Modeling and Predicting Cumulative Population of Bangladesh

Md. Rafiqul Islam

Dept. of Population Science and Human Resource Development, University of Rajshahi, Bangladesh

Abstract   The purpose of the present study is to build up some mathematical models for male and female forward cumulative population of Bangladesh. For this, the secondary data for population of Bangladesh have been taken from census of 2001. To check up the validity of the models, the model validation technique, cross-validity prediction power (CVPP), is employed in this paper. It is seen that male and female forward cumulative population of Bangladesh follow four parameters polynomial model. It is found that the parameters of these fitted models are highly significant with large proportion of variance explained. Moreover, these models are stable more than 99%. The stability of R^2 of these models is more than 99%. Hence, the fittings of these models are well.

Keywords   Male and Female Population of Bangladesh, Polynomial Model, Cross- Validity Prediction Power (CVPP), T-Test, F-Test

1. Introduction

In Bangladesh, population census usually carried out at ten year interval of time is the most important source of data on size, structure and distribution of population by region or any other demographic and socio-economic phenomena. The Government and Non-government organizations do require the up-to-the-minute information on population. As a consequent, one can easily utilize to formulate realistic devices for development for the present as well as future time. The Government along with demographers is not only the users of population but also trade unions, social organizations, university and other social research institutes and centers, market research analysis, housing developers and business communities frequently do need these estimates for their individual purposes. Population or population by region, religion, age and sex is used to serve as the denominator for the estimation of socio-demographic, health and development related indicators. For that reason, population should be accurate as well as error free for the estimation of different indicators. But not only in Bangladesh but also in any developing countries census population are tremendously affected by various types of errors. So, these might be removed by fitting mathematical model. Islam (2003) mentioned that age structure for male, female and both sexes population of Bangladesh follows either negative exponential model or modified negative exponential model. Islam et al. (2003)

reported that age related population for male of Bangladesh in 1991 follows modified negative exponential model. Islam (2004) showed that proportion of married women of Bangladesh in the reproductive life span follow 3rd degree polynomial model. It was observed that age structure for population of both sexes of Bangladesh follows negative exponential model (Islam et al., 2005). In Islam (2007), it was investigated that the age specific marital fertility rates (ASMFRs) in rural area of Bangladesh follows 2nd degree polynomial model and simple linear regression model where as forward cumulative ASMFRs follows 3rd degree polynomial model.

At this juncture, an attempt would be made to mull over the pattern for cumulative population of Bangladesh in 2001 census using mathematical model. Therefore, the fundamental objectives of this study are as follows:

i) to build up some mathematical models for male and female forward cumulative population of Bangladesh in 2001 census, and

ii) to apply CVPP to test the adequacy of the model.

This paper is divided into five divisions. Introduction is included in the first section. Second section is data and data source of the study. Section three contains methods and various methodological issues in which data smoothing, model building, validation technique of model and F-test are described. Results of model fittings and discussion are narrated in section four. As a final point, conclusion of this study is in division five.

2. Data and Data Source of the Study

To fulfill the objectives mentioned in the introduction

section of this manuscript, the secondary data on population for male, female and both sexes of Bangladesh by age group in years have been taken from 2001 census (BBS, 2003). These have been utilized as raw materials in the present study and shown in Table 1.

3. Methods and Various Methodological Issues

3.1. Data Smoothing

The cumulative population for male, female and both sexes by age group in years smoothed before going to fit the mathematical model to this data set. For this, using the Package Minitab Release 12.1 by the latest smoothing method named "4253H, twice" (Velleman, 1980) is employed here. After that, the smoothed data are used to fit model and these smoothed data have been launched in Table 1.

3.2. Model Building

If the forward cumulative distribution of population by age group is plotted in the graph paper then it can be seen that cumulative distribution of population are distributed by polynomial model in terms of ages. Therefore, an nth degree polynomial model is treated and the structural formation of the model is addressed by

$$y = a_0 + \sum_{i=1}^{n} a_i\, x^i + u \quad \text{(Montgomery and Peck, 1982)}$$

Where, x is the mean value of the age group; y is forward cumulative distribution of population; a_0 is the constant; a_i is the coefficient of x^i (i =1, 2, 3, ..., n) and u is the stochastic error term of the model. Here, a suitable n is chosen such that the error sum of square is minimum.

3.3. Validation Technique of Model

To probe the validity of these models, the CVPP, ρ_{cv}^2, is applied. The mathematical formula for CVPP is known by

$$\rho_{cv}^2 = 1 - \frac{(n-1)(n-2)(n+1)}{n(n-k-1)(n-k-2)}(1-R^2)$$

Where, n is the number of classes, k is the number of explanatory variables in the model and the cross-validated R is the correlation between observed and predicted values of the dependent variable (Stevens, 1996). The shrinkage of the model is equivalent to $\left| \rho_{cv}^2 - R^2 \right|$; where ρ_{cv}^2 is CVPP and R^2 is the coefficient of determination of the model. Moreover, 1-shrinkage is the stability of R^2 of the model. The estimated CVPP analogous to their R^2 and information on model fittings are presented in Table 2. It was informed that CVPP was also employed as validation method by Islam (2003 and 2007) and Islam et al. (2003 and 2005).

