Research Article

Determinants of fertility of women - multiple logistic regression approach: a case study of Kalaburagi district, Karnataka state, India

Vijaykumar Kulkarni*, K S Wali**

*Research Scholar, **Professor, Department of Statistics, Gulbarga University, Kalaburagi, Karnataka, INDIA.
Email: kumarviju_02@yahoo.com

Abstract The aim of this study is to assess determinants of fertility of women of urban and rural areas of Kalaburagi District, Karnataka, India. A **c**ross sectional descriptive study design was used. Systematic stratified random sample of 1500 women were selected in which, 750 were from rural and 750 were from urban areas. The data on different factors were collected through questionnaire with direct personal interview method. The causal relationships were established by multiple logistic regression model. The income, spouse is employed, practicing any family planning method and satisfactory medical facilities available have positive association with women fertility of rural area and age groups, age of spouse, and spouse is a blood relative have negative association with women fertility of rural area. Similarly, the knowledge about contraceptives, satisfactory medical facilities available and opinion about number of children for comfortable life have positive association with women fertility of urban area and age and practicing any family planning method have negative association with women fertility of urban area.
Keywords: Logistic regression modeling, women fertility, fertility determinants.

***Address for Correspondence:**
Dr. Vijaykumar Kulkarni, Research Scholar, Department of Statistics, Gulbarga University, Kalaburagi, Karnataka, INDIA.
Email: kumarviju_02@yahoo.com
Received Date: 01/06/2015 Revised Date: 08/06/2015 Accepted Date: 10/06/2015

Access this article online	
Quick Response Code:	Website: www.statperson.com
	DOI: 12 June 2015

INTRODUCTION

Fertility is an important component of population dynamics which plays a major role in the size and structure of a given population. Uncontrolled fertility has adversely influenced the socio-economic, demographic and environmental development of the country. Studies pertaining to fertility of women are important to policy makers in view of their participation in labour and the participation in the women self- help groups, for the economic development of the household (Salikumar, 1992; Joseph Raj, 1996; Chiranjeevulu, 2003; Sampoornam, 2003; Ramachandran, 2005, K. Senthamarai Kannan and V. Nagarajan, 2008). Biswas (1972) has examined the linear model with single equation, Fertility rate increases upto 20-24 for women and slowly declines and also currently majority of the married women feel that they don't want more than one child (NFHS-3). Kannan *et al*. (1998) have analyzed birth interval in human fertility. In view of the importance acquired by the women force in the development of the country, the study is taken in Kalaburagi district. The regression methods have become an integral component of any data analysis concerned with the explanation of relationship between a response variable and one or more explanatory variables called covariates (Javali S. B. 2009). Many different types of linear models have been seen in the literature and their use is discussed in many areas including demography. The use of logistic regression modeling has been exploded during the past few decades. This method is now commonly applied in many fields including population research, biomedical research, medicine, agriculture and medical epidemiology, etc. The logistic regression model is an important method to understand the principle that the goal of an analysis is the same as that of traditional model building technique used in statistical theory to find suitable description of relationship between response

How to site this article: Vijaykumar Kulkarni, K S Wali. Determinants of fertility of women - multiple logistic regression approach: a case study of Kalaburagi district, Karnataka state, India. *International Journal of Recent Trends in Science and Technology* June 2015; 15(2): 351-357 http://www.statperson.com (accessed 15 June 2015).

variable and a set of covariates. In traditional linear regression techniques it is assumed that dependent variable must be continuous or quantitative. In logistic model, we consider situations where the response variable is a categorical random variable, having only two possible outcomes called binary or dichotomous. This difference between logistic and linear regression is reflected both in the choice of a parametric model and in the assumptions. In this paper, the number of children (women fertility) is considered as dichotomized response variable. Since, the response variables are dichotomous, it is inappropriate to assume that they are normally distributed. Thus, the data cannot be analyzed using the traditional linear regression methods. It is convenient to denote one of the outcomes of response of women fertility as ≤2 children and ≥3children. It is a standard practice to let Y_1 (women fertility) to be two binary or dichotomous response variables, which are defined as

$$Y_1(\mathit{women fertility}) = \begin{cases} 1, & \mathit{if}\ Y_1\ (>/=3\ \text{children}) > 0 \\ 0, & \mathit{if}\ Y_1\ (</=2\ \text{children}) = 0 \end{cases}$$

In this article our main aim is to use multiple logistic regression models for modeling women fertility data.

