International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Application of Logistic Regression to Determine Important Socioeconomic and Demographic Factors Affecting the Incidence of Malaria in Sennar State, during the Year 2020

Halima Abdullah Alno Eissa Abdulrahman

Jazan University

Abstract: ***Malaria is considered one of the most serious diseases, with a high-frequency record in health institutions in Sudan in general and Sennar State in particular. This research aims to determine the most important economic and social factors and basic family characteristics that are associated with the incidence of malaria in Sennar State. A logistic regression model was used utilizing data collected randomly from 250 families in Sennar state. The models were evaluated using different statistical techniques. The results revealed that factors that lead to increasing the probability of malaria infection in Sennar state are: (father alive - the marital status - main occupation - relationship with the head of the family - the presence of awareness programs to defeat malaria in the area - the presence of home furniture - availability of food and drinking). All mentioned variables were statistically significant at (p<0.05). The research concluded with a number of recommendations, the most important of which are: Paying attention to environmental health and public health programs in Sennar state by providing health care centres and activating environmental sanitation programs and mosquito spray campaigns in the autumn season in particular, increasing the level of awareness among citizens of the dangers of malaria infection and the necessity of medical examination and take the dose.***

Keywords: binary logistic regression - malaria diseases - economic and social factors

1. Introduction

Malaria is one of the chronic health problems in Sennar State, especially Sudan in general, where the highest rates of visitors to health facilities are recorded, as well as the largest rates of deaths. The rate of malaria infection is estimated at 7.5 million cases and 35 thousand deaths. These numbers are equal to 50% of injuries and 70% of deaths in the east average (WHO).

Malaria constitutes a burden on health institutions, as it was found to represent (20-40) % of the inpatients, (10-40) % of the inpatients, and (10-15)% of the recorded deaths, as well as on mothers and their children, as we find that 37.2% of the maternal deaths are as follows: It causes 18.1% of the incomplete births of children, which causes an increase in the infant mortality rate in Sudan [2]. We also find that there are economic effects of malaria in all states of Sudan, especially the middle ones, which have not been studied, but what is known of them is considered significant [3].

Malaria is considered one of the most prevalent diseases in Sudan, according to the reports of the Federal Ministry of Health, and its spread was helped by the semi-tropical climate of Sudan and the availability of most of the factors of infection with this disease processing.

Many studies have been conducted in Sudan assessing predictors of Malaria. Research conducted by Ibtihal Abed Mahjoub (2012) assessed the factors affecting the incidence of malaria, in Khartoum state" and revealed that gender _ educational level, number of infection times. As for the rest of the factors, they do not have a significant effect, such as infection with enlarged liver - anemia - kinship...etc. Another study conducted by Othman Quality Study (2013 AD): "Single time series models of malaria injuries and deaths in the state of Khartoum - Sudan (1990 - 2011)" The study aimed to identify the best and most efficient statistical model to be used in prediction and to clarify the steps to be taken after determining the model, and the results of data analysis showed that the appropriate model is The model (1,2,0) ARIMA, and based on this model, the numbers of malaria infections and deaths were predicted for the next four years, and the predictive values were consistent with the real values, which indicates the efficiency of the model.

On the other hand Logistic regression has been widely used for health and medical applications. Using these models, we can predict the outcome variable based on the predictive variables. Additionally, these models can provide insights into health outcome variables, as each predictive variable's coefficient indicates its relative contribution to the outcome variable while accounting for the influence of other predictive variables (also known as confounders).

In this study we hypothesized that there a significant association between a number of socioeconomic and demographic factors that significantly increase malaria infection risk in Sennar state. Hence the main objective of this study is to determine the most important economic and social factors and basic family characteristics affecting the incidence of malaria in Sennar state. The importance of the research stems from its use of binary logistic regression models, which are commonly used in medical studies, as they are used to calculate probabilities known as risks. Institutions and voluntary and governmental organizations.

2. Methods

2.1 Study design and sampling procedures

The research relied on the analytical descriptive approach, where the descriptive side represented finding descriptive statistics for various factors, frequency tables, and percentages of data, while the analytical side represented building a binary logistic regression model and examining the accuracy of its features and the quality of matching the data.

The research community means the total group of elements that the researcher seeks to generalize the results related to the problem studied. In this research, the research community consists of family members in Sennar State, which consists of seven localities. Which was chosen in the state of Sennar (Sennar locality, Singa locality, East Sennar locality, Abu Hajjar locality).

