Int. J. Nonlinear Anal. Appl.
Volume 12, Special Issue, Winter and Spring 2021, 793-798
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2021.5447

# Use of a logistic regression model to analyze some variables in the incidence of dental caries disease

Sabah Hasan Jasim Alsaedi

*Mathematic Department - College of Basic Education, Misan University, Iraq*

*(Communicated by Madjid Eshaghi Gordj)*

**Abstract**

**Logistic regression** is commonly statistical methods used in empirical study including categorical dependent variables .In accurate study of reviews dental carries associated with some variables as sex, drink milk, smoking, Extract teeth and diabetics. Aim of study is examine common practices of the results and interpretation of logistic regression data .Determinate some variables (dependent variables that was effect on denial carries. In results and by used two methods as enter methods and steps wise forward to determine sex and extract teeth which more dependent variables on independent variables (yi) dental carries.

*Keywords:* Logistic regression, Dental caries, odds ratio

## 1. Introduction:

Dental caries is a process development of infection in tooth because of some factors and relation with microorganism in oral environment (2.1). Multiple factors, which was causes dental caries as having Carbohydrate sugar , microorganisms, smoking and others, which was effect on dental caries (2.2). Logistic regression known as logistic model or logit model techniques .It is between multiple independent variables and a dependent variables and estimate probability of occur of causes by fitting results and logistic curve .There are many types of regression model are very important in all searches as binary logistic regression and multiple regression . Dependent on variables is not dichotomous and include more than two variables it called multiples regression models (3.1) or between two dichotomous categories (0,1) in binary logistic regression .

*Corresponding author
*Email address:* sabah.h.alsaedi@gmail.com (Sabah Hasan Jasim Alsaedi )

*Received:* January 2021 *Accepted:* July 2021

## 2. Concept of Logistic regression model

It is one of the regression models where the relationship between the dependent variable and the independent variable is nonlinear the causes are always a curved shape of the letter S, but mathematically as;

$$\widehat{\rho_i} = E\left(\frac{y_i}{x_i}\right) = \frac{EXP^{(\delta_\circ + \ \delta_1 x_1 + \ \delta_2 x_2 + \cdots\cdots\cdots + \ \delta_n x_n)}}{1 - \ EXP^{(\delta_\circ + \ \delta_1 x_1 + \ \delta_2 x_2 + \cdots\cdots\cdots + \ \delta_n x_n)}} \tag{2.1}$$

The logistic model is more flexible than traditional regression models, where we can create a relationship between the dependent variable and other independent variables, by estimating the probability of an event occurring and also characterized by its ease of transformation into a linear form using logit transformation as ;

$$l_i \ = \ln(\widehat{\rho_i}/1 - \ \widehat{\rho_i}) \ = (\delta_\circ + \ \delta_1 x_1 + \ \delta_2 x_2 + \cdots\cdots\cdots + \ \delta_n x_n) \tag{2.2}$$

And $(\widehat{\rho_i}/1 - \ \widehat{\rho_i})$ it is mean odds ratio
$l_i$ Mean natural logarithmic (3.2)
**Odds and odds ratio**: Odds an event of odds probability, which was, occurred to the probability that it with not occur a probability of an event occurring is $(\widehat{\rho_i})$ , and probability event which not occur $(1 - \ \widehat{\rho_i})$ and odds rate is a value given by,
Odds of event= $(\widehat{\rho_i}/1 - \ \widehat{\rho_i})$
In logistic regression must be calculate the probability of events occur ring the probability of an event are not occurring the impact of independent variables is usually called as odds .Logistic regression the mean of response variables $(\rho \ _i)$ with explanatory variables $(X_i)$ as an equation $\rho \ _i \ = (\delta_\circ + \ \delta_1 x_1 + \cdots\cdots\cdots + \ \delta_n x_n)$ and there is not good model due to extreme value of $(X_i)$ will give $(\delta_\circ + \ \delta_1 x_1 + \cdots\cdots\cdots + \ \delta_n x_n)$that don't fall between $(0 \ and \ 1)$ .(3.2)

## 3. Estimation of logistic regression

The coefficients of the regression model estimated using the least squares method, which was been used to reduce the errors to the least possible, as in the linear regression model. In the case of logistic regression, the parameters estimated using the maximum likelihood method as the following;

$$L(Bi) = l_i(\widehat{\rho_i}) = \ \prod_{i=1}^{m} \tau_{y_i}^{n_i} \ \widehat{\rho}_i^{y_i} \ (1 - \widehat{\rho_i} \ )^{n_i - \ y_i} \tag{3.1}$$

