Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 23, Issue 3(2021) ISSN$_{Print}$: 1816-9171

The use of binary logistical regression models to classify the standard of living of some countries in the word

Teacher. Namareq qasim husain
Department of Economics, Faculty of Management and Economics/University of Kerbala, Iraq
Corresponding Author: Namareq qasim husain Email: namariq.q@uokerbala.edu.iq

Abstract:*The purpose of this research is to study and analyses, from a statistical point of view, the most important determinants of the country's situation (poor, rich), and to identify the most important moral factors that make sense using the logistic regression of the binary logistics technology and reject the hypothesis for coefficients whose moral level is greater than where we have adopted a significance level = 0.05.The results of the study showed that using a two-response logistic regression gave logical results behind the studied phenomenon*.

Keywords*: Logistic Regression Model, Classification determinants, logistical decline, binary logistics.*

Introduction

The main goal in most studies and research is to analyze and evaluate the relationships between a group of variables for the purpose of obtaining a formula that describes these relationships, and regression analysis is used to determine the form of the relationship between the variables by finding a mathematical model that shows the relationship between the dependent variable and the explanatory variables in the case that the dependent variable is discontinuous and not continuous The regression method cannot be applied. Rather, we use other statistical methods such as distinct analysis or logistic regression, as there are many models in which the dependent variables (response variables) are descriptive and specific (qualitative and limited dependent variables) that is, those expressed in a specific set of characteristics or facts. And it is determined (bounded) by restrictions that cannot be ignored, as it is not always the dependent variable that takes a numerical value that is confined between as in a normal reference. There are many applications that use both specific and descriptive response variables and discontinuous selection models in several fields including economics, finance, marketing, social sciences, ... etc.

Binary and multiple response models are a special case of regression models.

Research problem:
In the last two decades, many phenomena that suggest the economic and social changes of countries around the world have been monitored, which led to the emergence of economic and social disparities in most countries of the world, and many studies dealt with trends in economic and social levels, but studies that dealt with public income were few and rare. The two researchers decided to conduct this field study in order to determine the standard of living in light of the high prices of most basic materials, which is one of the most important challenges facing society and economic decision-makers because of their negative social, educational, health and economic repercussions.

Research Objective:
The aim of this research is to study and discuss the theoretical and practical importance of the logistic regression technique in the analysis due to the nature of the data collected and the dependent variable was of the binary type the country condition (poor, rich)

Research Methodology:
In order to achieve the desired goals of this research, the descriptive and analytical approach was adopted in the description of the logistic regression with a focus on how to estimate its features and address its most important characteristics, through books and references related to the research topic, as well as analyzing the indicators and data obtained from the research sample regarding indicators Related to the research topic.

Study hypotheses
This research relies on the following hypotheses:
The first hypothesis: The significance of the regression coefficient b_1 for the income variable is zero in the community
$H_0 : b_1 = 0$

That is, the income variable has no significant effect on the dependent variable.

The second hypothesis: The significance of the regression coefficient b_2 for the variable cost of living is equal to zero in the society

$H_1 : b_2 = 0$

That is, the cost of living variable does not have a significant effect on the lack of adequate income, and so on with respect to all the explanatory variables.

Previous studies:

The study of the economic and social determinants of income insufficiency is a relatively recent topic, as many studies have focused on topics close to this topic such as studies (household income and expenditures, consumer behavior patterns, determining the economic and social structure, etc.) and in general interest has increased in studying the factors affecting income since the beginning of the century. The current, which has been accompanied by an increase in prices and insufficient incomes, and the global literature that has dealt with the issue of economic and social determinants of insufficient income is rich and multiple, unlike the case in Arab countries, as previous local studies were limited to researching the elasticities of spending on commodities and the characteristics of the food system and their needs in general. Most of these studies focused on multivariate statistical analysis, building income standard models, and the scarcity of studies that used logistic regression to study socio-economic issues.