The sample size is calculated according to the following equation:

$$n = \frac{Z^2 p(1-P)}{(d)^2} deff$$

By applying the equation, the size of the families to be studied was 250 families distributed over the localities of (Sennar - Sinja - East Sennar - Abu Hajjar).

2.2 The proposed logistic Model: Theoretical Perspective

Suppose we have two independent variables, one of them is x and the other is dependent, and we want to study the relationship between them using the regression method, we suppose that the model that connects them is the following:

$$y = b_0 + b_1 x + e \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

y: Observations of a continuous variable, we assume that the mean of the values of$y$ Observation at a certain value of the variable x is E (y/x) and the variable e represents the error, which is the difference between the observed y and the estimated regression line. The model can be written as follows:

$$E(\frac{y}{x}) = \hat{b}_0 + \hat{b}_1 x \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

It is known in regression that the right-hand side of this model can take values from$(-\infty$ □□□ $\infty)$But when we have two variables, one of them is binary, the simple regression is not appropriate because (predict the mean) of the variable y, in this case, is equal to the probability that y = 1, which is equal to

$$E(\frac{y}{x}) = p(y=1) = p \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Thus, the value of the right side is confined between the zero and one number, including (the two numbers), and the model is not applicable from the point of view of the regression.

To overcome this problem, it was found that the ratio$\frac{p}{1-p}$ It can take values from zero to +∞ Also the amount$ln\ ln\ \left(\frac{p}{1-p}\right)$ can take values from∞ me∞ This is consistent with what happened in normal regression, and accordingly, the regression model can be written in the case of dependent variables as follows:

$$Eln\left(\frac{p}{1-p}\right) = \hat{b}_0 + \hat{b}_1 x \ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

This model is called the logistic regression model, and it is called the shunt from $E\ (y/x)$to me $ln\left(\frac{p}{1-p}\right)$ Convert it to a logit, and the previous equation can be converted into the following form:

$$p = \frac{exp\ (\hat{b}_0 + \hat{b}_1 x)}{1 + exp\ (\hat{b}_0 + \hat{b}_1 x)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Where: exp It is the inverse of the (natural) logarithm, and it is commonly used in medical studies. For example, it is used to calculate the probabilities that are defined as the risks of a hospital case of a specific disease during a period of time during which we are exposed to conditions called the risk factor and known to be related to the disease.

$(\hat{b}_0, \hat{b}_1$(form parameters),(x the independent variable,(y_iThe dependent variable is a descriptive variable with two modes that follow the Bernoulli distribution (0,1).

The ordinary least squares method is not suitable for application to logistic regression Now we ask why this method is not applied to logistic regression to find out the reason[ ], we must distinguish between several models, including the following models:

Linear Probability Model:

Let's say we have the model:

$$y = b_0 + b_1 x + e \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

The variable y is a dummy and binary dependent variable that takes the values (0, 1). $b0$is the truncated part, b_1 is the coefficient of the independent variable X, and e is the random error.

If we estimate the parameters of the model, it is possible to obtain a model called the linear probability model, which gives expected probabilities corresponding to certain values of the independent variables, but we usually encounter several problems, including The error in this model is not homogeneous because the variance of the dependent variable takes different values from the variance of the independent variables, so the variance of the error is:

$$V(e) = P(1 - P)$$

where P is the probability that the dependent variable takes the value one.

P depends on the variable x and this violates one of the conditions of normal regression in normal regression. It is assumed that the error does not depend on the independent variable (x)

* The error e is not normally distributed because P takes only two values, thus violating an important condition of the least-squares method, which is used to estimate parameters [????].

If we apply the linear probability model, the expected possibilities as a result of applying the model can be greater than one or less than zero, and this is illogical and does not agree with the properties of the probability, which will result in a problem if we want to use the expected probabilities in subsequent statistical analysis.

Logistic regression model:
Using the logit model can give a solution to the problems we encountered with the linear probability model. To clarify this, we suppose that we have the following model:

$$ln\ ln\left(\frac{p}{1-p}\right) = y = b_0 + b_1 x + e \ldots\ldots\ldots\ldots\ldots\ldots.(7)$$

or

$$ln\ ln\left(\frac{p}{1-p}\right) = exp\ exp\ [b_0]\ exp[b_1][e] \ldots\ldots\ldots (8)$$

where in is the logarithm of the base e, P is the probability that the variable y does not take the value one, that is, P (Y = 1) = P-value $\frac{p}{1-p}$It is called the preference ratio. Amount$ln\ ln\left(\frac{p}{1-p}\right)$ The logarithm is called the advantage ratio.