DEPENDENT VARIABLES AND INDEPENDENT COVARIATES

The women fertility is the dependent variable. For analysis purpose, the number of children are grouped as 0 if ≤2children and 1 if ≥3children. Apart from response variable, the data set on independent covariates like residence (urban =0, rural=1), age groups (>=31yrs=0, <=30 yrs=1), education (literate=0, illiterate=1), income (high income=0, low income=1), occupation (house-wife=0, employed =1), age of spouse (>=41yrs= 0, <=40yrs = 1), duration of marriage (21-40 yrs= 0, 1-20yrs = 1), spouse is employed (no=0, yes=1), age respondent at marriage (>=21yrs = 0, <=20 yrs = 1), age of spouse at the time of marriage (<=25yrs = 0, >=26yrs = 1), delivery at maternity hospitals (no=0, yes=1), satisfied with the facilities (no=0, yes=1), abortions (no=0, yes=1), Knowledge about contraceptives (no=0, yes=1), practicing any family planning method (no=0, yes=1), satisfactory medical facilities available (no=0, yes=1), opinion about number of children for comfortable life (One =0, >=Two =1), Spouse is a blood relative (no =0, yes >= 1) and opinion what should be the age at marriage for girls (<=20yrs = 0, >=21yrs =1) are obtained by structured questionnaire and interview method. The objective of this paper is to utilize logistic regression model for modeling response variable and to determine the important determinants.

Formulation and Fitting of Multiple Logistic Models

Consider a collection of independent covariates (at least interval scale) denoted by the vector

$$x' = (x_1, x_2, \ldots, x_p)$$

Let the conditional probability that the response variable is present be denoted by p[Y=1 | x] = π (x). The logit of the multiple regression model is given by the equation

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots.. + \beta_p x_p .$$

with the logistic model given by

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

If some of the covariates such as gender, socio-economic status etc., are discrete that are measured in nominal scale and so forth is inappropriate to include them in the model unless if they are interval scale variables. The number used to represent the various levels of these nominal scale variables are merely identifiers and have no numeric significance. In general, if a nominal scaled variable has k possible values, then k-1 design variables are needed. Thus, the logit for a model with p covariates and the j^{th} variable being discrete would be

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \sum_{l=1}^{k_j - 1} \beta_{jl} D_{jl} + \beta_p x_p$$

The likelihood of β is given by

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

where

The log of likelihood function is given by

$$\log l(\beta) = L(\beta) = \sum_{i=1}^{n} \{ y_i \log[\pi(x_i)] + (1 - y_i) \log[1 - \pi(x_i)] \}$$

Here we get (p+1) likelihood equations that are obtained by differentiating the log likelihood function with respect to the p+1 coefficient. Then, the method of estimating the variance and co-variances of these estimated coefficients follows from the well developed theory of maximum likelihood estimation (Rao, 1973; David and Stanley, 2000). This theory states that estimators are obtained from the matrix of partial derivatives of the log likelihood function.

Analysis using multiple logistic regression models

The dichotomized women fertility data are analyzed. Measures of central tendency and dispersion are computed for both continuous and nominal variables. The multiple logistic regression model is constructed between the binary response variables i.e. women fertility with covariates independently. In the first step, the multiple logistic full model is constructed for considering all covariates and in the second step, the stepwise called multiple logistic reduced model is performed by considering only significant covariates from the full model. In the selection procedure using the stepwise multiple logistic model analysis, we first select the variable having a greatest influence power. Then the effect of this variable is eliminated from the information content of all the other variables. The variable, which then has the greatest power of influence after the above elimination procedure, is ranked as the second etc. Thus, the variables are listed in decreasing order with respect to probability of additional information gained from including further variables was less than 0.05. Lastly, the variable having the weaker influence power may be dropped from the final analysis. In order to weigh the significance of each chosen variable with respect to their influence, its correlation with multiple logistic regression model and odds ratios, 95% confidence intervals of each covariate are computed.