What distinguishes this research is the focus on the use of the logistic regression technique on some economic applications, noting that this technique is commonly used in many medical and social studies due to the privacy of data in these two fields, which are often of the binary type, in addition to the parameter estimates according to the model Logistic is acceptable in the absence of some restrictions on linear and logarithmic regression.

The Theoretical Side

Logistic Regression Model [7],[11]

The logistic regression model is one of the important models used in statistical applications and data analysis. The logistic regression model is used in general to explain the relationship between one or more explanatory variable and the dependent variable, as it is widely used in many economic, medical and social sciences.

The importance of the logistic regression model stands out for several reasons, including:

1. There are no pre-assumptions about the explanatory variables.
2. The model determines the belonging of the new observations to which societies belong and the probability of this belonging. The model can also be used to analyze the binary and multiple dependent variable.

The Maximum Probability (ML) method is used to estimate the model parameters.

In the following, we will discuss in detail the use of the two-response logistic model:

The Regression Logistic model (binary response) [2]:

This model is generally used when the dependent variable is a two-response response (event occurrence, non-occurrence of the event), as the logistic regression model is characterized by its flexibility, and the logistic regression model is based on the basic assumption that the dependent variable is descriptive with two sides and not a continuous variable knowing that there are many statistical methods to treat Such cases, including the method of multiple linear regression, but this method faces certain problems, including: (error variance is not distributed naturally, the inability to interpret the predicted values probably any values whose value is between 0 and 1. Therefore, a model such as the two-response logistic model that is distinguished This model, in addition to what was mentioned above, that the probability estimates of the occurrence of the event calculated using logistic regression are acceptable results, especially after overcoming some problems (bias of estimates, inefficiency of estimates, inaccuracy of conclusions based on the logistic model, which reduces the usefulness of the model) that arise as a result of the absence of some Hypotheses of the model, the two-response logistic model in which the relationship between the dependent variable and the explanatory variable is non-linear, which called Researchers transformed the paradigm into general linear form, using what is known as Logit transformation.

1. Logit Transformation [9],[11].

In simple linear regression, whose explanatory variables and response variable take continuous values, the model linking the variables is as follows:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (1)$$

when a certain value of the variable (X) is $E\ (Y)$, the model can be written as follows:

$$\boldsymbol{E(y \mid x) = \beta_0 + \beta_1 X} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (\mathbf{2})$$

It is known in regression that the right-hand side of this model takes values from $(-\infty)$ to (∞), but when it is a double-response variable (y), the simple linear regression model is not suitable because:

$$\boldsymbol{E(y/x) = \pi(y = 1) = \pi} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (\mathbf{3})$$

Thus, the value of the right-hand side is confined to the two values $(1,0)$ and thus the model is not applicable from the point of view of the regression.

One of the ways to solve this problem is to introduce an appropriate mathematical transformation on the dependent variable (y). Berkson (1944) proposed converting the logistic model into a linear function by taking the natural logarithm, the value of $(0 < \pi < 1)$ and then the ratio $\left(\frac{\pi}{1-\pi}\right)$ is a positive expression $\mathbf{(}0 \le \frac{\Pi}{1-\Pi} \le \infty\mathbf{)}$ and by taking the natural logarithm of base (e) to convert $\left(\frac{\pi}{1-\pi}\right)$, its value field will be limited to $\left(-\infty \le \mathrm{LOG_E}\left(\frac{\Pi}{1-\Pi}\right) \le \infty\right)$ Therefore, the regression model can be written in the case of one explanatory variable as follows:

$$\boldsymbol{log_e\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (\mathbf{4})$$

Formula (4) represents the logistic regression model in linear form, with a sign of what is known as (logit).