The rest of the quantities in the model have the same meaning as we knew earlier.

The logistic regression model is simply a non-linear transformation of the linear regression and therefore it is appropriate to use the properties of the logistic distribution in our case, this restricts the estimated probabilities and makes them confined between zero and one. An estimated probability is as follows:

$$\hat{p} = \frac{exp\ (\hat{b}_0 + \hat{b}_1 x)}{1 + exp\ (\hat{b}_0 + \hat{b}_1 x)} \ldots\ldots\ldots..\ldots\ldots\ldots (9)$$

From it we note the following**:**
1- When is $\hat{b}_0 + \hat{b}_1 x = 0$ van $\hat{p} = .05$
2- when $\hat{b}_0 + \hat{b}_1 x$ Great, the probability value $\hat{p}$ approaching one.
3- when$\hat{b}_0 + \hat{b}_1 x$ small, then the probability value P^ approaches zero [????]

Logistic model response function: simple logistic response function

The response functions of the logistic model take the general form:

$$E(Y) = \frac{exp^{(\hat{b}_0+\hat{b}_1 x)}}{1-e^{(\hat{b}_0+\hat{b}_1 x)}} \ldots\ldots\ldots\ldots\ldots\ldots (10)$$

Among its characteristics are: It takes the form of either an incremental or decreasing curve depending on the j. signalβIn addition, it will often be linear to the extent that it is$E(Y)$ Between (0.2, 0.8) gradually it leads to (0,1) at the end of the x range, and also from its properties it can be converted to linear easily if we put:
$g(x) \ldots\ldots\ldots\ldots..(11) = E(X/Y)$

The shunt can be used:

$$g(x) = ln\ ln\left(\frac{\pi(x)}{1-\pi(x)}\right) \ldots\ldots\ldots\ldots\ldots\ldots (12)$$

Where$\pi(x)$ is the probability that the phenomenon exists while$1-\pi(x))$ is the probability of the absence of the phenomenon

$$g(x) = \beta_0 + \beta_1 x \ldots\ldots\ldots\ldots (13)$$

g(x) is called logistic and the significance of the logistic model is due to the magnitude ($e^{\beta i}$) gives what we call the risk ratio (odds ratio) to clarify the idea of the risk ratio as follows: Let x take one of the two values (0,1)

$$\text{Odds}\ (1) = \frac{\text{The number of those who have the disease among those who carry the malar}}{\text{The number of those who don'thave the diseasee among those eho c}}$$
x=1

$$\text{Odds}\ (0) = \frac{\text{The number of those who have the disease among those who don}}{\text{The number of those who don't have the disease among those who c}}$$
x=0

The odds ratio is defined as the ratio of the two:

$$\text{Odds Ratio} = \frac{Odds\ (1)}{Odds\ (0)} \ldots\ldots\ldots\ldots\ldots(14)$$

This ratio represents the risk of developing the disease for people who have the trait compared to those people who do not have the disease and we can analyze logistic regression directly from knowing the risk ratio.

2.3 Estimation of logistic regression parameters and test of significance

The parameters of the logistic model are estimated by a commonly used estimation method called the greatest possibility method: It is one of the most famous methods of estimation in statistics, as the maximum possible function is known to measure the observed probabilities for a number of n independent variables. $(P1, P2, \ldots..\ Pn)$ Which falls in the sample and goes through a number of steps, namely: We start with the initial values of the coefficients of the parameters, we get the expected value of the logit, we transform the logit into a probability, and we substitute the values for (Y) and we calculate the logarithm of the weighting function and we repeat the steps for each case of the sample (0,1) and we take the estimates that generate the highest value of the logarithm (close to zero)), it is represented by the product of the probabilities and the function of greatest possibility is as follows:

$$L = Prob\ (P1P2 \ldots\ldots\ Pn)$$

An estimate is chosen for the various possibilities involved in this function by obtaining their values. If they were substituted into the function of the greatest possibility, it would have given a maximum limit for the function. In another way, it is that which makes the logarithm of the function of greatest possibility as large as possible, or we make negative 2 multiplied by the logarithm of the function

$(-2\,Log$ "L"As small as possible, which is an iterative method that relies on the repetition of arithmetic operations to improve the estimates until the differences between the sum of the squares of errors (SSE) in step (i+1) and step (i) become as small as possible.