Table 1. Observed Population for Male, Female and Both Sexes of Bangladesh in 2001 Census by Age Group and Their Forward Cumulative Distribution (in thousands)

Age Group	Observed			Smoothed			Predicted		
	Male	Female	Both Sexes	Male	Female	Both Sexes	Male	Female	Both Sexes
0-4	8362	7724	16086	8425	7903	16273	8518	7264	15775
5-9	8822	7956	16778	16728	15658	32259	16789	15833	32452
10-14	8420	7432	15852	24437	22933	47105	24182	23421	47297
15-19	6292	5672	11963	31009	29532	60117	30748	30087	60418
20-24	4859	6057	10916	36470	35541	71497	36535	35891	71922
25-29	4895	5865	10760	41272	40972	81673	41594	40893	81915
30-34	4313	4436	8748	45681	45603	90644	45975	45152	90506
35-39	4204	3795	7998	49660	49278	98259	49726	48730	97800
40-44	3426	2774	6200	53051	52081	104445	52898	51685	103906
45-49	2610	1991	4601	55772	54221	109305	55540	54078	108930
50-54	2175	1826	4002	57880	55897	113099	57703	55969	112979
55-59	1309	1047	2356	59494	57209	116034	59435	57416	116161
60-64	1529	1300	2829	60732	58198	118254	60787	58481	118583
65-69	814	629	1443	61686	58937	119941	61808	59223	120351
70-74	926	700	1626	62417	59513	121244	62548	59702	121573
75-79	358	258	616	62994	59994	122302	63057	59978	122356
80+	581	496	1076	63519	60446	123280	63383	60111	122807

Table 2. Information on Model Fittings and Corresponding CVPP of These Fitted Models

Models	n	K	R^2	ρ_{cv}^2	Shrinkage	Parameters	p-value	Cal. F	Tab. F (at 1% level)
1	17	3	0.99989	0.999842	0.0000478	a_0 a_1 a_2 a_3	0.000 0.000 0.00000 0.000000	39389.61	5.74 with (3, 13) d.f.
2	17	3	0.99953	0.999326	0.0002044	a_0 a_1 a_2 a_3	0.000 0.000 0.0000 0.00000	9215.53	5.74 with (3, 13) d.f.
3	17	3	0.99989	0.999842	0.0000478	a_0 a_1 a_2 a_3	0.000 0.000 0.0000 0.00000	39389.61	5.74 with (3, 13) d.f.

3.4. F-test

To find out the measure of overall significance level of the fitted models as well as the significance of R^2, the F-test is employed here. The F-test is given by $F = \dfrac{R^2/(l-1)}{(1-R^2)/(n-l)}$ with (l-1, n-l) degrees of freedom (d.f.).

Where, l = the number of parameters is to be estimated, n is the number of cases and R^2 is the coefficient of determination of the model (Gujarati, 1998).

4. Results of Model Fittings and Discussion

The fitted model for forward cumulative distribution of population for male of Bangladesh in 2001 is:

$$y = 4038.093 + 1839.170x - 19.0402x^2 + 0.06626x^3 \qquad (1)$$

t-stats   (20.533)   (91.438)   (-34.6107)   (15.56793)

The fitted model for forward cumulative population for female of Bangladesh in 2001 is:

$$y = 2592.746 + 1921.665x - 21.4303x^2 + 0.079857x^3 \qquad (2)$$

t-stats   (6.440)   (46.668)   (-19.0283)   (9.164796)

The fitted model for forward cumulative distribution of population for both sexes of Bangladesh in 2001 is:

$$y = 6715.965 + 3722.203x - 39.8433x^2 + 0.14282x^3 \qquad (3)$$

t-stats   (17.505)   (94.860)   (-37.1255)   (17.20011)

The information on model fittings and estimated CVPP, ρ_{cv}^2, corresponding to their R^2 of these models is shown in Table 2. From this table it appears that all the fitted models (1) - (3) are highly cross- validated and their shrinkages are 0.0000478, 0.0002044 and 0.0002044 respectively. Moreover, it is found that the parameters of the fitted model (1) - (3) are highly statistically significant with large proportion of variance explained. Moreover, these models are stable more than 99%. The stability of R^2 of these models is more than 99%.

The calculated values of F statistic for the models (1) - (3) are 39389.61, 9215.53 and 39389.61 with (3, 13) d.f. respectively whereas the analogous tabulated values are only 5.74 at 1% level of significance. Therefore, from these statistics it is seen that these models and their analogous R^2 are highly statistically significant. Hence, the fits of these models are well.

5. Conclusions

In this study, it is observed that forward cumulative distribution of male, female and both sexes population follow 3rd degree polynomial model, i. e., cubic polynomial models containing four parameters. Hope one might be used these predicted population of Bangladesh in 2001 census for further higher study as more smoothened as well as more reliable data than that of observed data aggregate.

REFERENCES

[1]   BBS (2003). Bangladesh Population Census 2001, National Report, Government of the People's Republic of Bangladesh, Dhaka

[2]   Gujarati, Damodar N. (1998). Basic Econometrics, Third Edition, McGraw Hill, Inc., New York

[3]   Islam, Md. Rafiqul (2003). Modeling of Demographic Parameters of Bangladesh-An Empirical Forecasting, Unpublished Ph.D. Thesis, Rajshahi University

[4]   Islam, Md. Rafiqul, Islam, Md. Nurul, Ali, Md. Ayub & Mostofa, Md. Golam (2003). Construction of Male Life Table from Female Widowed Information of Bangladesh, International Journal of Statistical Sciences, Vol. 2, Dept. of Statistics, University of Rajshahi, Bangladesh, Page 69-82

[5]   Islam, Md. