

Use of Binary Logistic Regression Technique to Study some of the Factors Affecting Women's Use of Family Planning Methods in Iraq

Dr. Worood Basim Noor ^a

^a The University of Mashreq, Baghdad, Iraq. Email: Wurood.basim@uom.edu.iq

Abstract

This study dealt with employing the technique of binary logistic regression to study some of the factors affecting women's use of family planning methods in Iraq, represented by (the current age of the woman, the age of the woman at the first marriage, the number of children alive, the number of girls, the number of children who died, the educational level, the level of living, Functional difficulties, geographical area, housing environment (urban, rural)) through employing a dual logistic regression model based on the data of the Multiple Indicator Cluster Survey in Iraq for the year (2018) that was implemented by the Ministry of Planning / Central Bureau of Statistics and the Kurdistan Region Statistics Authority in cooperation With the Iraqi Ministry of Health as part of the Global Multiple Indicator Cluster Survey Program As the survey covered all regions of Iraq (central, southern and northern Iraq and the Kurdistan region), The results of the analysis showed that the use of family planning methods by married women aged (15-49) years is affected by a number of factors (the number of live children, the number of girls, the geographical area, the age of the woman at the first marriage, the educational level of the woman), and the analysis did not show any significant or effect. For the variables (the woman's current age, the number of children who died, the level of living, job difficulties, the housing environment).

Keywords: Binary Logistic Regression, Family Planning Methods, Odds Ratio.

Introduction

The Binary Logistic Regression Model is one of the nonlinear regression models that deal with binary response data and whose prediction value either takes the value (1) or takes the value (0), and this is not found in the linear regression model where the value is the predictor. It is between $(-\infty, \infty)$, and the first to use this model is the scientist (Verhulst) [Shaefer, R.L 1979] who used it to describe the growth of society and called it the growth function, and in 1920 the researchers [Gramer, J.S. (1991)] (Pearl & Reed) used it in calculating population growth and launching it as the logistic function. The importance of this model is evident in the possibility of using it in the analysis of many phenomena.

Research Methodology

The research relied on data obtained from (Multiple Indicator Cluster Survey 2018 in Iraq), which was implemented by the Central Bureau of Statistics (Ministry of Planning) and the Kurdistan Regional Authority in cooperation with the Iraqi Ministry of Health as part of the Global Multiple Indicator Cluster Survey Program. The determinants of family planning methods included [the current age of the woman, the age of the woman at the first marriage, the number of children alive, the number of girls, the number of children who died, the educational level, the level of living, difficulties, the geographical area, the housing environment (urban, rural)] As for the dependent variable in the study, family planning methods, and it may include two types of response (non-use of family planning methods, use of family planning methods). Where a binary logistic regression model was employed in analyzing the data using the statistical program SPSS version 24.

The Importance of Studying

The freedom of women to have children and to determine the size of the family by limiting the number of children is a basic human right recognized in the international conventions ratified by Iraq. Where family planning is one of the main pillars of reproductive health services that improve women's health, and since the role of women in Iraq is of great importance through their participation in many political, economic and social fields, Accordingly, it is necessary to provide facilities to them through improving health services, especially those related to the use of family planning methods. Reducing the incidence of unwanted childbearing helps women reach their reproductive desires, which in turn is reflected in the optimal use of the resources available to the state. As some studies indicate that childbearing has a role in every country as a whole, and partial effects on the family and woman level. As the change in fertility rates at the state level leads to a change in the age structure of the population, while the individual effect comes through helping the family and the woman to achieve reproductive desires. Hence the importance of research in studying a number of determinants that affect women in using family planning methods in Iraq.

Research Objective

The main objective of this research is to identify a number of factors that are believed to affect the use of family planning methods among married women whose ages range between (15-49) years, in other words, to identify the most important determinants affecting the prevalence of family planning use. Among Iraqi women, by using a binary logistic regression model.