Wald test

The test is performed by the father's statistic, the test statistic is known as:

It is distributed as a chi-square distribution with one degree of freedom, which is simply the statistic square (t) to evaluate the performance of the model. There are many statistical methods and they are used to compare different models or to evaluate the performance of only one model, and it tests the significance of a certain parameter in the model as it tests the significance of the parameters:

Null hypothesis $H_{0:\,\beta j=0}$

against the alternative hypothesis $0 \neq \beta j\ H_{1:}$

The test is based on statistics:

$$wald = \left[\frac{\hat{b}_1}{s(\hat{b}_1)}\right]^2 \ldots\ldots\ldots (15)$$

The maximum possible ratio or chi-square statistic:

The greatest possibility ratio is used in conducting a test that is used in the null hypothesis test.

Test stats

$$LR(i) = -2(logL(\hat{b}_0) - logL(\hat{b}_0\hat{b}_1) \ldots\ldots\ldots (16)$$

whereas: i: represents the number of independent variables in the model Log $L(\hat{b}_0\hat{b}_1)$It is the logarithm of greatest potential.

For a form with two parameters. Log $L\hat{b}_0$ It is the logarithm of the maximum potential of a model with one parameter. The test statistic is distributed according to a chi-square distribution with one degree of freedom. It is noticeable that the Log $L(\hat{b}_0\hat{b}_1)$ It is called the logarithm of the greatest possibility of the unconstrained model, which is the logarithm of the function of the greatest possibility after substitution by the estimation of the two parameters.

Log $L(\hat{b}_0\hat{b}_1)$ It is called the logarithm of the restricted model, and it is the logarithm of the maximum possible function after substituting the parameter estimate $\hat{b}_0$The test is conducted to reject or accept whether the form is moral or not.

As for testing the adequacy of the entire model and its quality, we used its (F) and (F) statistics.(R^2) In linear regression, but in the case of the logistic model, the log-likelihood ratio that follows the chi-square distribution is used according to the relationship:

$$\chi^2 = 2(log_e\,L_0 - log_e\,L_1) \ldots\ldots\ldots (17)$$

L_0: The value of the maximum possible function that contains (i) a variable.

L_1: The value of the maximum possible function that contains $(i-1)$ a variable.

If the value of (Sig) is less than (0.001), which confirms the significance of the fully adapted model.

2.4 Study variables and operational definitions

Personal factors, **i**ncluding (age, gender, educational level, marital status, and income level). Health factors and involving some diseases such as Pneumonia and Viruses Pneumonia. Environmental factors and includes human behavior include some habits and practices, the most important of which are:

1) Use mosquito nets at night when sleeping and cover windows and doors with narrow wire mesh.
2) The necessary health propaganda of the disease.
3) Spray swamps with a larvicide such as petroleum jelly.
4) The use of insecticides against mosquitoes, such.

Finally economic and Social Factors and this group include the level of houses, live sewage, electricity, and poverty, ... Etc., it has a great impact, knowing that malaria is a disease in developing countries, and malaria also affects the economic and social situation and human activities such as building reservoirs, constructing roads, wars and displacement. Study Variables are shown in Table 1

Table 1 shows the variables and symbols of a group of

Symbol	**Variable name**	Symbol	**Variable name**
X_{21}	Food and drinks	X_{11}	Zone type
X_{22}	Electricity and water	X_{12}	local
X_{23}	fuel	X_{13}	Relationship with the head of the family
X_{24}	Education (daily expenses, lessons...)	X_{14}	Type
X_{25}	Health (interviews and examination of treatments)	X_{15}	the age
X^{26}	social contributions	X_{16}	the condition
X27	Telecom	X_{17}	The educational level of the individual
X28	gifts or giveaways	X_{18}	Is the mother alive?
X29	Expenses for children (other than education)	X_{19}	Is the father alive?
X210	clothes (not school)	X_{110}	The period of residence in the current location (domestic) in years
X211	home maintenance	X_{111}	Currently working
X212	Education (fees, clothes)	X_{112}	main occupation
X213	Property tax only	X_{222}	toilet type
X214	family events	X_{223}	power supply
X215	cart maintenance	X_{224}	Drinking water source
X216	Home furnishings	X_{225}	Is there a pool or pool near the house?
X217	home ownership	X_{226}	Is there a visitor at home?
X218	building type	X_{227}	Is there periodic spraying of pesticides in the neighborhood/ village?