If we have more than one explanatory variable, then the logistic regression model becomes as follows:

$$\boldsymbol{log_e\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (\mathbf{5})$$

$$\mathbf{i = 1, 2, \ldots\ldots, n}, \qquad \mathbf{j = 1, 2, \ldots\ldots, p}$$

As:

β_j : Unknown features to be graded.

x_{ij}: explanatory variables.

$$\boldsymbol{\frac{\pi}{1-\pi} = \frac{\pi(y=1)}{1-\pi(y=1)} = e^{\left(\beta_0 + \sum_{j=1}^{n} \beta_j X_{ij}\right)}} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (\mathbf{6})$$

The formula for response probabilities for a two-response logistic regression model is written as follows:

$$\pi = \frac{1}{1 + \boldsymbol{e}^{-\left(\beta_0 + \sum_{j=1}^{n} \beta_j X_{ij}\right)}}$$

That the form of the relationship between the explanatory variables (Xij) and the probability of response P_i cannot be linear and it takes a curved form, and the diagram in Figure (1) shows that

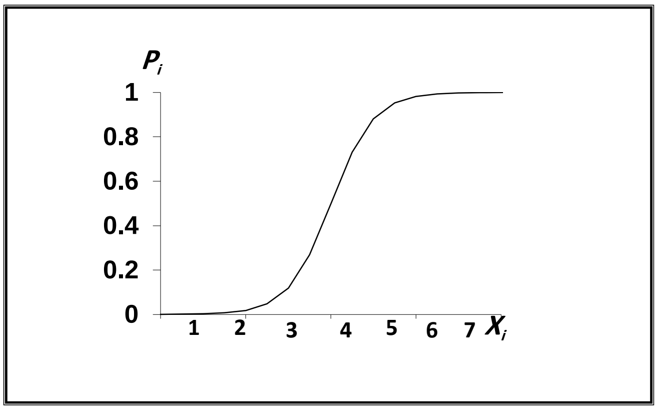

Figure [6] (1) The relationship between the response probability and the explanatory variable

Through the drawing, the characteristics or characteristics of the logistic regression function are shown, as it is found that the function is continuous and takes values between (0,1) and Y approaches zero whenever the left side approaches ($-\infty$) and Y approaches (1) whenever the right side approaches (∞).

2. Assumptions of logistic regression [2]

The variable dependent on the binary logistic model is divided into two classes (descriptive variable) for this model takes one of the two values (0 or 1), while the explanatory variables can be continuous or discrete and descriptive binary or multiple and the conditions for this model are: -

- The random error is not assumed to be distributed normally.
- Assuming a large sample size, the reliability of estimating the model using the model decreases as the sample size decreases.
- There is no correlation between errors and explanatory variables, and the expectation of errors equals zero.

3. Estimating the parameters of the two-response logistic model [1] [9]

In logistic regression, the value of the dependent variable (Y) is either (zero) or (one). Therefore, the least squares method cannot be used. Therefore, Berkson suggested using the Maximum Likelihood method, which is one of the most popular estimation methods in statistics because of its statistical properties. Desirable instead of using the least squares method, which aims to reduce the error squares to the least possible, as is usual in the case of linear regression, as estimation by (ML) method includes finding the values of the parameters that maximize the possibility of a single observation according to the following formulas:

$$L_i = P(Y_i = y_i) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (7)$$

$$L_i = \pi_i^{yi}(1-\pi_i)^{1-y_i} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (8)$$

$$L_i = \text{EXP}\{Ln(\pi_i^{yi}(1-\pi_i)^{1-y_i})\} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (9)$$

$$L_i = \text{EXP}\{y_i Ln(\pi_i) + (1-y_i)Ln(1-\pi_i)\} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (10)$$

The formula for the greatest possibility function for (n) from the observations is: Y1

$$L = P(Y_1) . P(Y_2)\ldots\ldots\ldots P(Y_n) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (11)$$

$$L = \Pi_{i=1}^{n} P_r(Y_i=y_i) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (12)$$

$$L = \Pi_{i=1}^{n} L_i$$

$$L = \Pi_{i=1}^{n} EXP\,\{y_i Ln(\pi_i) + (1-y_i)Ln(1-\pi_i)\} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (13)$$

$$L = exp(\sum_{i=1}^{n}(y_i Ln(\pi_i) + (1-y_i)Ln(1-\pi_i))) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (14)$$

4. Numerical methods maximum likelihood [11]

Since we can obtain the greatest probabilities of the linear probability model by making the log likelihood function equal to (zero) and by solving these equations the estimations of the parameters are determined. Numerical methods are often used in solving nonlinear models to find the estimations that maximize the function of the logarithm of the possibility equation by setting the derivative of the logarithm of the probability function to zero.