In order to obtain the coefficients that maximize, we derive the coefficients and make them equal to zero, so that we produce,

$$ln \ \ ( \ \widehat{\rho_i}) = \ \sum_{i=1}^{m} ( \ ln \ \ \tau_{y_i}^{n_i} + y_i \ ln\widehat{\rho_i} + (n_i - \ y_i \ ) \ ln(n_i - \ y_i \ ) \tag{3.2}$$

## 4. Evaluate the quality of the model

In the logistic regression model $(R^2 \ )$, replace the coefficient used as a fit of the proposed regression model for the data of study Nagelkerke $R^2$,Cox and Snell $R^2$. They have the same objective $(R^2)$ as the linear regression as,

$$r^2 = 1 - \left( \frac{l_\circ}{l_1} \right)^{(2/n)} \tag{4.1}$$

$(l_\circ)$: The possible function in the case of the model contains only a constant
$l_1$: Possible function for all variables.
$n$ : Sample size
$0 \leq r^2 \leq 1$ if $r^2$ near to one therefore quality and suitable model .as the formula

$$\breve{r}^2 = \frac{r^2}{r_s^2} \tag{4.2}$$

$$r_s^2 = 1 - (l_\circ)^{(n/2)} \tag{4.3}$$

Wald test for quality of modeling to show the importance of the logistic regression coefficients, the mean and Wald test is used, which has a chi-square distribution, as the probability value of a significant parent is compared with the predetermined level, and the function is significant if the probability value of a Wald test is less than the significance level.

## 5. Descriptive statistics

In table (1) show statistic description for each variables as independent variables $(X_i)$ and $(y_i)$ as dependent variables. According to sex distribution female were recorded 66(62.3%) with mean (1.63) as a high rate than male. Based on residence, patients with live in city area was 94(88.7%) with mean (1.1) than rural area was lower rate. Drinking milk related with dental caries patient having milk was recorded a high rate 56(52.8%) with mean (1.96) .Smoking as another factors associate with dental caries but in this study no smoking patients was 99(93.6%)with mean (0.963) than from smoking patients with dental caries according to sample size of study and habits of patients . Diabetics as another variables which was effect on dental caries was documented 100(94.3%) with mean (0.999) .Extract teeth in patients with dental caries was 105(100%) with mean (0.093). Finally education was recorded a high rate and associate with dental caries was 97(91.5%) with mean (2.9).

| Variables | No of cases | Percentage% | Mean | S.D |
|---|---|---|---|---|
| **Sex** | | | | |
| **Male** | 39 | 36.8 | 1.63 | 0.486 |
| **Female** | 66 | 62.3 | | |
| **Age** | 105 | 100 | 32.86 | 10.86 |
| **Residences** | | | | |
| **City** | 94 | 88.7 | 1.1 | 0.308 |
| **Rural** | 11 | 10.4 | | |
| **Drink milk** | | | | |
| **Yes** | 49 | 46.2 | 1.53 | 0.501 |
| **No** | 56 | 52.8 | | |
| **Smoking** | | | 1.94 | 0.233 |
| **Yes** | 6 | 5.7 | | |
| **No** | 99 | 93.4 | | |
| **Diabetics** | | | | |
| **Yes** | 5 | 4.7 | 1.95 | 0.233 |
| **No** | 100 | 94.3 | | |
| **Extract tooth** | 105 | 100 | 1.65 | 0.214 |

Table 1: Statistical description associated with variables

## 6. Model estimation results and sample selection

By Spss Software statistical program ( **version 26**) was used to estimate logistic regression model , In table (2) was reported constant rate $B$ (1.061) and Wald test (22.573) with significance rate (0.00) , it is low rate from ($p - value \leq 0.05$) in step(0) therefore use alternative hypothesis and refused null hypothesis as in table(2) . In table (3) sex variable was enrolled score (5.278) with significance (0.022) and variable Extract teeth was significance (0.023) with score (5.201).

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 0 | Constant | 1.061 | .223 | 22.573 | 1 | .000 | 2.889 |

Table 2: Variables in the Equation

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 0 | Variables | Sex | 5.278 | 1 | .022 |
| | | Smoking | .193 | 1 | .660 |
| | | Drink milk | .072 | 1 | .788 |
| | | Age | .020 | 1 | .887 |
| | | Extract teeth | 5.201 | 1 | .023 |
| | Overall Statistics | | 9.341 | 5 | .096 |

Table 3: Variables not in the Equation

In table (4) Use likehood method by enter method show two variables which effect on dental caries as sex and extract teeth of patients (-2log likehood was recorded low rate (114.556, 109.652) for the equality of the model.