X219	The number of rooms	X_{228}	Are there awareness programs to defeat malaria in the region?
X_{220}	Floor-type of rooms and halls	X_{229}	What do you think of the health services in the region?
X_{221}	roof type		

3. Results (Model application)

A two-response logistic regression model was built for the study variables, where the variables were divided into groups, each group testing a hypothesis of the research, where a two-response logistic regression model was applied to each group, as follows:

3.1 First Model: Family member characteristics

Table (2) shows the parameters of the optimal **first** model that we obtained in the fourth cycle and their estimations, in addition to the standard error for each parameter and the statistic (Wald) for each parameter with the number of degrees of freedom and its statistical significance. Through the table, the significance of the following variables was shown. (x_3), (X6), (X9), X12)) and its impact on the incidence of malaria, and we note the high value of (Wald) for the moral variables, and we find that the variable (Relationship (name) to the head of the family Significance score 041. (Positively affects the incidence of malaria, while Marital status..021, Is (name's) natural father alive 004., main occupation 033.) has a negative effect on the incidence of malaria.

For this model the number of iterative cycles of the derivatives of the maximum possible function of the basic data variables to obtain the lowest value of the negative twice the logarithm of the greatest possibility function we obtained the lowest value, which is ((369.873). We stopped at this cycle because we then obtained the least differences between the transactions, where these differences reached less than (0.001), and it is considered the best result because the negative double of the logarithm of the function of the greatest possibility is at its lowest end at this cycle.

Table 2: Shows the regression coefficient, standard error, and the Wald statistic with their degrees of freedom and significance for the model of family member characteristics variables.

		B	SE	Wald	df	Sig.	AOR	95.0% CI for AOR	
								Lower	Upper
Step 1(a)	X_{11}	.098	.300	.107	1	.743	1.103	.613	1.985
	X_{12}	-.135	.141	.912	1	.340	.874	.663	1.152
	X_{13}	.212	.109	3.797	1	.041	1.236	.999	1.530
	X_{14}	-.058	.357	.026	1	.872	.944	.469	1.900
	X_{15}	007	.013	.three hundred fifty	1	.554	1.008	.983	1.033
	X_{16}	-.613	.266	5.325	1	.021	.542	.322	.912
	X_{17}	-.005	.100	.003	1	.959	.995	.818	1.210
	X_{18}	-.155	.291	.282	1	.596	.857	.484	1.516
	X_{19}	-.494	.171	8.368	1	.004	.610	.437	.853
	X_{110}	.004	.011	.143	1	.705	1.004	.982	1.027
	X_{111}	.018	.397	.002	1	.964	1.018	.468	2.216
	X_{112}	-.240	.113	4.554	1	.033	.786	.631	.981
	Constant	4.131	1.293	10.199	1	.001	62.212		

When assessing the goodness of the fit of the model we found that the value of the chi-square from the model waws (37,259) and that the probabilistic value of the chi-square statistic is (000.) which is significant at the level of statistical significance (0.05), and thus we reject the null hypothesis and accept the alternative hypothesis that is, the model is significant and represents the data well. On the other hand the results of the Hosmer and Lamshaw test was calculated to ensure the quality of the agreement of the first model, by using the chi-statistic to test the significant difference between the observed values and the expected values and thus testing the estimated model from which the expected observations were calculated as shown in (**appdendix1**), where it was found that the test was not significant, and this can be seen from which indicate a good degree of fitness of the model.

Further we assessed how well the model is reconciled with variables Characteristics of family members **whereas** null hypothesis: The model represents the data well vs the alternative hypothesis: The model does not represent the data well. Referring (a**ppendix2**) for the quality of the model fit, we find that the value of the chi-square is (12.085) and that the probabilistic value of the chi-square statistic is (163.), which is not significant at the level of statistical significance (0.05) and degrees of freedom (8). Thus, we accept the null hypothesis, meaning that the model represents the data well

Table (3) shows the classification rate of people infected with malaria in Sennar state amounted to (63.8%), and the percentage of incorrect classification reached (36.2%). The percentage of correct classification of those infected and not infected with malaria amounted to (70.7%), and the percentage of incorrect classification amounted to about (29.3%).