5. Tests of the Logistic Model Parameters: -

There is a group of statistical tests used in logistic regression, including:

5.1 Wald test [7], [8]

Among the important tests that are used to demonstrate the significance and significance of the effect of the explanatory variable on the dependent variable in non-linear models we use the ((wald) test, which is one of the most important tests to know this and has a chi-square distribution (χ^2). If the degree of freedom is one, then it equals a square (t), and the significance of the value of (Wald) statistic is compared with the level of significance that is determined by the researcher to find out whether the variable is significant or not.

The formula for the (Wald) statistic is as follows:

$$wald = \left(\frac{\hat{b}_i}{S.E(\hat{b}_i)}\right)^2 \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (15)$$

Where:

$\hat{b}$:Represent parameter estimator
$SE\ (\hat{b})$: standard error of the parameter

We must mention that the value of the (Wald) statistic is questionable in the event that the absolute value (of the regression coefficient is large) as the value of the standard error is very large, which results in a small value for the statistic and thus leads to the result that the test for the variable in question is not significant, and the problem will be The same in the case of small-sized samples because the standard error is large and leads to the same problem mentioned above, as we will test through (Wald) statistics the following hypothesis:

$H_0: b_{0j} = 0$
$H_1: b_{0j} \neq 0$
i = 1.2, 5 and b_i refer to regression coefficients of independent variables

5.2 Score Test [7]

The score test, which is also known as the Lagrange Multiplier Test, is a statistical test that tests the parameters in the framework of the null hypothesis $(H_0: \beta = 0)$, as this test is one of the most powerful tests when the value of the parameters approaches the value of (B_0). The main feature of this test is It is his lack of need to estimate information under the alternative hypothesis ($H_0: \beta \neq 0$), as the statistic of this test approaches the chi-square distribution (χ_p^2)and is compared with its tabular value.

5.3 Tests of conformity to the form: -

The quality of fit means how well the statistical model is suitable for the study sample data. The measures of the quality of fit measure the convergence between the observed and expected values of the model. The following are some important tests for the quality of fit.

5.4 The Hosmer – Lemeshow Test [8],[11]

To measure the Goodness of Fit for a logistic regression model, the Hosmer-Lemeshow test is used. This test is used widely to assess the goodness of fit for the model. It allows any number of explanatory variables, which may be continuous or discontinuous. This test is somewhat similar to the (χ^2)test for good fit. This test groups the sample cases based on the expected probability values. Hosmer-Limeshow has suggested using one of two collection strategies in this test:

A- Grouping of cases based on percentages of expected probabilities.
B- Grouping cases based on fixed values of expected probabilities.

The observed and expected values of the cases are summed according to the two values of the dependent variable Y (0 ,1), as well as in each of the ten groups, after which the Hosmer – Limeshow statistic denoted by the symbol **H** is calculated according to the chi-square statistic (χ^2) of the table (g*2) For the observed and expected frequencies, as the **H** statistic follows the chi-square distribution with degrees of freedom equal to (g-2).

6. Classification table [3],[4],[5]

It is a method used to summarize the results of the logistic regression or to clarify the observed and expected responses to the dependent variable, as the rows represent the observed responses and the columns represent the expected responses, as these tables try to estimate the percentages of accuracy of prediction for the response variable (adopted) in both classes, as well as the overall ratio and control relative to the sample size and for each A step in building the model, a special classification table is created for that model. The accuracy of this table depends on the Sensitivity scale, which is the correct prediction ratio for the response variable class (the dependent) in which the event appears, and the specificity scale, which is the correct prediction ratio for the response variable class (the dependent) In which the watch appears.