| Iteration | | -2 Log likelihood | Coefficients Constant | sex | extract |
|---|---|---|---|---|---|
| Step 1 | 1 | 115.108 | -.350- | .811 | |
| | 2 | 114.558 | -.549- | 1.019 | |
| | 3 | 114.556 | -.564- | 1.034 | |
| | 4 | 114.556 | -.564- | 1.034 | |
| Step 2 | 1 | 111.281 | -.494- | .710 | .188 |
| | 2 | 109.702 | -.821- | .915 | .304 |
| | 3 | 109.652 | -.881- | .945 | .334 |
| | 4 | 109.652 | -.883- | .946 | .335 |
| | 5 | 109.652 | -.883- | .946 | .335 |

Table 4: -2log Likelihood Enter method

Table (5) show Cox and Snell R square with Nagelkerke R square were enrolled (0.093) and (0.137) respectively for Variance in logistic regression. In table (6) discuses Wald test for variables in step(1) by enter method was recorded sex variables with Wald test more significance p-value(0.05) and Extract teeth variable was Signiant (0.04) with Wald test (4.213) and all variables smoking, drinking milk and age factors are not effect as variables on dental caries .

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 109.429[a] | .093 | .137 |

Table 5: Model summary

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1 | Sex | .947 | .482 | 3.858 | 1 | .050 | 2.578 |
| | smoking | -.003 | .995 | .000 | 1 | .998 | .997 |
| | Drink milk | -.224 | .481 | .218 | 1 | .641 | .799 |
| | Age | -.002 | .022 | .008 | 1 | .928 | .998 |
| | Extract teeth | .342 | .166 | 4.213 | 1 | .040 | 1.407 |
| | Constant | -.475 | 2.463 | .037 | 1 | .847 | .622 |

Table 6: Variables in the equation

Table (7) Use a method of step-wise forward for enter dependent variables(X) as Smoking , drinking milk and age ,all variables were not effect by score parameter without statistical Significance . Table (8) in step (1) was reported sex variable with Wald test (5.087) with significance p-value (0.024) and extract teeth was enrolled Wald test (4.136) with p-value(0.042) therefore concluded sex variables and extract teeth were more variables from others which was more statistical parameters effect on dental caries in a current study .

| | | | Score | df | Sig. |
|---|---|---|---|---|---|
| Step 1 | Variables | Smoking | .008 | 1 | .927 |
| | | milk | .123 | 1 | .726 |
| | | Age | .020 | 1 | .889 |
| | | Extract teeth | 4.352 | 1 | .037 |
| | Overall Statistics | | 4.531 | 4 | .339 |
| Step 2 | Variables | smoking | .004 | 1 | .952 |
| | | milk | .214 | 1 | .644 |
| | | Age | .001 | 1 | .973 |
| | Overall Statistics | | .222 | 3 | .974 |

Table 7: Variables not in Equation

| | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 1 | sex | 1.034 | .458 | 5.087 | 1 | .024* | 2.812 |
| | Constant | -.564- | .732 | .595 | 1 | .441 | .569 |
| Step 2 | sex | .946 | .469 | 4.070 | 1 | .044* | 2.576 |
| | Extract teeth | .335 | .165 | 4.136 | 1 | .042* | 1.399 |
| | Constant | -.883- | .769 | 1.316 | 1 | .251 | .414 |

Table 8: Step-Wise forward for variables dental caries(p-value$\leq$0.05)*

## 7. Conclusions

This search by use logistic regression in step wise forward for variables of dental caries was been concluded some factors as sex and teeth extraction was very effected on dental carries diseases and statistical significance were sex p-value(0.024) and teeth extract p-value(0.042) .

## References

[1] Axelsson P.,*An introduction to risk prediction and preventive dentistry*, Quintessence, 1999, 1–7.
[2] Derry C., K. Toumba.,*Diagnosis and Prevention of Dental caries*, Pediatric Dentistry, Third Edition, Edited by Richard Welbury, Monty Duggal, Marie-Terese Hosey, Oxford.,2005.
[3] Kleinbaum, D. G., & Klein, M. (2010), *Logistic regression (statistics for biology and health)* (3rd ed.). New York, NY: Springer-Verlag New York Inc.
[4] Shariff A.A .,Zaharim Aand Sopian K ., 2009, *The Comparison Logit and probit regression analysis in estimation the strength of gear teeth* .European Journal of scientific research.27,548-5.
[5] Berkson,J., 1955, *Maximum likelihood and minimum* $\chi^2$ *estimates of the logistic function*, J. Amer. Statist. Asso , Vol. 50, 130–162 (1955).