Table 3: The correct classification of the family member characteristics variables model

	Observed		Predicted		
			Has the name been infected in the past 30 days?		Percentage Correct
			yes	no	
Step 1	Infected in the past 30 days?	yes	44	25	63.8
		no	101	350	77.6
	Overall Percentage				70.7

The basic variables model can be formulated as follows:

$$\mathbf{Log}\left(\frac{\hat{P}}{1-\hat{P}}\right) = \mathbf{Y = 4.1320 + 0.09X_{11} - 0.13X_{12} + 0.21X_{13} - 0.05X_{14} + 0.007\ X_{15th} - 0.61X_{16} - 0.005X_{17} - .15X_{18} - 0.49X_{19} + 0.004X_{110} + 0.01X_{111} - 0.24X_{112}}$$

From the basic family characteristics group model, we note the clear significance of the variables (parent is alive) It ranked first in the model with statistical significance (0.004), and its regression coefficient was (-0.494), standard error (171.0), Wald's statistic (8.368), and (marital status) ranked second in terms of significance with statistical significance (021). And its regression coefficient (-0.613), standard error (266.0), and Wald's statistic (5.325), (A).for the main profession ranked third in terms of significance, with statistical significance (033.0), and its regression coefficient (-0.240), standard error (113.0), and Wald's statistic (4.554), while we find that the variable (Relationship with the head of the family) ranked fourth in terms of significance, with statistical significance (041.0) and its regression coefficient (0.210), standard error (109.0) and Wald's statistic (3.797). Sennar, but its effect is less than that of the significant variables.

3.2 Second Model: Economic and Social Factors:

From the number of iterative cycles of the derivatives of the maximum potential function of the basic data variables to obtain the lowest value of the negative twice the logarithm of the maximum potential function, to obtain the optimal estimate of the parameters of the model for the derivative of the negative twice the function of the greatest possibility, we note that in the seventh cycle of the derivative of the negative twice the function of the greatest possibility **(-2** Log Likelihood**)** We get the lowest value which is (116.625). We stopped at this cycle because we then obtained the least differences between the transactions, where these differences reached less than (0.001) and are considered the best result because the negative double of the logarithm of the function of the greatest potential is at its lowest end at this cycle.

Table 4 shows the parameters of the optimal model that we obtained in the fourth cycle and their estimations, in addition to the standard error for each parameter and the statistic (Wald) for each parameter with the number of degrees of freedom and its statistical significance. Food and drinks, Having home furniture, Are there awareness programs to defeat malaria in the region?) and its effect on the incidence of malaria, and we note the high value of (Wald) for the moral variables, and we find that the variable (Are there awareness programs to defeat malaria in the region?) has a positive effect on the incidence of malaria, while (Food and drink, home furnishing negative impact on the incidence of malaria.

Table 4: The regression coefficient, standard error, and the Wald statistic with their degrees of freedom and significance for the model of economic and social variables

Variables in the Equation									
		B	SE	Wald	df	Sig.	AOR	95.0% CI for AOR	
								Lower	Upper
Step 1[a]	X_{21}	0.001	0.001	4.263	1	0.039	1.001	1	1.002
	X_{22}	-.002-	7	0.132	1	0.717	0.998	0.985	1.011
	X_{23}	0.001	0.005	0.002	1	0.963	1	0.99	1.01
	X_{24}	0.002	6	0.114	1	0.736	1.002	0.99	1.015
	X_{25}	0.0003	0.001	8	1	0.927	1	0.998	1.002
	X_{26}	-.010-	6	2.396	1	0.122	0.99	0.978	1.003
	X_{27}	-.003-	0.004	0.425	1	0.515	0.997	0.989	1.006
	X_{28}	0.003	0.004	0.471	1	0.493	1.003	0.995	1.01
	X_{29}	-.001-	0.002	0.513	1	0.474	0.999	0.995	1.002
	X_{210}	0.001	0.001	1.36	1	0.243	1.001	0.999	1.003
	X_{211}	0	0.001	0.029	1	0.864	1	0.999	1.002
	X_{212}	0.001	0.001	0.888	1	0.346	1.001	0.999	1.003
	X_{213}	0	0.001	0.048	1	0.827	1	0.999	1.001
	X_{214}	0.0005	0.001	0.001	1	0.981	1	0.999	1.001
	X_{215}	0.0006	0	0.085	1	0.77	1	0.999	1
	X_{216}	0.004	0.002	5.438	1	0.02	1.004	1.001	1.007
	X_{217}	0.077	0.696	0.012	1	0.912	1.08	0.276	4.227
	X_{218}	0.176	0.649	0.074	1	0.786	1.193	0.334	4,256
	X_{219}	0.126	0.178	0.5	1	0.48	1.134	0.8	1.606
	X_{220}	-.369-	0.443	0.691	1	0.406	0.692	0.29	1.649
	X_{221}	0.228	0.337	0.457	1	0.499	1.256	0.649	2.429