Application aspect

Introduction:

The data were collected from the Central Bank of Information [12], and it is a sample that includes two types of country living conditions (poor / rich), which represent the adopted variable.
The study sample included 50 countries for some countries of the world, and the following variables were studied for each country.
(Income, cost of living, cost of education, unemployment rate, and geographical location) and the SPSS statistical program was used to extract the results.

Defining variables

We have relied on a number of variables and since part of the variables are discrete variables and others are related.
A- The intentional variable (Y) which represents (the country's living situation):
Poor = 0 Rich = 1
B- The explanatory variables represent the following:
(X_1 = income, X_2 = cost of living, X_3 = cost of education, X_4 = unemployment rate,
X_5 = Geolocation: We have given codes for each geographical location to facilitate data analysis (Asia = 1, Africa = 0))

Table (1) A table representing the sample size and the missing data

Case Processing Summary			
Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	50	100.0
	Missing Cases	0	.0
	Total	50	100.0
Unselected Cases		0	.0
Total		50	100.0
a. If weight is in effect, see classification table for the total number of cases.			

Table No. (1) summarizes the data entered in the analysis, the size of the studied sample, and the missing data

Table (2) encoding the adopted variable

Dependent Variable Encoding	
Original Value	Internal Value
Poor	0
Rich	1

Table No. (2) represents the code and symbols for the values of the dependent variable.

Table (3) **Iteration History**[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients					
			Constant	x_1	x_2	x_3	x_4	$x_5(1)$
Step 1	1	48.770	-1.773-	.000	-.017-	.126	-.007-	1.816
	2	47.181	-3.507-	.000	-.012-	.209	-.017-	2.163
	3	46.997	-4.508-	.000	-.005-	.254	-.028-	2.262
	4	46.993	-4.677-	.000	-.004-	.261	-.030-	2.280
	5	46.993	-4.681-	.000	-.004-	.261	-.030-	2.280
	6	46.993	-4.681-	.000	-.004-	.261	-.030-	2.280
a. Method: Enter								

b. Constant is included in the model.
c. Initial -2 Log Likelihood: 68.029
d. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Table No. (3) shows the number of iterations of the derivatives of the greatest probability function to obtain the lowest value of the negative double of the logarithm of the greatest possibility function to obtain the optimal estimate of the model parameters for the derivative of negative twice the greatest possibility function.
In the sixth session of the derivative of a negative factor twice the greatest probability function, we got its lowest value, which is equal to (46.993), meaning $-2\ Log\ likelihood\ =\ 46.993$, and we stopped at this cycle because the change in the coefficients became less than 0.001. In fact, the change in the estimated parameters became slow. Very after the third cycle, and as we note this from Table No. (3), therefore it can be said: The capabilities of the features in the courses (4,5,6) are similar with very small differences and we stopped at the sixth session and considered its features the best result that can be obtained for the milestones as Negatively double the logarithm of the greatest possibility function is at its lower end at this cycle,
Table No. (4) summarizes the parameters of the ideal model that we obtained in the sixth session of Table No. (3).

Table (4) illustrates the variables included in the model

Variables in the Equation									
		B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**95% C.I.for EXP(B)**	
								Lower	**Upper**
Step 1[a]	x1	.000	.000	2.051	1	.152	1.000	1.000	1.000
	x2	-.004-	.048	.007	1	.931	.996	.906	1.095
	x3	.261	.245	1.142	1	.285	1.299	.804	2.098
	x4	-.030-	.073	.171	1	.679	.970	.842	1.119
	x5(1)	2.280	.801	8.102	1	.004	9.778	2.034	47.004
	Constant	-4.681-	4.758	.968	1	.325	.009		
a. Variable(s) entered on step 1: x1, x2, x3, x4, x5.									

