invention of remote control equipments and less exercise are the causes of incidence of obesity in the sampl data analysized. Moreover, genetical and environmental behaviors are also the reasons of obesity. e reference group for obesity was taken as non-obese persons. e O.R for obesity was 3.017, which shows that an obese person has 3.017 times more chance of getting signi cant GDM, as compared to non-obese person keeping all other factors constant. As the O.R for obesity was greater than 1 and the 95% con dence interval for obesity did not include 1, therefore, obesity has a positive association with GDM, and was statistically signi cant. e reference group for F.H was taken as absent of F.H persons. e O.R for F.H was 2.489, which means that a pregnant woman in Ebonyi State with positive F.H has 2.489 times more chance of getting a signi cant GDM, as compared to a pregnant woman in which F.H of GDM was absent. erefore, F.H was signi cantly di erent from reference group, and was positively associated with GDM. e reference group for, exercise was sedentary life style. e O.R for exercise was 0.519, which is less than 1 because by general rule, if O.R is less than 1 and chi-squa is signi cant, then there is a protection of exposure against outcome; also 95% con dence interval for exercise did not include 1, therefore, O.R for exercise was signi cantly di erent from reference group, and shows that the person who take light exercise have 0.481 probability of protection against GDM. In the light of the above analysis for the 1000 sampled pregnant women, since it turns out that 3 risk factors, obesity, F.H and exercise were signi cant, that means empirical ndings con rm concept and theory of risk factors. So clinicians and public health personal should take appropriate measures to control these risk factors, and prevention programs should be started against GDM. In the remaining 5 risk factors; age, category of women, income, hypertension and D.H, empirical ndings do not con rm the concept and theories of risk factors. e theme of every study started with past literature and studies done by experts. According to the literature, these ve variables were also the risk factors of diabetes in di erent regions of the world.

Multivariate Version with Interaction Terms

All the interactions terms were calculated separately and tested for signi cance at 5% level of signi cance (Table 4).

In the sample analysis, the main e ect factors: category of women, age, obesity and F.H were signi cant risk factors. Besides the independent factors age was interacted with gender (P=0.005), exercise (P=0.000), and D.H (P=0.016) showed signi cant e ect. Similarly,

Ebonyi State and Health Ministries, with the collaboration of WHO, should arrange the maximum number of seminars and conferences on diabetes. To educate and aware the people against GDM, media should play its signi cant role. Non-Government Organizations (N.G.O's) can also play their role with the help of well- trained health care team, educating both patients and general public with the consequences and complications of this chronic disease. In rural areas, special arrangements should be made for educating the people about balance diet and about this disease. Further studies are needed to specify the change associated with psychosocial problems in Ebonyi State, and to study the genetic components of individually as well as collectively e ect of those risk factors, which are associated to GDM.

References

1. Agresti A (2007) An Introduction to categorical data analysis. (2nd Edn), Wiley, New York, USA.
2. Alonzo TA, Pepe MS (2002) Distribution-free ROC analysis using binary regression techniques. Biostatistics 3: 421-432.
3. Hosmer DW, Lemeshow S (2000) Applied logistic regression. (2nd Edn), Wiley-Interscience Publication, New York, USA.
4. 3 H S H 0 7 K H V W D W L V W L F D O H Y D O X D W L R Q R I P prediction. Oxford University Press, New York, USA.
5. Chou P, Liaq MJ, Tsai ST (1994) Risk factors of diabetes. Diabetes Res Clin Pract 26: 229-235.
6. Jafar TH, Chaturvedi N, Pappas G (2006) Prevalence of overweight and obesity and their association with hypertension and DM in an Indo-Asian population. CMAG 175: 1071-1077.
7. Hagura R, Matsuda A, Kuzuya T, Yoshinaga H, Kosaka K (1994) Family history of diabetic patients in Japan. Diabetes Res Clin Pract S69-S73.
8. Fox J (2005) Maximum-likelihood estimation of the logistic regression model. UCLA/CCPR Notes.