International Journal of Contemporary Mathematical Sciences
Vol. 15, 2020, no. 1, 1 - 10
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ijcms.2020.91228

Using Logistic Regression to Measure Unmet Need Levels in Some Rural Upper Egypt Governorates

H. E. Semary

Department of Statistics and Insurance, Faculty of Commerce
Zagazig University, Egypt

Abstract

The main objective of this paper is to measure the levels of unmet need in the rural areas of some Upper Egypt governorates, namely the governorates of Minya, Assiut and Sohag, also the research aims to identifying the most significant influential variables using the binary response logistic regression approach. The study sample included 4000 women. The results resulted in the entry of 7 independent variables in the expected model of a total of 29 variables representing demographic, social and economic characteristics at 5% significant level. The determination coefficient (Nagelkerke R^2) has reached 0.752, so the ability of the independent variables to interpret any changes happens for the dependent variable is a good explanation percentage. The percentage of the total correct classification was %82.63 which is considered a high and good percentage exceeding the minimum division efficiency of 73%. Thus, the model is generally good as the model's significance is less than the allowable error level of 1%.

Keywords: Logistic Regression, Fertility, Unmet Need, The determination coefficient, Upper Egypt Governorates, Total Correct Classification, Demo- graphic Health Survey

1. Introduction

The population problem needs more research, study and careful handling of all its data, the reality is that Egypt is a developing country that needs a more precise response to that problem and leave no field or port unless and must work through it to solve this problem. Effective and positive family planning programs that focus on reducing high population growth rates should be adopted. Egypt is working by providing a package of family planning methods that have led to acceptable results in the area of fertility reduction. The results of the 2008 Demographic Health Survey (DHS) confirm that the use of family planning methods is sometimes hampered by some obstacles Such as "discontinuation" of

use, "misuse" which invalidates its effectiveness or "failure to comply" with the timing of use.

If we aim to reach the total fertility rate to 2 children in 2019, then the rates of practice should rise to about 67% instead of 62%, which is the current rate. Some studies and follow-up of the behavior of women confirmed that there are quite a few women with unmet needs in family planning, and this percentage increases in rural areas. A good deal with this percentage of women may increase the rates of practice because they are actually targeted and thus reduce the number of births and fertility of women and improve their reproductive health. The reports issued by the Central Agency for Public Mobilization and Statistics in Egypt confirm that the proportion of rural population in Egypt is about 57.4% While the proportion of their births represent 47.5% of the total births in Egypt, Consequently, it is a large proportion that entails studying the behavior of rural women.

2. Research Problem

Despite the relatively low levels of reproduction at the level of the Republic, but there are some governorates such as Minya, Assiut and Sohag still suffer from high rates of reproduction and unmet need is still high, this inhibits the achievement of the target total reproduction rate in Egypt of 2.1 births per woman. In addition, the use rates in these governorates are not at the required level.

Indeed, by examining the position of service provision in those areas and linking it to practice ratios this puts our hands on some of the catalysts in explaining and understanding the high unmet need rates which rise to 15.4% in rural Upper Egypt, Hence the main problem of research is the knowing of the causes of high fertility and unmet need methods may contribute to achieving the goals of family planning strategies and reduce fertility levels, This is due to a clear relationship between the use rate and the fertility rate.

Therefore, the importance of this study aimed at monitoring the most important social, economic, demographic and health features of the governorates Minya, Assiut and Sohag as a sample of the governorates of Upper Egypt which suffer from low population characteristics to identify the reasons for the high level of fertility and the unmet need in the countryside of these governorates.

