Modelling Age at First Marriage of Female Adolescents in Bangladesh

Md. Tofazzal Hossain[1,*], Md. Ahshanul Haque[2] and Mohammed Nasser[3]

[1]Department of Statistics, Bangabandhu Sheikh Mujibur Rahman Science and Technology University
Gopalganj, Bangladesh.

[2,3]Department of Statistics, Rajshahi University
Rajshahi, Bangladesh.

[*]*Corresponding author's email: tofazzal.stat [AT] gmail.com*

ABSTRACT— ***Age at first marriage has an important impact on population growth rate. Modelling age at first marriage is very crucial since the growth rate of Bangladesh is very high. Here logistic regression model has been used for modelling age at first marriage using Bangladesh Demographic and Health Survey (BDHS-2011) data. The BDHS data was collected under two stage stratified sampling plan. That is why cluster correction and weight adjustment is necessary for analyzing this data. An extensive literature search fails to find the usage of these adjustments for modelling age at first marriage. In this paper, these two adjustments have been used for modelling and the prediction performance of the model has been measured by the training and test error.***

Keywords— Age at first marriage, cluster correction, weight adjustment and logistic regression analysis.

1. INTRODUCTION

Marriage is very common in Bangladesh. The legal age of marriage for female in Bangladesh is 18 years. But Bangladeshi females are getting married before reaching 18 years old especially in rural areas. The age at first marriage in Bangladesh is low relative to the other developing countries in the world [1]. Early marriage has an important effect of fertility. Women with early marriage generally get a longer period of exposure of pregnancy which leads to higher fertility. Women in Bangladesh and the other developing countries in Asia and Africa where the age at first marriage is low have higher fertility. And higher fertility leads to higher growth rate of population. On the other hand, the delay in the age of first marriage leads to low fertility [2]. In Bangladesh, the main objective of marriage is to have children. Early marriage leads to early childbearing and early childbearing leads to health risk of both mother and child even to death of both mother and child. It is necessary to control early marriage by controlling the factors that influences the early age at first marriage. So modelling age at first marriage is very important.

A lot of work has been done for the variable 'Age at First Marriage' for BDHS data such as [1-4]. In these articles, cluster correction and weight adjustment have not been considered, although the data is collected under two stage stratified sampling. In [5] cluster correction and weight adjustment have been introduced for predicting the number of children ever using BDHS (2011) data. In this paper, we have use logistic regression model making these adjustments for modelling age at first marriage and measure the prediction performance of the model using training and test errors.

There are four sections in this paper. In Section one the problem and what we want to do have been introduced. Section two discusses data source, data nature and statistical technique that is used in this article. Section three provides results and discussion. Finally, in section four the conclusion and the findings have been described.

2. METHODS AND MATERIALS

In this section we discuss the nature and sources of data, the variables under study and the statistical technique which is used for the analysis.

2.1 Data description

In this study, we have used secondary data. The source of the data is the Bangladesh Demographic and Health Survey (BDHS) - 2011. The National Institute for Population Research and Training (NIPORT) of the Ministry of Health and Family Welfare conducted BDHS 2011. Mitra and Associates, a Bangladeshi research firm, implemented this survey. The BDHS-2011 data were collected under two stage stratified random sampling. In first stages, each division was

subdivided into urban and rural areas. The urban areas of each division were further subdivided into: city corporations and other than city corporations. In the first stage, 600 Enumeration Areas (EA) were selected, with probability proportional to the EA size and with independent selection in each sampling stratum. In the second stage of selection, a fixed number 30 household per cluster was selected with an equal probability systematic selection from the newly created household listing.

With this design, 18,000 residential households were selected, and was expected to result in completed interviews with about 18,000 ever married women. In addition, ever-married men age 15-54 in every third household were eligible for the male survey.

2.2 Variables under study

There are a number of socio-economic and demographic variables in BDHS 2011 data. With the help of previous studies [4,6] we have considered only seven explanatory variables namely respondent's place of residence, husband's education, wealth index, religion, respondent currently working, and food security which highly related to the response variable. The data of the response variable age at first marriage was collected in the name of age at first cohabitation. For our study purposes, we divide it into two categories: 1 for the first age of marriage less than 18 years and 0 for 18 or more years.