Depending on the parameters of the model in the above table, the logistic regression equation can be written. Table No. (4) includes all the parameters of the estimated model ((constant, b_1, .., b_5) and the standard error for each parameter.
And a (wald) statistic for each of the model's features, the number of degrees of freedom, and the significance of the features
We notice that column B contains the parameters of the attached form, which are in Log-odds, and the form equation is as in the following figure:

$$log\left(\frac{\hat{P}}{1-\hat{P}}\right) = -4.681 - 0.004X_2 + 0.261X_3 - 0.030X_4 + 2.280X_5$$

Since $\hat{P}$ is the probability that the country is rich for new decisions, and these estimates indicate the relationship between the independent variables and the dependent variable in units of (logit). The second column represents the standard error of the transactions S.E according to the relationship

S.E (*Bˆi*) ??*hii*

The third column represents the wald statistic to test the significance of the transactions according to the relationship

$$wald = \left(\frac{\hat{b}_i}{S.E(\hat{b}_i)}\right)^2$$

It follows the distribution of χ ^ 2 with the degree of freedom d.f = 1, as we can see from Table (4)

$$wald\ for\ var2 = \left(\frac{-0.004}{0.048}\right)^2 = 0.007$$

As for Sig, it is a column that represents the significance of the coefficients corresponding to acceptance or rejection of the hypothesis H0 using the probabilities, since at α = 0.05 and Sig <0.05, the hypothesis H0 is rejected, meaning that the parameter is significant and it is not equal to zero in the population from which the sample is drawn, while the column Exp (B) is explained The value of the exponential function of the regression coefficient, and it expresses the multiplier with which the odds ratio changes [the probability of occurrence of the event p (y) the odds ratio [the probability of the occurrence of the event p (y) to the probability of not occurring -1 p (y) and for the second line in the table

$$Exp\ (-0.004) = e^{-0.004} = 0.996\ Odds\ Ratio$$

As for the last column, it represents the limits of confidence C.I for Exp (B). Returning to the interpretation of the regression coefficients in Table (4), we find the following:
X5 (the geographic location of the country) ranked first in influencing (the dependent variable, the state of the country Y) as the regression coefficient of this variable b_5 = 2.280, and this parameter showed a high Significance of the dependent variable at the level of significance α <0.001 with a degree of freedom df = 1 and the wald statistic statistic = 8.102 and the standard error is 0.801
And in the second place came the variable X3 (the cost of educational education) in terms of importance in the effect Y, as the regression coefficient of this variable b1 = 0.261, and it is explained that the change of the value of b1 = 0.261
As for the rest of the variables X1, X2, X4 (income, cost of living and unemployment rate), they were not significant in affecting the variable Y.
As for testing the adequacy of the model in full and its quality (Goodness of fit), we used the statistic (F, R^2) in linear regression, but in the case of the logistic model, the greatest likelihood ratio (log likelihood ratio) is used, which follows the distribution of χ^2 according to the relationship: $\chi^2 = 2(log_e\ L_0\ -\ log_e\ L_1)$

Table(5)

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	21.036	5	.001
	Block	21.036	5	.001
	Model	21.036	5	.001

And that the value of χ^2 = 21.036 in the above table, which is significant at the level of significance of α = 0.05, as Sig. = 0.001, which confirms the significance of the whole model as shown in Table (5) with a degree of freedom 5.

Table (6)

Contingency Table for Hosmer and Lemeshow Test						
		y = poor		y = rich		Total
		Observed	Expected	Observed	Expected	
Step 1	1	5	4.810	0	.190	5
	2	5	4.484	0	.516	5
	3	1	4.336	4	.664	5
	4	5	4.130	0	.870	5
	5	5	3.516	0	1.484	5
	6	3	2.565	2	2.435	5
	7	3	2.023	2	2.977	5
	8	2	1.638	3	3.362	5
	9	0	1.141	5	3.859	5
	10	0	.358	5	4.642	5

As for the table (6), it is also a non-parameter test of the quality of model fitment, as it is based on computing the χ^2 statistic for the difference between observed and expected values. Hosmer and lemeshow have suggested using the χ^2 distribution to detect deviations of the logistic model. The statistic for this test consists of an observed part not based on a theoretical model and the other (Expected) calculated from the logistic model estimates, and the χ^2 statistic is calculated for the contingency table.