	X_{222}	-.106-	0.347	0.094	1	0.759	0.899	0.456	1.773
	X_{223}	-.897-	0.621	2.084	1	0.149	0.408	0.121	1.378
	X_{224}	0.663	0.488	1.846	1	0.174	1.94	0.746	5.046
	X_{225}	-.881-	0.611	2.082	1	0.149	0.414	0.125	1.371
	X_{226}	-.339-	0.807	0.177	1	0.674	0.712	0.146	3.466
	X_{227}	1.4	0.719	3.786	1	0.052	4.054	0.99	16,607
	X_{228}	-1.857-	0.664	7.81	1	0.005	0.156	0.042	0.574
	X_{229}	0.542	0.514	1.114	1	0.291	1.719	0.628	4.704
	Constant	1.489	3.535	0.178	1	0.673	4.435		

We find that the value of the chi-square for this model is (466.927) and that the probabilistic value of the chi-square statistic is (0.021), which is significant at the level of statistical significance (0.05), and thus we reject the null hypothesis and accept the alternative hypothesis that is, the model is significant and represents the data well(**appendix3**). Along with this the results of the Hosmer and Lamshaw test to ensure the quality of the agreement of the second model, by using the chi-statistic to test the significant difference between the observed values and the expected values indicated that the test was not significant, and this can be seen from (**appendix4**) signifying that the differences between the observed and expected values are minor. Moreover Table 5 shows the classification rate of people infected with malaria in Sennar state amounted to (68.1%) and the percentage of incorrect classification reached (31.9%). The percentage of correct classification of those infected and not infected with malaria was (75.4%), and the percentage of incorrect classification amounted to about (24.6%)

Table 5: The correct classification of the model economic and social variables

Classification Table					
	Observed		Predicted		
			Has the name been infected in the past 30 days?		Percentage Correct
			yes	no	
Step 1	Injury in the past 30 days	yes	47	22	68.1
		no	78	373	82.7
	Overall Percentage				75.4

The model of economic and social factors variables can be formulated as follows:

$$\text{Log}\left(\frac{\hat{P}}{1-\hat{P}}\right) = Y = 1.489 + 0.001X_{21} - 0.002X_{22} + 0.001X_{23} + 0.002X_{24} + 0.0003X_{25} - 0.010X_{26} - 0.003X_{27} + 0.003X_{28} - 0.001X_{29} + 0.001X_{210} + 0.0005X_{211} + 0.001X_{212} + 0.0006X_{213} + 0.0007X_{214} + 0.0008X_{215} + 0.004X_{216} + 0.077X_{217} + 0.176X_{218} + 0.126X_{219} - 0.369X_{220} + 0.228X_{221} - 0.106X_{222} - 0.897X_{223} + 0.663X_{224} - 0.881X_{225} - 0.339X_{226} + 1.40X_{227} - 1.857X_{228} + 0.542X_{229}$$

From the model of the set of economic and social factors variables, we note the clear significance of the variable (Existence of awareness programs to defeat malaria in the region) ranked first in the model with statistical significance (0.005), and its regression coefficient was (-1.857), standard error (.664) and Wald's statistic (7.810), and the variable (existence of home furnishings) It ranked second in terms of significance, with statistical significance (0.020), its regression coefficient (0.004), standard error (0.002) and Wald's statistic (5.438), and we find that the variable (providing food and drink) It ranked third in terms of significance, with statistical significance (039.0) and its regression coefficient (0.001) and standard error (001.0) and Wald's statistic (4.263), and we note the insignificance of the rest of the variables within the model of economic and social factors variables, as they may affect the incidence malaria in Sennar state, but its effect is less than the effect of the significant variables.