2.3 Logistic regression

Logistic regression is a well-known technique for categorical data analysis. In this regression, the dependent/response variable is categorical and the independent variables may either be categorical or numeric [7]. In this paper, we have used binary logistic regression model. The binary logistic regression model for the binary dependent variable Y and the independent variables $X_1, X_2, \dots, X_r$ is given by

$$p = P(Y = 1 | X1, X2, \dots, Xr) = \frac{e^{\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_rX_r}}{1 + e^{\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_rX_r}}$$

The above relationship can be expressed alternatively in terms of log odds associated with p as follows:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1X_1 + \beta_2X_2 + \cdots + \beta_rX_r.$$

One key assumption of the logistic regression model is that the log odds of the response variable is linearly related to the explanatory variables.

3. RESULTS AND DISCUSSION

Here we show the proportions for our study variables which are risk factors for early marriage for all the BDHS data in Table 1.

From Table 1, we observe that for the variable respondent's education and husband's education, the categories primary, secondary and higher are increasing and the category no education is decreasing day by day. It indicates that the education rate (both respondent and husband) is increasing over time. The poor are becoming poorer and the rich are becoming richer. The working status of women are increasing except in 2011. In 2011, the proportion of working women has fallen down. One of the reasons behind this may be women are being educated but not getting jobs proportionally. The results of the logistic regression analysis are given in Table 2.

From the Table 2 it is observed that all (except wealth index: middle) the predictor variables have statistically significant effect on the response variable. Rural females are 1.14 times more likely to have early marriages than urban females. Females whose husband's education level is primary are 21% less likely to have early marriages than those females whose husband's education level is no education. Again females whose husband's education level is secondary are 45% less likely to have early marriages than those females whose husband's education level is no education. And females whose husband's education level is higher are 80% less likely to have early marriages than those females whose husband's education level is no education. Females who are currently working are 31% less likely to have early marriages than those who are not currently working. Rich females are 27% less likely to have early marriages than those who are poor. Muslim female are 2.03 time more likely to have early marriages than those with other religions. Females whose families are secured in food are 15% less likely to have early marriages than those whose families are unsecured in food.

From our analysis, we observe that rural female are being married early compared to urban female. So, social awareness should be increased especially in rural areas. Educated husband do not want to marry early. Education rate should be increased to lessen early marriages. Female employment rate should be increased because unemployed female are getting married early. Awareness of Muslim families should be increased by taking programs describing the bad

effects of early marriages. Food security should be ensured because families with unsecured in food want to reduce their family member through the early marriage of their female children.

In order to measure the prediction performance of the model we randomly divide the whole data set into two parts: training data and test data. The training data set consist of 75% and the test data set consist of 25% of the whole data set. Using the 75% data we fit the model and calculate the misclassification error which is called training error. Again using the fitted model we predict the outcomes for the 25% data and calculate the misclassification error which is called test error. These errors are given in Table 3.

Table 1: Proportions of our categorical study variables for different BDHS data.

Variables	1993, 1994	1996, 1997	1999, 2000	2004	2007	2011
	N=9640	N=9127	N=10544	N=11440	N=10996	N=17842
Place of residence						
Urban	1466 (15.21)	1449 (15.88)	3150 (29.87)	3904 (34.13)	4151 (37.75)	6196 (34.73)
Rural	8174 (84.79)	7678 (84.12)	7394 (70.13)	7536 (65.87)	6845 (62.25)	11646 (65.27)
Respondent's education						
No education	5431 (56.34)	4899 (53.68)	4575 (43.39)	4419 (38.63)	3525 (32.06)	4639 (26)
Primary	2693 (27.94)	2530 (27.72)	2997 (28.42)	3381 (29.55)	3268 (29.72)	5332 (29.88)
Secondary	1302 (13.51)	1403 (15.37)	2415 (22.9)	2949 (25.78)	3345 (30.42)	6406 (35.9)
Higher	214 (2.22)	295 (3.23)	557 (5.28)	691 (6.04)	855 (7.78)	1465 (8.21)
Husband's education						
No education	4290 (44.54)	4015 (44.06)	3920 (37.18)	4124 (36.05)	3602 (32.78)	5197 (29.13)
Primary	2307 (23.95)	2253 (24.72)	2339 (22.18)	2903 (25.38)	2881 (26.21)	4834 (27.09)
Secondary	2086 (21.66)	1901 (20.86)	2593 (24.59)	2947 (25.76)	2900 (26.39)	5175 (29)
Higher	882 (9.16)	859 (9.43)	1481 (14.05)	1456 (12.73)	1598 (14.54)	2627 (14.72)
Don't know	75 (0.7)	99 (1.08)	211 (1.87)	10 (0.09)		9 (0.05)
Wealth index						
Poorest				2048 (17.9)	1775 (16.14)	3096 (17.35)
Poorer				2058 (17.99)	1995 (18.14)	3345 (18.75)
Middle				2147 (18.77)	2095 (19.05)	3428 (19.21)
Richer				2276 (19.9)	2201 (20.02)	3777 (21.17)
Richest				2911 (25.45)	2930 (26.65)	4196 (23.52)
Religion						
Islam	8430 (87.45)	8109 (88.85)	9135 (86.64)	10182 (89)	9924 (90.25)	15845 (88.81)
Christianity	1163 (12.06)	13 (0.14)	27 (0.26)	47 (0.41)	26 (0.24)	48 (0.27)
Hinduism	32 (0.33)	975 (10.68)	1293 (12.26)	1192 (10.42)	1011 (9.19)	1913 (10.72)
Buddhism	13 (0.13)	28 (0.31)	86 (0.82)	13 (0.11)	23 (0.21)	36 (0.2)
Other	2 (0.02)	10 (0.1)	3 (0.01)	6 (0.05)	12 (0.1)	
Respondent currently working						
Yes	1588 (16.47)	3322 (36.4)	2282 (21.64)	2515 (21.98)	3233 (29.4)	2374 (13.31)