3. Research Importance

This study is closely related to the strategic planning and institutional development work carried out by the National Population Council in dealing with the population problem as a national problem, the Egyptian government attaches the highest priority through activating decentralization and activating the role of localities. Therefore, this study aims to monitor the most important social, economic, demographic and health features of rural governorates Minya, Assiut and Sohag as a sample of the governorates of Upper Egypt that suffer from low population characteristics. This study also aims to achieve the following:-

- Identifying the unmet needs in the field of family planning and integrating them within the framework of future policies and strategies for reproductive health in Egypt through increasing the rate of use of family planning methods and reducing the level of fertility.
- Identify the most important indicators that can be useful for decision makers to reduce reproductive health risks and improve the health status of women this contributes to their empowerment in the fields of education, employment and social participation.
- Contributing to the improvement of the quality of family planning services by revealing the most important strengths and weaknesses in the current efforts in family planning and reproductive health services.

4. Research Objectives

In general, the study aims to comparing some demographic, social and economic characteristics in the rural areas of the three governorates under study, which have high fertility levels and unmet need levels. The study also aims to find out whether each governorate has its own independent characteristics or whether there are special characteristics of Upper Egypt countryside in general. The last main objective includes the following sub-objectives:-

- Recognition of reproductive level in rural governorates Minya, Assiut and Sohag during the past years.
- Measuring the demand for the use of family planning methods and the reasons for discontinuing their use in rural areas of the last governorates.
- Study of reproductive details.
- Identify the most important demographic, economic and social factors affecting the unmet needs in the countryside of the last governorates.
- Identify the most important characteristics that distinguish between fertility of women in the countryside of the last governorates.

5. Research Methodology

This study is based on an in-depth vision on the analysis and identification of the most important causes of unmet need and its percentage in the field of family planning, whether in spacing or stopping in the rural areas of Minya, Assiut and Sohag. The study relied on the descriptive analytical approach as the most appropriate approach in dealing with the subject of this study to reach indicators and trends of value-added and practical evidence in order to achieve the objectives of the study. This will be done through relying on the raw data of the surveys conducted by the National Population Council and the DHS 2017 to study the causes of high fertility and methods of meeting unmet needs in family planning services in rural Egyptian governorates.

The study also relied on methods of statistical analysis appropriate to the quality of data to achieve the objectives of the study to achieve a high level of

accuracy and quality of results. Among the methods used in the analysis (basic measures - statistical models) such as Logistic regression in order to identify the characteristics of each governorate separately and whether they have an impact on the proportion of unmet need, In addition to identify the characteristics of women with a met need and compare them with women with an unmet need in each governorates and to know whether each governorate has its own characteristics or whether there are special characteristics of rural Upper Egypt in general, which will be useful in the development of special policies for each governorate commensurate with their characteristics.

6. Logistic Regression

The logistic linear regression model is one of the most important models used to formulate the function of discrimination; this model is suitable for many uses. In general the Logistic regression is used to analyze the relationship between what happened in the past and what might happen in the future. Suppose that $x_1, x_2, x_3, ..., x_p$ represent the number of independent variables in the model, the dependent variable (y_i) also indicates the probability of a particular event (θ). The Logistic regression model arises when there is a relationship between the probability (θ) and explanatory variables x_i where $i = 1, 2, 3, ..., p$, this relationship can be represented by the following linear model;

$$\theta_k = \beta_0 + \beta_1 x_{1k} + \beta_2 x_{2k} + ... + \beta_p x_{pk} \qquad (1)$$

From the previous equation we find that the left side is finite $(0 < \theta_k < 1)$ while the right side is infinite which forces us to use the logit transform, then the previous equation becomes as follows;

$$P_r(Y_k = 1/X) = \frac{\exp(\beta' X)}{1 + \exp(\beta' X)} = P_r(X) \qquad (2)$$

The logit transform, which is the basis of this model is defined as;

$$g(X) = \ell n\left(\frac{P_r(X)}{1 - P_r(X)}\right) = (\beta' X) \qquad (3)$$

This transformation has many characteristics necessary for the shape of linear regression where it has the following advantages:

1- The logit value $g(X)$ is linear in the parameters $\beta' = \{\beta_0, \beta_1, ..., \beta_p\}$.