Table (7)

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	7.465	8	.487

The model significance test shows after entering the (independent) explanatory variables of the model to include the fixed limit and explanatory variables using Hosmer and Lemeshow's test. 0.05) that the model is appropriate when introducing the explanatory variables and the fixed term, and this indicates the good fit for the model, as according to Hosmer and Lemeshow's hypothesis, if the observed values converge with the predicted values, this means that the model represents the data well according to the following hypothesis: -

H_0: The observed values are close or equal to the expected values

H_1: The observed values do not converge or equal with the expected values

Table (8)

Classification Table[a]					
	Observed		Predicted		
			y		Percentage Correct
			0	1	
Step 1	y	0	24	5	82.8
		1	4	17	81.0
	Overall Percentage				82.0
a. The cut value is .500					

Table (8) shows the percentage of correct classification of observations of the adopted variable for the first step after entering the independent variables of the model. Watching is only wrongly classified and that the probability of total error is around (0.18), meaning that the error rate is (18), which is a good rate.

Conclusions:

1. The results of the study showed that using a two-response logistic regression gave logical results behind the studied phenomenon.
2. The null hypothesis which states $H_0 : b_1 = 0$ and the alternative hypothesis $H_1 : b_2 = 0$ was rejected, as the most explanatory variables appeared (income, cost of living, cost of education, unemployment rate) were not significant and did not affect the dependent variable As for the geographical location variable, it appeared to be significant, meaning that it affects the dependent variable
3. The percentage of correct classification of the observations of the adopted variable appeared for the first step after entering the independent variables of the model. Up to (0.18), meaning that the error rate is (18), which is a good rate.

Recommendations:

Based on the above, and in light of the study, the researchers suggest the following recommendations:

1. Expanding the use of binary logistic regression in economic and social studies, as its previous uses were limited to the medical and educational sciences.
2. Those in charge of the Central Bureau of Statistics must use advanced statistical methods to study household income and expenditures and other important studies.
3. Conducting continuous studies on income at the state level, especially with the emergence of recent economic and social variables, by introducing more factors affecting income that affect the determination of the socio-economic fabric