Table 2: Results of logistic regression model

Predictor variables	Odds ratio	Std. Err.	t-test	P-value	95% C.I
Place of residence					
Urban	Reference				
Rural	1.14	0.072	1.99	0.047	1.002-1.287
Husband's Education					
No education	Reference				
Primary	0.79	0.053	-3.55	0.000	0.687-0.898
Secondary	0.55	0.039	-8.44	0.000	0.481-0.634
Higher	0.20	0.016	-20.62	0.000	0.168-0.229
Respondent currently working					
No	Reference				
yes	0.69	0.044	-5.93	0.000	0.606-0.778
Wealth Index					
Poor	Reference				
Middle	1.00	0.072	0.02	0.981	0.870-1.153
Rich	0.73	0.053	-4.39	0.000	0.629-0.838
Religion					
Others	Reference				
Islam	2.03	0.188	7.64	0.000	1.691-2.434
Food security					
Insecure	Reference				
Secure	0.85	0.047	-2.91	0.004	0.762-0.949

Table 3: Approximate misclassification error for training (75%) data and test (25%) data.

Misclassification Error	
Training Error	20.95%
Test Error	21.25%

4. CONCLUSION

Since BDHS (2011) data were collected under two stage stratified sampling, cluster correction is necessary for any analysis using this data set. The prediction performance of the logistic regression model is good enough after cluster correction since the training and test errors are approximately 20.95% and 21.25% respectively. From our analysis we can conclude that the policy makers of the government should take necessary steps to increase the social awareness of the rural people, to increase education rate, to increase female employment and to ensure to food security of the people for reducing early marriages of our country.

5. ACKNOWLEDGEMENT

Special thanks to Nishith Kumar, assistant professor, department of Statistics, Bangabandhu Sheikh Mujibur Rahman Science and Technology University Gopalganj, Bangladesh for his inspiration to do this type of work.

6. REFERENCES

1. Bosch, A. M, F. J. Willekens, A. Baqui, J. K. S. Van Ginneken and I. Hutter. 2008. “Association between age at menerche and early life nutritional status in rural Bangladesh.” Journal of Biosocial Science, 40: 223-237.

2. Hango, D.W., & Bourdais, C.L. (2007) 'Early union formation in Canada: links with education', European Journal of Population, vol. 23, no. (3-4), pp.339-368.
3. S. M. Mostafa kamal, Che Hashim Hassan, Gazi Mahabubul Alam and Yang Ying Child Marriage in Bangladesh: Trends and Determinants. Journal of Biosocial Science, Available on CJO 2014 doi: 10.1017/S0021932013000746
4. Abdul Goni and Mustafizur Rahman (2012), Age at First Marriage in Bangladesh: Socioeconomic Differentials and Determinants. Middle East Journal of Age and Ageing Volume 7, Issue 5.
5. Haque A, Hossain T, and Nasser M (2015): Predicting the Number of Children Ever Born Using Logistic Regression Model. Biom Biostat Int J 2(4): 00034. DOI: 10.15406/bbij.2015.02.00034.
6. Martin E. Palamuleni, (2011): Socioeconomic determinants of age at marriage in Malawi, International Journal of Sociology and Anthropology Vol. 3(7), pp. 224-235.
7. Bursac Z, Gauss CH, Williams DK, Hosmer DW (2008) Purposeful selection of variables in logistic regression. Source Code Biol Med 3:17.