2- Using this transform indicates that the values of g (x) are a continuous function.

3- For each right-hand value $(-\infty, \infty)$ there is only one corresponding value of $P_r(X)$ which is limited by the period $(0,1)$.

6.1. The Assumptions of the Logistic Linear Regression Model:

The logistic linear regression model has the following assumptions;

1) The dependent variable is a binary or multiple descriptive variable, the conditional prediction of this variable $E\left(y/x\right)$ is a variable limited by the period $(0,1)$ but the explanatory variables can be continuous or discrete, binary or multiple, also all variables are assumed to be measured without any errors.

2) There is a functional relationship between the dependent variable and the explanatory variables take the form of equation (2).

3) The expected value of random error equal zero $E\left(u_k=0\right)$, the variance of the random error is constant and equal $\left(P_r(X)\left[1-P_r(X)\right]\right)$, the random error $\left(u_k\right)$ follows the binomial distribution with a probability determined by the conditional mean.

4) There is no correlation between random errors (error independence).

5) There is no correlation between random error and explanatory variables.

6) There is no correlation between explanatory variables themselves, where the variables which have a complete correlation between them must be deleted.

7. Application on Data

In logistic regression, it is not as important to estimate parameters as they are used to calculate the probability of a dependent variable which is considered here as an unmet need as a descriptive dependent variable having two faces (there is a need - saturated). Using a stepwise regression method, the results showed a statistically significant regression model at 5% level by using the statistical package programs of SPSS22.

The following section shows a summary of the results obtained. Based on the study framework, 29 independent variables were introduced in the model which appears in table (1) below;

The best model was chosen which it's significant was (probability of error) (0.000), also the efficiency of the correct division of the model was 87.3% which is considered a high and good percentage exceeding the minimum division efficiency of 73%. Seven independent variables were entered into the model represented in (HGOV, W104, W201, W318A, W318B, AGE5, CS).

The following table (2) shows the quality measures of logistic regression model which is considered one of the most important outputs of logistic regression analysis. These values or indicators show the importance of the independent variables and how they interpret the changes that occur to the dependent variable.

Table (1): Explanatory Variables of the Study

Number	Code	Variable Name
1	HGOV	Governorate.
2	W102	Age in years.
3	W103A	Number of times married.
4	W104	Age at first wedding.
5	W105	Before your wedding were you working?
6	W106	After your wedding, you stayed alone? Or with the husband's family?
7	W107	Educational Status.
8	W108	Do you work now?
9	W109	The current educational situation for your husband.
10	W110	s your current husband working?
11	W111	Current husband's profession.
12	W112	Kinship between you and your husband.
13	W201	Did You give birth before that?
14	W208T	The total number of males and females you have given birth.
15	W225A	The number of children you want to have throughout your life.
16	W225B	The number of male children you want to have throughout your life.
17	W225C	The number of female children you want to have throughout your life.
18	W301	Have you ever used family planning methods?
19	W303	What method of family planning do you use?
20	W305	Are there any problems with using this method?
21	W308	Are you satisfied with the level of service and consulting provided by the source?
22	W310A	s this source located in the same village or outside the village?
23	W312	When you first used the method, how many children did you have?
24	W315	What is the first method you have used?
25	W317	What is your age when using the first method?
26	W318A	Have you ever been visited by a rural leader?
27	W318B	Does the rural leader visit you frequently?
28	AGE5	Age group.
29	CS	Number of surviving children.

Table (2): Quality Measures of Logistic Regression Model

χ^2	Degrees of freedom	Significance	similarity of the determination coefficient	The Correct Division Ratio
1143.34	7	0.000	0.743	%87.3

From the previous table it is clear that the similarity of the determination coefficient has reached 0.743. Thus, the ability of independent variables to explain any change in the dependent variable by 74.3% is a good explanation. Also, as can be seen from the table, the value of the χ^2 (1143.34) with a degree of freedom (7) with a significant value of (0.000) which is less than the allowable error level, therefore, we reject the null hypothesis and accept the alternative hypothesis that is; at least there is a variable with a coefficient not equal to zero. The efficiency of the correct division rate was 87.3%, so the model in general is good as its significance is less than the allowable error level.