RESULTS

In this section, we illustrate the nature of distribution of study subjects according to different characteristics. Comparisons in terms of log likelihood functions are carried, discussed and presented. Table 1 presents the distribution of women according to different characteristics. A total of 1500 women are included in the study of which 57.73% belong to >=31yrs of age groups and 42.27% belong to <=30 yrs of age groups with overall mean age 33.56±7.60. Similarly, 77.33% are illiterates, 67.60% have high income, 62.07% are employed, 65.33% of spouses belong to <=40yrs of age group. However, 86.07% of women completed 1-20yrs as a duration of marriage, 98.60% of their spouses are employed, 44.27% belong to <=20yrs of age at the time of marriage, 61.80% spouses belong to <=25yrs of age at the time of marriage, 86.13% of women delivered in maternity hospitals and 72.33% felt that they are satisfied with hospital facilities, and 86.13% had abortions. Further, 74.80% have knowledge about contraceptives, but 55.07% are practicing any family planning method, 82.53% felt that satisfactory medical facilities are available in the hospital, 59% of the women felt that, one child is enough for comfortable life. In the present study most of the selected women have married non blood relative, 64.00% women told that, the age at marriage for girls should be >=21yrs.

Table 1: Distribution of study subjects according to different characteristics

Factors	No of respondents	% of respondents
Residence		
Urban	750	50.00
Rural	750	50.00
Age groups		
>=31yrs	866	57.73
<=30 yrs	634	42.27
Education		
Literate	340	22.67
Illiterate	1160	77.33
Income		
High income	1014	67.60
Low income	486	32.40
Occupation		
House-wife	569	37.93
Employed	931	62.07
Age of spouse		
>=41yrs	520	34.67
<=40yrs	980	65.33
Duration of marriage		
21-40 yrs	209	13.93
1-20yrs	1291	86.07
Spouse is employed		
No	21	1.40
Yes	1479	98.60
Age respondent at marriage		

>=21yrs	836	55.73
<=20 yrs	664	44.27
Age of spouse at the time of marriage		
<=25yrs	927	61.80
>=26yrs	573	38.20
Delivery at maternity hospitals		
No	208	13.87
Yes	1292	86.13
Satisfied with the facilities		
No	415	27.67
Yes	1085	72.33
Abortions		
Yes	1292	86.13
No	208	13.87
Knowledge about contraceptives		
No	378	25.20
Yes	1122	74.80
Practicing any family planning method		
No	674	44.93
Yes	826	55.07
Satisfactory medical facilities available		
No	262	17.47
Yes	1238	82.53
Opinion about number of children for comfortable life		
One 0	890	59.33
>=Two 1	610	40.67
Spouse is a blood relative		
No	1125	75.00
Yes	375	25.00
Opinion what should be the age at marriage for girls		
<=20yrs 0	540	36.00
>=21yrs 1	960	64.00
Total	**1500**	**100.00**

Comparison of full and reduced logistic regression models with women fertility data as a whole

A total of 19 covariates are included in the model for the total samples, in which 9 covariates (70.00%) have found to be significant. Five covariates such as place of residence, age respondent at marriage, knowledge about contraceptives, satisfactory medical facilities available and opinion about number of children for comfortable life have positive association with women fertility. These significant covariates exhibited positive regression coefficients, indicating that these are supporting to increase the women fertility. However, four covariates such as age groups, age of spouse, practicing any family planning method and opinion what should be the age at marriage for girls have negative association with women fertility. These significant covariates exhibited negative regression coefficients, indicating that these are supporting to decrease the women fertility as a total. The log likelihood ratio of full model is -519.4437 and of reduced model is -526.1831. Based on log likelihood, the full and reduced logistic regression models fit equally better for women fertility data as a whole. Thus, there is no advantage in excluding some of the covariates from the model for assessment of significant determinants of women fertility. The estimated odds ratios for different covariates are as follows. The estimated odds ratio of residence (OR=1.88, 95% CI: 1.19-2.96), education of women (OR=1.30, 95% CI: 0.76-2.23), income (OR=1.37, 95% CI: 0.87-2.15), age of respondent at marriage (OR=2.51, 95% CI: 1.56-4.04), age of spouse at the time of marriage (OR=1.06, 95% CI: 0.69-1.64), delivery at maternity hospitals (OR=1.21, 95% CI: 0.67-2.21), satisfied with the medical facilities (OR=1.18, 95% CI: 0.76-1.82), experience of abortions (OR=1.17, 95% CI: 0.75-1.84), knowledge about contraceptives (OR=2.17, 95% CI: 1.25-3.79), satisfactory medical facilities available in the hospital (OR=8.95, 95% CI: 4.99-16.03), opinion about number of children for comfortable life (OR=10.38, 95% CI: 7.26-14.85), spouse is a blood relative (OR=1.28, 95% CI: 0.89-1.84) and opinion what should be the age at marriage for girls (OR=0.59, 95% CI: 0.39-0.91). The covariates like residence, education of women, income, age respondent at marriage, age of spouse at the time of marriage, delivery at maternity hospitals, satisfied with the facilities, experience of abortions, knowledge about contraceptives, satisfactory medical facilities available in the hospital, opinion about number of children for comfortable life, spouse is a blood relative and opinion what should be the age at marriage have significant association with women fertility (Table 2 and 3).

Comparison of full and reduced logistic regression models with women fertility data of rural samples

A total of 18 covariates are included in the model, in which 7 covariates (38.88%) have found to be significant, in which four covariates such as income, spouse is employed, practicing any family planning method and satisfactory medical facilities available have positive association with women fertility of rural samples. These significant covariates exhibited positive regression coefficients, indicating that these are supporting to increase the women fertility of rural samples. However, three covariates such as age groups, age of spouse, and spouse is a blood relative have negatively associated with women fertility of rural samples. These significant covariates exhibited negative regression coefficients, indicating that these are supporting to decrease the women fertility of rural samples. The log likelihood ratio of full model is -202.368 and of reduced model is -209.3165. Based on log likelihood, the reduced logistic regression model better fits as compared to full model. Thus, there is an advantage in excluding some of the covariates from the model for assessment of significant determinants of women fertility of rural samples. The estimated odds ratio of income (OR=3.56, 95% CI: 1.80-7.01), age respondent at marriage (OR=4.94, 95% CI: 1.83-13.31), delivery at maternity hospitals (OR=1.52, 95% CI: 0.23-9.96), knowledge about contraceptives (OR=1.26, 95% CI: 0.19-8.30), satisfactory medical facilities available in the hospital (OR=4.83, 95% CI: 1.71-13.62), opinion about number of children for comfortable life (OR=13.20, 95% CI: 7.43-23.46), and spouse is a blood relative (OR=1.21, 95% CI: 0.66-2.22). The covariates like income, age respondent at marriage, delivery at maternity hospitals, knowledge about contraceptives, satisfactory medical facilities available in the hospital, opinion about number of children for comfortable life and spouse is a blood relative have significant association with women fertility of rural samples (Table 2 and 3).

Comparison of full and reduced logistic regression models with women fertility data of urban samples

A total of 18 covariates are included in the model, in which 5 covariates (33.33%) have found to be significant, in which three covariates such as knowledge about contraceptives, satisfactory medical facilities available and opinion about number of children for comfortable life have positive association with women fertility of urban samples. These significant covariates exhibited positive regression coefficients, indicating that these are supporting to increase the women fertility of urban samples. However, two covariates such as age groups and practicing any family planning method are negatively associated with women fertility of urban samples. These significant covariates exhibited negative regression coefficients, indicating that these are supporting to decrease the women fertility of urban samples. The log likelihood ratio of full model is -282.6340 and of reduced model is -294.7643. Based on log likelihood, the reduced logistic regression model better fits as compared to full model. Thus, there is an advantage in excluding some of the covariates from the model for assessment of significant determinants of women fertility of urban samples. The estimated odds ratio of education of women (OR=1.44, 95% CI: 0.80-2.60), income (OR=1.07, 95% CI: 0.43-2.69), age respondent at marriage (OR=2.80, 95% CI: 0.22-35.74), age of spouse at the time of marriage (OR=1.68, 95% CI: 0.95-2.99), delivery at maternity hospitals (OR=1.33, 95% CI: 0.78-2.25), satisfied with the facilities (OR=1.07, 95% CI: 0.53-2.13), experience of abortions (OR=1.36, 95% CI: 0.78-2.36), knowledge about contraceptives (OR=1.37, 95% CI: 0.62-2.99), experience of any family planning method used (OR=2.08, 95% CI: 1.12-3.86), opinion about number of children for comfortable life (OR=14.14, 95% CI: 6.42-31.15), spouse is a blood relative (OR=10.85, 95% CI: 6.56-17.96) and opinion what should be the age at marriage for girls (OR=1.13, 95% CI: 0.68-1.88). The covariates like education of women, income, age respondent at marriage, age of spouse at the time of marriage, satisfied with the facilities, experience of abortions, knowledge about contraceptives, experience of any family planning method used, opinion about number of children for comfortable life, spouse is a blood relative and opinion what should be the age at marriage for girls have significant association with women fertility of urban samples (Table 2 and 3).