Findings and Recommendations:

4. Conclusion

Application of the logistic regression models revealed that certain factors contribute to an increased risk of malaria in Sennar state. Family characteristics such as marital status, main occupation, and relationship with the head of the family, as well as the father's status of being alive, were found to be important factors. Additionally, economic and social factors such as the presence of awareness programs to fight malaria in the region, the availability of home furniture, and access to food and drink were also significant. In order to combat malaria in Sennar state, we recommend that the government and official authorities prioritize environmental health and public health programs. This can be achieved through the provision of healthcare centers, the activation of environmental sanitation programs, and the implementation of mosquito spray campaigns during autumn. Raising awareness among citizens about the dangers of malaria and the importance of medical examination and treatment is crucial. Furthermore, controlling the mosquito population by filling ponds and swamps or using insecticides can also be effective.

Declarations

Competing interests
The authors have no actual or potential competing financial or non-financial competing interests.

Funding
NIL

Authors' contributions
MSM and conceptualized and designed the research project, and carried out the practical work. AA and MSM performed the statistical analysis and finalized the manuscript. All

authors provided significant input in the manuscript, read and approved the final version of it.

References

[1] Abu Sidra Fathi Saleh, Statistics and Econometrics, National Book House, Benghazi, 1999 AD.
[2] Ahmed Hafez, Diseases endemic to Africa and Asia, this series is published with the assistance of the Supreme Council of Science in 1961 by a group of authors.
[3] Azhari Abdullah Al-Hajj, Statistical Analysis of Variables, University of Khartoum Press, 2010.
[4] Amin Ibrahim Adam, Basic Statistical Principles in Applied Nonparametric Methods, Makkah Al-Mukarramah, 2005
[5] Thomas Kinnear, James Taylor, Marketing Research, An Applied Introduction, Part One, translated by Abd al-Rahman Daalah and Abd al-Fattah al-Sayed, Dar al-Marikh Publishing, Riyadh, 1983 AD.
[6] Jalal Mustafa Al-Sayyad, Statistical Inference, Dar Al-Marikh, 2nd Edition, Riyadh, 1993.
[7] Mr. Hassan Saleh, Ahmed Ibrahim Idris, Health education and its impact on combating endemic diseases in the national capital, Khartoum.
[8] El-Sherbiny Shawky El-Sayed, Estimates of Multi-Level Models, Mansoura University, vol. 1, 2005.
[9] Salah al-Din Mahmoud Allam, Inferential statistical methods in analyzing psychological, educational and social research data, Egypt, Al-Azhar University, 1426 AH, 2005 AD.
[10] Adel Ahmed Hussein Babtain, Logistic Regression and How to Use It, Umm Al-Qura University, 1, 1430 AH.
[11] Abdulaziz Al-Tarih, environment and human health.
[12] Adnan Ghanem, Fridkhil Al-Jaouni, Logistic Regression Technique, Damascus University, 1, 2011 AD.
[13] Firas John, Logistic Regression, New York, Part One, 2003 AD.
[14] Farid Kamel Abu Zina, Abdel Hafez Al-Shayeb, Imad Abania, Muhammad Abdel-Al, Scientific Research Methods, Statistics in Scientific Research, Amman Arab University, 1984.
[15] 15th. kleenbaum caper & Muller, translated by Abdel Fattah El-Sayed, The Application of Regression Analysis to Multiple Variable Methods, New York, 1988.
[16] Muhammad Othman Al-Amin Nouri, Statistics and Measurement in the Social and Behavioral Sciences, Part Two, Khartoum, 2008.
[17] Mahmoud El-Derini, Polynomial Logistic Regression Model, Tanta University, Egypt, c. 1998 AD.
[18] Nitrogon Azman, William Cutter, Applied Linear Statistical Models, Regression Analysis of Variance and Design of Experiments, Part One, translated by Kangwanis Ismail and Al-Zeer Abdul Hamid, King Saud University, Riyadh, 1990. secondly: Foreign references:
[19] Harvey Motulsky&Arther chistopoulos-Fitting Model to -1-Davied Hosmer- Applied Logistic Regression-second edition- New York- 2000.
[20] Draper&smith - Applied Regression analysis - New York-1981.
[21] Hosmer Lemeshow &Klor –Goodness of fit testing for the logistic model when the estimated-1 .probabilitiesare small- sedny-1988.