The following table (3) shows the Wald's test for goodness of fit for the parameters of the logistic regression model and its significance.

Table (3): The Parameters of the Logistic Regression Model and Their Significance

VAR Code	Estimation	S.td Error	Wald Test	D.f	Sig.	Odds Ratio	P_r
HGOV	**0.153**	**0.051**	**10.107**	**1**	**0.001**	**1.176**	**0.54**
W104	**0.106**	**0.018**	**18.587**	**1**	**0.000**	**1.079**	**0.52**
W201	**-2.974**	**0.327**	**61.986**	**1**	**0.000**	**0.076**	**0.07**
W318A	**1.743**	**0.153**	**102.513**	**1**	**0.000**	**4.694**	**0.82**
W318B	**1.324**	**0.116**	**93.384**	**1**	**0.000**	**3.077**	**0.75**
AGE5	**-0.303**	**0.044**	**30.277**	**1**	**0.000**	**0.785**	**0.44**
CS	**0.649**	**0.048**	**137.699**	**1**	**0.000**	**1.749**	**0.64**
Constant	**-1.979**	**0.597**	**9.710**	**1**	**0.002**	**0.156**	**0.13**

It was clear from the previous table (3) the significance of all coefficients of the variables under study. Thus, we accept the hypothesis that there is a Significant effect was statistically significant for the seven variables that proved to be significant on the unmet need or not.

From the results of the previous table (3), the equation of the estimated logistic regression model is;

$$\ell og\ odds = -1.979 + 0.153(HGOV) + 0.106(W104) - 2.974(W201) + 1.743(W318A) + 1.324(W318B) - 0.303(AGE5) + 0.649(CS) \quad (4)$$

The following table (4) shows the results of the goodness of fit test for the residuals of the entire logistic model.

Table (4): The Goodness of Fit Test for the Residuals of the Logistic Regression Model

χ^2	Degrees of freedom	Significance
8.347	**7**	**0.357**

From the previous table we note that, the value of χ^2 was (8.347) and the significance was equal to (0.357), this leads to acceptance of the null hypothesis that there are no statistically significant differences between actual and estimated values, this confirms the goodness of fit of the entire model.

Also, the following table (5) presents the efficient division of the logistic regression model. It is clear from the table that there are 2390 units has been properly divided by percent 85.6% from 2790 total units from those who responded were (no need and saturated). Also there are 906 units has been properly divided by percent 75.68% from 1197 total units from those who responded were (there is a need–spacing/stop). It is also clear that the number of 3296 items was divided properly by percent 82.63% of the sample size, which is 4000 units.

Due to the inefficient quality of the overall model, some independent variables were excluded because there were statistically significant differences between estimated and actual values.

Table (5): The Efficient Division of the Logistic Regression Model

Observed		Expected		The Correct Division Ratio
		Unmet Need		
		No Need and Saturated	There is a Need (Spacing/Stop)	
Unmet Need	No Need and Saturated	2390	402	%85.6
	There is a Need (Spacing/Stop)	291	906	%75.68
Total ratio				%82.62

The following table (6) and figure (1) show the area under the rocker curve. From this area, the sensitivity of the model can be measured in classification

Table (6): Area under the Rock Curve

Area	S.td Error	Significance	95% Confidence Interval	
			Lower Limit	Upper Limit
0.924	0.004	0.000	0.904	0.931

The above table shows that the area under the rock curve was 0.924 approximately with a confidence level 95% and a confidence interval ranging between 0.904 and 0.931. This percentage is considered very high, besides it is significant as the test significant was 0.000, what leads us to reject the null hypothesis and accept the alternative hypothesis is that the area of the curve differs from 0.5.