No.	Country	Annul Person Income ($)	Life expectancy (Years)	School life expectancy (Years).	Unemployment rate	Continental location (Asia=1 , Africa=0)	Economic status (poor=0 , rich=1)
		X1	X2	X3	X4	Cont.	Y
1	Algeria	3,974.0	72	13	8.1	1	0
2	Bahrain	23,504.0	75	13	5.6	0	1
3	Djibouti	3,414.9	57	6	54.6	1	0
4	Egypt	3,019.2	72	12	4.9	1	0
5	Iraq	5,955.1	68	12	16.2	0	0
6	Jordan	4,405.5	72	12	11	0	0
7	Kuwait	32000.5	74	13	2	0	1
8	Lebanon	7,583.7	71	13	8.6	0	0
9	Libya	7,685.9	73	16	7.6	1	0
10	Mauritania	1,679.4	57	8	23.9	1	0
11	Morocco	3,204.1	70	11	8.4	1	0
12	Oman	15,343.1	71	13	1.9	0	1
13	Qatar	62,088.2	79	12	0.2	0	1
14	Saudi Arabia	23,139.8	73	14	3.5	0	1
15	Somalia	1,26.9	50	3	26.1	1	0
16	Sudan	4,41.5	60	10	18.7	1	0
17	Syria	2,032.6	74	12	5.7	0	0
18	Tunisia	3,317.5	73	14	11.9	1	0
19	United Emirates	43,103.3	76	11	2	0	1
20	Yemen	774.3	65	11	12.4	0	`0
21	Cambodia	14,478	70	12	10.5	0	1
22	Colombia	47,551	50	11	10	1	0
23	Cuba	11,249	70	13	12	1	0
24	Hondrus	7,912	60	14	11	0	1
25	American	77.4	80	12	30	0	0
26	Angola	36.6	76	12	11	1	0
27	Anguilla	44.7	53	13	12	1	1
28	Argentina	13.46	81	11	13	0	0
29	Armenia	43.03	80	12	12	0	1
30	Aruba	56.3	77	14	10	1	0
31	Australia[1]	54.7	78	16	15	0	1
32	Austria	657.1	84	15	14	1	0
33	Azerbaijan[2]	65.7	84	12	11	0	1
34	Bahamas	8.99	74	11	12	1	1
35	Bahrain	12.7	79	13	13	1	0
36	Finland [8]	34.6	72	14	14	0	1
37	France	70.5	83	12	15	1	0
38	Guiana	76.8	85	12	16	0	1
39	cuba	87.8	81	12	17	1	0
40	Gabon	45.8	78	15	15	0	1
41	Gambia	32.9	64	16	14	1	0
42	Georgia [9]	44.8	60	13	13	0	1
43	Germany	87.9	77	12	12	1	0

44	**Ghana**	**45.7**	**83**	**13**	**23**	**0**	**1**
45	**Gibraltar**	**43.8**	**66**	**14**	**12.5**	**1**	**0**
46	**Greece**	**87.5**	**83**	**11**	**23**	**0**	**0**
47	**Honduras**	**87.9**	**83**	**12**	**12**	**0**	**1**
48	**Hungary**	**45**	**70**	**12**	**14**	**1**	**1**
49	**Iceland**	**80.5**	**80**	**16**	**11**	**1**	**0**
50	**india**	**43.8**	**65**	**12**	**12**	**1**	**1**

The References

1. Anderson, J.A., "Separate Sample Logistic Discrimination", Biometrika, 59, 19-35, (1972).
2. Burns, Robert and Burns, Richard, "Business Research Methods and Statistics uAsing SPSS", Five extra advanced chapters, chapter 24 Logistic Regression: pp 568 -575, (2008).
3. Dallal, Gerard E. (2001). Logistic Regression. Available at: www.tufts.edu/~gdallal/logistic.htm
4. David G. and Kupper, L., "Applied Regression Analysis and other Multivariate Methods", The University of North Carolind and Chapel Hill, (1978).
5. Ferrer, Alvaro J. Arce and Wang, Lin (1999). Comparing the Classification Accuracy among Nonparametric, Parametric Discriminant Analysis and Logistic Regression Methods. Paper Presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23,1999).
6. Hardle, W. and Simar, L., "Applied Multivariate Statistical Analysis", Berlin and Louvain-la-Neuve, Germany, p332, (2003). 48
7. He, X. and Fung, W.K., "High breakdown estimation for multiple with applications to discriminant analysis", Journal of Multivariate analysis, 72, 151-162, (2000).
8. Hosmer, D., Lemeshow, S. & Sturdivant , R. ,(2013)," Applied Logistic Regression", 3rd edition ,New York: willey,WSIPS, http : // ihmsi.org.
9. kleinbaum , D.G, Kupper , L.L. and muller. k. E. (1988): “Applied Regression Analysis and other Multibariable methods” DWS- KENT publishing company , a division of wadsworth P 317
10. McLachlan, G. J., "Discriminant Analysis and Statistical Pattern Recognition", New York: Wiley, (1992).
11. Schlotzhauer, D.C, "Some issues in using proc logistic for Binary logistic regression", the technical for SAS software users vol. No. 4. (1993).
12. https://unstats.un.org/unsd/demographic/products/socind/default.htm.