Table 2: The estimated odds ratios of covariates from full logistic regression model to women fertility

Covariates	Total			Rural			Urban		
	Odds ratio	95% confidence interval		Odds ratio	95% confidence interval		Odds ratio	95% confidence interval	
Residence	1.88	1.19	2.96						
Age groups	0.15	0.10	0.23	0.07	0.03	0.17	0.19	0.11	0.35
Education	1.30	0.76	2.23	0.84	0.10	7.40	1.44	0.80	2.60
Income	1.37	0.87	2.15	3.56	1.80	7.01	1.07	0.43	2.69
Occupation	0.74	0.53	1.04	0.66	0.37	1.16	0.86	0.54	1.36
Age of spouse	0.45	0.29	0.69	0.34	0.18	0.62	0.54	0.28	1.03
Duration of marriage	0.63	0.34	1.15	0.43	0.15	1.30	0.58	0.26	1.29
Spouse is employed	0.32	0.06	1.62	-	-	-	-	-	-
Age respondent at marriage	2.51	1.56	4.04	4.94	1.83	13.31	2.80	0.22	35.74
Age of spouse at the time of marriage	1.06	0.69	1.64	0.63	0.24	1.66	1.68	0.95	2.99
Delivery at maternity hospitals	1.21	0.67	2.21	1.52	0.23	9.96	1.33	0.78	2.25
Satisfied with the facilities	1.18	0.76	1.82	0.84	0.37	1.91	1.07	0.53	2.13
Abortions	1.17	0.75	1.84	0.98	0.53	1.81	1.36	0.78	2.36
Knowledge about contraceptives	2.17	1.25	3.79	1.26	0.19	8.30	1.37	0.62	2.99
Practicing any family planning method	0.30	0.20	0.45	0.73	0.35	1.54	2.08	1.12	3.86
Satisfactory medical facilities available	8.95	4.99	16.03	4.83	1.71	13.62	0.20	0.12	0.34
Opinion about number of children for comfortable life	10.38	7.26	14.85	13.20	7.43	23.46	14.14	6.42	31.15
Spouse is a blood relative	1.28	0.89	1.84	1.21	0.66	2.22	10.85	6.56	17.96
Opinion what should be the age at marriage for girls	0.59	0.39	0.91	0.08	0.03	0.21	1.13	0.68	1.88
Log likelihood	**-519.4437**			**-202.368**			**-282.6340**		

Table 3: The estimated odds ratios of covariates from reduced logistic regression model to women fertility

Covariates	Total			Rural			Urban		
	Odds ratio	95% confidence interval		Odds ratio	95% confidence interval		Odds ratio	95% confidence interval	
Residence	0.14	0.09	0.22	-	-	-	-	-	
Age groups	0.40	0.27	0.58	0.06	0.03	0.14	0.13	0.08	0.19
Income	-	-	-	2.60	1.44	4.71	-	-	-
Age of spouse	2.85	1.93	4.21	0.27	0.15	0.46	-	-	-
Age respondent at marriage	2.08	1.27	3.40	5.90	2.80	12.44	-	-	-
Knowledge about contraceptives	0.30	0.21	0.45	-	-	-	1.73	1.03	2.89
Practicing any family planning method	7.54	4.36	13.03	-	-	-	0.19	0.12	0.31
Satisfactory medical facilities available	11.34	8.06	15.96	2.89	1.35	6.20	14.43	6.86	30.35
Opinion about number of children for comfortable life	0.66	0.44	0.97	15.06	8.64	26.24	11.57	7.27	18.44
Opinion what should be the age at marriage for girls	1.61	1.09	2.39	0.10	0.04	0.23	-	-	
Log likelihood	**-526.1831**			**-209.3165**			**-294.7643**		