The Theoretical Side

Logistic Regression Model Binary

The idea model of logistic regression is based on principle that the dependent variable (y) is the response variable that we are interested in studying. It's a binary variable that goes like this: the Bernoulli distribution, and takes (1) with a probability of (π) and (0) with a probability of $(1-\pi)$ i.e. the occurrence of the response or its non-occurrence, and as is known in the simple The model that links the variables is as follows: linear regression model with continuous explanatory factors and response variable.

$$y = \beta_0 + \beta_1 + \epsilon \quad (1)$$

Since (y) represents a continuous variable, and assuming that the average of the observed or actual (y) values at a certain value of the variable (x) is $E(y)$, the model can be rewritten as follows:

$$E\left(\frac{y}{x}\right) = \beta_0 + \beta_1 x \quad (2)$$

The right-hand side of a regression model is known to take values from $(-\infty)$ to (∞) , however when we have two variables, one of which is binary (y), the simple linear regression model is not appropriate because:

$$E\left(\frac{y}{x}\right) = P_r(y = 1) = \pi \quad (3)$$

Thus, the value of the right side is confined between the two peaks [1,0] and therefore the model is not applicable from the point of view of the regression. And one of the ways to solve this problem is to introduce a mathematical transformation on the dependent variable (y), and it is known that $(0 \leq \pi \leq 1)$ and then the ratio $\left(\frac{\pi}{1-\pi}\right)$ is a positive expression between $(0 \text{ and } \infty)$, meaning that $\left[0 \leq \left(\frac{\pi}{1-\pi}\right) \leq \infty\right]$ by taking the natural logarithm to the base (e) for the transformation $\left(\frac{\pi}{1-\pi}\right)$ then the range of its values becomes confined between $[-\infty \leq \log_e\left(\frac{\pi}{1-\pi}\right) \leq \infty]$, and therefore, the regression model can be rewritten in the case of one independent variable as follows:

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

The ratio $\left(\frac{\pi}{1-\pi}\right)$ is called the odds of success, and the expression $\log_e\left(\frac{\pi}{1-\pi}\right)$ is called the log odds ratio or the logit. (El-Din, & Maher (2020)) (Sari, & Muhammad. (2017)).

And formula No. (Khalaf (2019)) represents the binary logistic regression model in a linear form, in terms of what is known as (logit).

We note from Figure (1) that the logistic function is a continuous function whose range ranges between (1,0) as it approaches zero as the right side of it approaches $(-\infty)$ and approaches one as this side approaches $(+\infty)$.

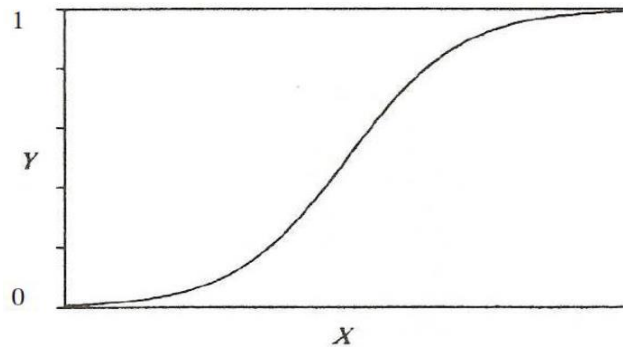


Figure 1: Explain the Relationship between Two Logistic Regression Variables

Significance Tests of Model Parameters

These tests are conducted for the purpose of knowing the statistical significance of each of the independent variables separately. The Wald test or the adds ratio is used as follows:

Wald

Wald Statistics is used to test the statistical significance of all logistic regression coefficients. Wald test tests the null hypothesis which states that the independent variable (X) has a logistic regression coefficient of zero, meaning that hypothesis $H_0: b = 0$

Wald's statistic is calculated as follows:

$$W = \frac{b}{S.E} \quad (5)$$

Where as

b: indicates the value of the independent variable's logistic regression coefficient (X).

S.E: indicates the magnitude of the independent variable's standard error of the logistic regression coefficient (X).