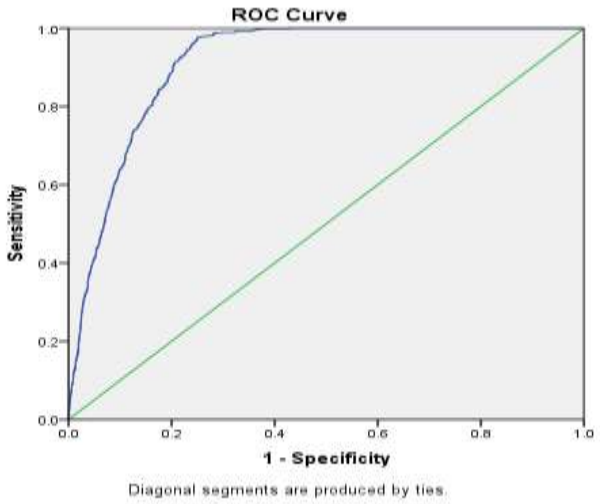

Figure (1): Area under the Rock Curve of the Logistic Model Sensitivity

We can see from the previous figure (1) that the area under the rock curve is 0.924 and this means that the logistic regression is classified in a more significant and better than chance.

8. Recommendations

- Attention to the governorates of Upper Egypt in general in the direction of health services and family planning services and awareness of popular leaders.
- The need to work hard and to benefit from community leaders, informal leaders, clerics, mayors, and outlets to provide literacy services in activating the conviction and use of family planning methods and reduce the number of women with unmet need.
- The need for attention to the nature and characteristics of each governorate separately and identify the most important problems facing family planning services in each governorate.
- Reduce the number of women with unmet need by working to improve the economic, social and demographic characteristics of each governorate.
- The need to pay attention to the role of rural pioneer and activate its role, where the results of the logistic regression showed the importance of the role of rural pioneer in the three governorates under study.

References

[1] A. M. El-Abbassi, *Advanced Statistical Analysis Using SPSS*, Institute of Statistical Studies and Research, Cairo University, Egypt, 2010.

[2] B. H. Chang, Silipsitz and C. Waternaux, Logistic Regression in Meta-Analysis using Aggregate Data, *Journal of the Applied Statistical*, **27** (2000), 411-424. https://doi.org/10.1080/02664760050003605

[3] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, 2nd edition, New York: John Wiley and Sons, 2000. https://doi.org/10.1002/0471722146

[4] El-Zanaty et al., *Health Population Survey*, Egypt, Ministry of Health, 2014.

[5] James Jaccard, *Interaction Effects in Logistic Regression*, Thousand Oaks, CA: Sage Publications, Quantitative Applications in the Social Sciences Series, Vol. 135, 2001. https://doi.org/10.4135/9781412984515

[6] J. F. Hair, W. C. Black, B. J. Babin and R. E. Anderson, *Multivariate Data Analysis*, 7th edition, Upper Saddle River: Prentice Hall, 2010.

[7] L. G. Grimm and P. R. Yarnold, *Reading and Understanding Multivariate Statistics*, Washington, DC: American Psychological Association, 2008.

[8] National Population Council, Causes of High Fertility Level and Methods of Meeting Unmet Needs in Family Planning Services in Minya Governorate, 2011.

[9] National Population Council, Reasons for High Fertility Level and Methods of Meeting Unmet Needs in Family Planning Services in Rural Beni Suef Governorate, 2011.

[10] National Planning Institute, Egypt, Human Development Report, 2010.

[11] R. Elwazeer, H. Semary and Hanim, Methods of Data Mining: Parametric and Non-parametric Methods, *The Egyptian Journal of Population and Family Planning, Institute of Statistical Studies and Research, Egypt*, **45** (2012).

[12] SPSS, SPSS BASE 22.0, User's Guide, Chicago, IL: SPSS Inc, 2013.

Received: January 5, 2020; Published: January 30, 2020