DISCUSSION AND CONCLUSIONS

Different statistical methods i.e. regression analysis, multilevel modeling, logistic regression and ordinal regression techniques employed to analyze the women fertility data yield results having different focuses. The regression methods allow researchers to identify factors related to socio-economic-demographic and other factors related to fertility that contribute to overall status of response variables. These methods also permit researcher to estimate the magnitude of the effect of covariates. Usually, the nature of women fertility data is in count scale. Therefore, the women fertility data are converted into binary or dichotomous outcomes (<=2 children, >=3 children). Use of a dichotomous outcomes in traditional multiple linear regression model violates the assumption of normality and homoscedasticity. Hence, the main aim of the present paper is to utilize the applications of logistic regression model to assess the effect and relationship between covariates and binary response variables independently. Impact of different factors (determinants) on women fertility has already been investigated in several studies. The strength of the present study is the inclusion of a large number of detailed probable and prominent covariates or determinants for modeling women fertility. In the present study, one of the most significant factors was age of the woman, it was correlated significantly with fertility of women and similar finding was obtained in a study by Abdelkader SM *et al. (*2014). Women education had no significant effect on their fertility behavior in current study. Even Abdelkader SM *et al. (*2014) study showed that the effect of husband's education on fertility conducted in Riyadh showed not associated. Further, the factors such as age of spouse, practicing any family planning method and opinion what should be the age at marriage for girls are found to be significant and negative influence of women fertility. But the factors such as age of respondent at marriage, knowledge about contraceptives, satisfactory medical facilities available and opinion about number of children for comfortable life are found to be significant and positive influence of women fertility in the study. Further, in this paper, we compared performance of full logistic model with that of reduced logistic model using log likelihood estimate for women fertility data. The results show that reduced logistic regression model is slightly a better fit as compared to full logistic regression model applied to dichotomized women fertility data.

REFERENCES

1. Abdelkader SM et al.: Factors Affecting Fertility among Saudi Women. Sch. J. App. Med. Sci., 2014; 2(3C):1063-1069
2. Biswas, S.: On some models of the gestation period for live birth. Sankhya, Series B, 34: 79-88 (1972).
3. Chiranjeevulu, T.: Empowering women through self help Groups: Experiences in Experiments. The Journal of Kurukshetra, 51(5): 30-32 (2003).
4. David W. H., Stanley (2000). Applied Logistic Regression, Second edition, A Wiley Inter science Publication, John Wiley and Sons, Inc. New York
5. Javali S. B. 2009: Some Non-Parametric Tests for Location and Statistical Analysis of Determinants of Selected Diseases of Oral Health, Unpublished thesis
6. Joseph Raj, K.: Women's work participation in India. Southern Economists, 35 (15): 17-18 (1996).
7. K. Senthamarai Kannan and V. Nagarajan: Factor and Multiple Regression Analysis for Human Fertility in Kanyakumari District Anthropologist, 10(3): 211-214 (2008).
8. Prithiviraj, C.A and Senthamarai Kannan, K.: Birth intervals in human tertility Analysis, 13th Annual conference of Ramanujan Mathematical Society Department of Mathematics, Manonmaniam Sundaranar University, Tirunelveli (1998).
9. Ramachandran, T.: Self help Groups and Social Changes in Kanyakumari District - An Analysis. Journal of HRC, Annual vol. XIV: 92-95 (2005).
10. Rao C. R. 1973. Linear Statistical Inference and its application, second edition, Wiley Inc., New York.
11. Salikumar, P.N.: Determinants of Fertility in Rural Area - A Study of Selected Villages in Kanyakumari District. Ph. D. Thesis (unpublished), Madras University, Chennai (1992)
12. Sampoornam: An Economic study of Women Construction Workers in Kanyakumari District. Ph.D Thesis (unpublished), ManonmaniamSundarnar University, Thirunelveli (2003).

Source of Support: None Declared
Conflict of Interest: None Declared