Where Wald's statistics follow a distribution Z (Hosmer & Lemshow, 2000).

If Wald statistics are the null hypothesis that the value of the regression coefficient is equal to is rejected if the regression coefficient is statistically significant (0), meaning that the independent variable (X) has an effect on the dependent variable's value (Y). In the case that Wald's statistic is not Statistically significant, the value When the regression coefficient is equal to zero, it means that the variable (X) has no effect on the process of predicting the dependent variable (Y), and it can be removed from the model. (Lemeshow, & Sturdivant. (2000)).

Odds Ratio

It is a way of expressing the probability of an event occurring divided by the probability of not occurring, and it is often expressed as a ratio between the two quantities $\left(\frac{\pi}{1-\pi}\right)$. (Lemeshow, & Sturdivant. (2000)) (Qasim. R. (2011)).

Hosmer and Lemeshow Test

The (Hosmer & Lemeshow) test is used for the purpose of knowing the suitability of the binary logistic regression model for the data, as it depends on the collection of cases based on "fixed values of the expected probabilities", which is done by collecting the expected and observed values based on the two values of the dependent variable (Y=0, Y =1) for each of the classification groups using the following formula:

$$\hat{C} = \sum \sum \frac{(O-E)^2}{E^2} \quad (6)$$

Where as:

O: represents the observed values

E: represents the expected values

(Lemeshow, & Sturdivant. (2000))

Classification Table

The classification table is one of the methods used in the quality of matching the model with the data. The classification table depends on clarifying the number of cases of the desired trait as well as the cases of the undesirable trait that were classified correctly or incorrectly. The classification table proves that the model matches the observed data or does not match it. The higher the correct classification ratio, the more this indicates that the model has a good predictive ability, and it should be noted here that these tables are used for predictive rather than explanatory purposes.

Table 1: Classification Table

category		expected		total
observed	P (positive)	correct	incorrect	P
observed	N (negative)	incorrect	correct	P
Total		Q	Q	1

Assumptions of Logistic Regression

1. The dependent variable (y) represents a binary descriptive variable and its conditional mean is $E(Y/X)$, which is a variable with a period of [0,1]. The other independent variables are either descriptive (binary or multiple) or quantitative (continuous or discontinuous) variables.
2. The relationship between the dependent variable (y) and the other independent variables is nonlinear.
3. The error term follows a binomial distribution with expectation [0] and variance equal to $P(X)[1-P(X)]$.
4. There is no autocorrelation between the limits of random errors, that is, $E [e_i, e_j]$.
5. There is no correlation between the random error term and the independent variables, meaning that $E [e_i, x_i]$. (Park, H. (2013)).

Characteristics of the Logistic Regression Model

1. The regression model is more of a tool because it is a test of the coefficients function.
2. It gives an idea of the effect of each independent variable on the dependent variable (the binary response variable).
3. It arranges the effect of variables in order of importance, which allows us to conclude that one variable is stronger than another.
4. The least sensitivity towards deviations from the normal distribution of the study variables.
5. It is considered the best method in case the dependent variable is binary (Y).
6. The possibility method (ML) is used to estimate the parameters of the model, which provides the quality conditions for the estimates. (Sari, & Muhammad (2017)).

The Parameters of the Logistic Regression Model are estimated

The Maximum Likelihood Estimation approach was used to estimate the parameters of the binary logistic regression model, which is a method that uses iterative methods to solve differential rates numerically.

It should be noted here, however, that the method of greatest possibility is considered one of the iterative methods, and it requires a lot of calculations, so it was not used on a large scale until the emergence of statistical packages (SPSS) programs, which helped a lot in estimating the parameters of binary logistic regression models.

In order to illustrate the method of the greatest possibility, it is necessary to know the shape of the Likelihood Function in the logistic regression model:

$$L = \prod_{i=1}^{m_1} p(x_i) \prod_{i=m_1+1}^n [1 - p(x_i)] \quad (7)$$

Where as:

L: represents the likelihood function.

p (xi): represents the probability of the adjective.

[1-p(xi)]: is the probability that the adjective will not appear.

n: represents the sample size (total number of cases.

m_1 : represents the number of cases representing the adjective.

And the mathematical formula of the function is the greatest possibility in the event that the data is binary, it can be written as in the equation.

$$L[\beta_i] = \prod_{i=1}^n p_i^{y_i} (1 - p_i^{y_i}) \quad (8)$$

We substitute the value of (pi) with its equivalent in equation (8).

$$p_i = \frac{\exp(y)}{1 + (\exp(y))}$$

Using the natural logarithm method for both sides of the equation, we get:

$$\ln[L(\beta_i)] = \sum_{i=1}^n y_i(\beta_0 + \beta_1 x_i + \dots + \beta_k x_k) - \sum_{i=1}^n \ln(1 + e^{(\beta_0 + \beta_1 x_i + \dots + \beta_k x_k)}) \quad (9)$$

In order to get the estimates of the parameters of the model, we derive the function for the parameters and make them equal to zero, so we get (k + 1) from the natural equations, and because these equations are non-linear and can only be solved using iterative methods, including the (Newton Raphson) iterative method that gives appropriate estimates for the parameters. (Park, H. (2013)).

The Quality Criteria of the Fitment of the Model

The criteria for working the explanatory criteria of the model, these criteria are considered as alternatives to the coefficient (R^2) of the linear regression, as it is called in a special semi-commercial way, regression, and the most famous of these coefficients are (R-deux de Nagelkerke, R-deux de Cox & Snell). (Namik. (2015)).

Application Side

The research sample was extracted from the Multiple Indicator Cluster Survey carried out by the Central Statistical Organization (Ministry of Planning) with the Iraqi Ministry of Health and the Kurdistan Regional Statistics Authority and in cooperation with the Global Program for the Multiple Indicator Cluster Survey, which is an important source in the study of family planning methods, as this survey included All of Iraq (central and southern governorates, Kurdistan Region) (urban and rural), where the sample size selected from the survey data was (7745) married women in the age group (15-49) years.

Search Variables

The dependent variable (y) is a nominal variable of the binary type that takes the value (1) The woman uses family planning methods and takes the value (0) The woman does not use family planning methods.

X1: represents the current age of the woman and takes the values (class 15-25 = 1, class 26-35 = 2, class 36-45 = 3).

X2: represents the number of dead children of the woman and takes the values (category 0 = 1, class 1-2 = 2, class 3-6 = 3).

X3: represents the number of living children for the woman and takes the values (category 1-2 = 1, class 3-4 = 2, a class 5 and more = 3).

X4: represents the number of girls for a woman and takes the values (category 0 = 1, class 1-4 = 2, class 5-9 = 3).

X5: Geographical area (inhabiting the Kurdistan region = 0, inhabiting central or southern Iraq = 1).

X6: represents the age of the woman at first marriage and takes the values (category 15-25 = 1, class 26-35 = 2, class 36 and more = 3).

X7: represents the standard of living for women and takes the values (poor = 0, average = 1, rich = 2).

X8: Represents the educational level of women and takes the values (illiteracy = 0, primary = 1, intermediate = 2, middle school and above = 4).

X9: represents the functional difficulties of the woman and takes the values (no functional difficulties = 0, functional difficulties = 1).

X10: represents the living environment and takes the values (urban = 0, rural = 1) Independence variables.

It was considered that the last category is the reference category for each of the variables when conducting the analysis, which was used as the statistical program SPSS (version V24) for the purpose of data analysis and as follows:

Overall Evaluation of the Model

For the aim of determining the model's explanatory power, The findings of a model with only the fixed term (and no explanatory factors) are compared to those of a model with explanatory variables, as we note from Table (2) that the value of the function of the greatest possibility has stabilized at the value (8922.042) at the third iteration. And that this value will be compared with the maximum possibility function of the model that includes all the independent variables.

Table 2: Frequencies and the Value of the Logarithm of a Function and the Possibility without the Explanatory Variables (the Model includes the Fixed Term only)

Iterations	-2 Log likelihood	Coefficients Constant
1	8922.532	0.586
2	8922.042	0.603
3	8922.042	0.603

Table 3: Frequencies and the Value of the Logarithm of the Maximum Potential Function of the Model including the Independent Variables

Coefficients																				
Iterations	-2 Log likelihood	Constant	X1 ₁	X1 ₂	X2 ₃	X2 ₁	X2 ₂	X3 ₁	X3 ₂	X4 ₁	X4 ₂	X5 ₁	X5 ₂	X6 ₁	X6 ₂	X6 ₃	X7	X8	X9	X10
1	8240.340	.278	.087	.104	1.008	.867	-.616	.398	-.722	-.508	.658	.006	-.628	.008	-.032	-.575	-.350	-.233	-.001	.093
2	8223.619	.349	.118	.126	1.113	.945	-.620	.471	-.863	-.609	.822	.008	-.707	.011	-.040	-.689	-.419	-.282	-.005	.111
3	8223.579	.353	.120	.128	1.117	.948	-.621	.474	-.869	-.613	.833	.008	-.710	.011	-.040	-.694	-.422	-.285	-.006	.112
4	8223.579	.353	.120	.128	1.117	.948	-.621	.474	-.869	-.613	.833	.008	-.710	.011	-.040	-.694	-.422	-.285	-.006	.112

Table (3) shows the number of iterative cycles of the derivative of the logarithm of the greatest possibility function in order to obtain the lowest value of the logarithm of the greatest possibility function, which gives the optimal estimates for the parameters of the model. The change in parameter values became less than (0.001). In addition, the value of the derivative of the maximum possibility function is much less than its value in the case of the model, which includes only the fixed term, which is equal to (8922.042), and from it we conclude that there is a relationship between the independent variables and the dependent variable (Y), and the estimates obtained in the fourth cycle are the best possible estimates for the parameters Binary logistic regression model.

Match Quality Standards

They are statistical criteria used to measure the explanatory power of the model, and they are considered alternatives to the coefficient of determination (R^2) of the linear regression model, which is calculated by comparing the explanatory power of the model without explanatory variables with the model with the presence of explanatory variables, and the most famous of these coefficients are (Cox & Snell R Square: Nagelkerke R Square), where we

notice from Table (4) that the coefficients are (0.09, 0.13) and from it we conclude that the model that includes the explanatory variables is compatible with the data.

Table 4: Quality of Model Fit Using the Logarithm of the Likelihood Function

-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
8224.029	0.097	0.133

Table No. (5) shows the results of the (Hosmer and Lemeshow Test), which is used to determine the existence of significant differences between the observed and expected values, which uses the Chi-square statistic, where we note that the probabilistic values of the test are equal to 0.508), which is greater than the level of Significance is 0.05. Therefore, we accept the null hypothesis, which states that there are no significant differences between the observed and expected values, as we note that there is a slight difference between the observed and expected values, and from it we conclude that there is agreement between the model and the data, in other words, the model represents the data well. Hosmer and Lemeshow test.

Table 5: Hosmer and Lemeshow Test

	Do not use family planning		Use family planning methods		Total
	Observed	Expected	Observed	Expected	
1	474	462.451	182	193.549	656
2	356	351.487	334	338.513	690
3	269	269.909	373	372.091	642
4	252	256.550	435	430.450	687
5	224	237.751	482	468.249	706
6	200	218.931	499	480.069	699
7	201	196.860	489	493.140	690
8	152	156.120	463	458.880	615
9	166	153.295	532	544.705	698
10	134	124.646	650	659.354	784
Chi-square		df	Sig.		
7.264		8	0.508		

Estimate the Parameters of a Binary Logistic Regression Model

Table No. (6) shows the results of the binary logistic regression model's parameters, which include parameter estimation, standard error and (wald) statistics, and where we note the significance of the following variables (number of girls, number of live children, geographical area, age of women at first marriage, educational level of women) as for the rest The results of the other variables were not significant. The effect of each of the significant variables is shown below:

- With regard to the variable (number of girls), we note the probability that a woman uses family planning methods and has no daughters is (0.419) times lower compared to the woman who uses family planning methods and has (5-9) girls, as well as we note the possibility that women use family planning methods and then (1-4) Girls are (0.542) times less than women who use family planning methods and have (5-9) girls.
- Regarding the variable (geographical area), we note the probability that a woman uses family planning methods and lives in the Kurdistan region is (1,299) times higher than that of a woman who uses family planning methods and lives in central and southern Iraq.
- Regarding the variable (the woman's age at the first marriage), we note the possibility that the woman uses family planning methods, and her age at the first marriage ranges between (26-35) times less than (0.491) times than the woman who uses family planning methods, and her age at the first marriage ranges (36-36).
- With regard to the variable (educational level), we note the probability that a woman uses family planning methods and her level of illiteracy is (0.499) times lower than that of a woman who uses family planning methods. Her educational level is preparatory and above, and we also note the possibility that a woman uses methods of planning and her primary education level is less (0.656) times for the woman who uses family planning methods. Her educational level is preparatory and above. We also note that the woman uses family planning methods and her educational level is average (0.752) times lower than the woman who uses family planning methods. Her educational level is middle and above.

- With regard to the variable (number of living children), we note the probability that a woman uses family planning methods and has children whose number ranges between (0-2) children, 0.538 times less than the woman who uses family planning methods. She has children whose number ranges between (6-8) children.

Table 6: Estimates of the Parameters of the Maximum Possible Function

Variables in the Equation						
Variables	B	S.E.	Wald	df	Sig.	Exp(B)
The woman's current age			4.049	2	0.132	
15-25 years old woman's current age	0.120	0.090	1.777	1	0.183	1.127
Years (26-35) of the woman's current age	0.128	0.064	3.921	1	0.048	1.136
The number of deceased children			6.014	2	0.049	
Child (0) Number of children who died	1.117	0.673	2.759	1	0.097	3.056
Child (1-2), the number of children who died	0.948	0.677	1.961	1	0.161	2.580
The number of living children			138.210	2	0.000	
Child (1-2) the number of live children	-0.621	0.290	4.580	1	0.032	0.538
Child (3-4) the number of live children	0.474	0.270	3.070	1	0.080	1.606
number of girls			15.321	2	0.000	
Girl(0) Number of girls	-0.869	0.310	7.876	1	0.005	0.419
Girls (1-4) the number of girls	-0.613	0.299	4.208	1	0.040	0.542
(inhabits the Kurdistan Region) Region:	0.833	0.075	122.526	1	0.000	2.299
Woman's age at first marriage			17.048	2	0.000	
15-25 years old woman at first marriage	0.008	0.088	0.008	1	0.929	1.008
Years (26-35) of the woman's age at first marriage	-0.710	0.194	13.472	1	0.000	0.491
standard of living			0.451	2	0.798	
(poor) standard of living	0.011	0.073	0.022	1	0.881	1.011
(average) standard of living	-0.040	0.077	0.275	1	0.600	0.960
Scientific level			56.576	3	0.000	
(Umma) scientific level	-0.694	0.099	49.494	1	0.000	0.499
(Elementary) Academic level	-0.422	0.072	34.541	1	0.000	0.656
(Intermediate) Academic level	-0.285	0.079	12.852	1	0.000	0.752
(No difficulties) Functional difficulties	-0.006	0.125	0.002	1	0.965	0.994
(Urban) housing environment	0.112	0.067	2.789	1	0.095	1.118
Constant	0.353	0.757	0.217	1	0.641	1.423

Chi-square Test (χ^2)

We note from Table (7) the value of the (Chi-square) test equals 698.463 and that the probability value of the test is equal to (0.000) which is less than the level of significance. In the model, it has a statistically significant effect in the classification of women's use of family planning methods (use family planning methods or not).

Table 7

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	698.463	20	.000
	Block	698.463	20	.000
	Model	698.463	20	.000

Model Classification

We note from Table (8) that the ability of the binary logistic regression model to correctly classify the two groups of women who use family planning methods and those who do not use these methods, as the model was able to classify (641) women among those who do not use family planning methods correctly, with a percentage of (26.4%). At the same time, the model failed to classify (1787) women within their correct group, while (4116) women were correctly classified among those who use family planning methods at a rate of (92.7%), and this percentage is very good, and at the same time the model was classified (323). A woman was categorized incorrectly, and the general percentage of the model's ability to correctly categorize (using or not using family planning methods) was (69.3%), which is a good percentage that indicates the strength of the model and its ability to predict the values of the dependent variable.

Table 8: Classification Table for the Two Groups of Women who use Regulation Methods and those who do not use these Methods

observations	Do not use family planning	Use family planning methods	Correct rating ratio
Do not use family planning	641	1787	26.4
Use family planning methods	323	4116	92.7
Overall rating ratio			69.3

Conclusions

1. The variables that appeared to have an impact on women's use of family planning methods are (number of live children, number of girls, geographical area, age of the woman at first marriage, and the educational level of the woman).
2. The tests showed that the variables (current age of the woman, number of dead children, standard of living, functional difficulties, housing environment) were not significant and had no effect.
3. The correct classification rate for the total model of women who use family planning methods is (69.3%).
4. Women in Kurdistan use more family planning methods than women in central and southern Iraq.
5. The higher the educational level of women, the higher the use of family planning methods by women.
6. The possibility of building an appropriate data model that explains the most important factors that affect women's use of family planning methods.
7. The logistic regression model arranges the effect of variables according to importance, which helps the researcher to conclude that one variable is stronger than another.

Recommendations

1. Expanding the study of other variables, such as the economic and cultural variables affecting women's use of family planning methods.
2. Working to spread awareness and educate the importance of using family planning methods.
3. Attention to raising the scientific level of women, which has a role in increasing the use of family planning methods.
4. Studying the role of men in reproductive health and the extent of their participation in women's decision to use family planning methods.

References

- Faisal Naji Namik. (2015). "Using the two-response logistic regression technique and the multiple linear regression technique to diagnose the factors affecting the high repetition rates in higher education, a case study in one of the colleges of higher education", *Baghdad University Journal of Economic Sciences*. (43).
- Babiker, Anwar El-Din, & Faqiri, Maher (2020). "Using binary logistic regression to determine the most important health and medical factors that lead to women's deaths during pregnancy in Khartoum state during the year (2018)" *International Journal of Humanities and Social Sciences*, Issue (13), June 2020.
- Daish, Muhammad Amin, Sari, & Muhammad. (2017). Logistic regression model: concept, characteristics, applications" with an applied example of binomial logistic regression on SPSS program.
- Azar, Karim. Khalaf (2019), "Using multi-response logistic regression model to determine the factors affecting eye disease." *Magazine of college administration & economics for economic & administration & financial studies*, 11(3).
- David W., Hosmer, Lemeshow, S., & Rodney X., Sturdivant. (2000). *Applied logistic regression*. New York: Wiley.
- Gramer, J.S. (1991). *The Logit model: An Introduction for economists*, Edward Arnold, London.
- Park, H. (2013). An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing*, 43(2), 154-164.
- Qasim, B.A.R. (2011). Analyzing of impact of some variables of gingivitis by using the logistic regression. *Economic sciences*, 7(27).
- Shaefer, R.L 1979, "*Multicollinearty and logistic regression*", Ph.D. dissertation, university of Michigan, USA.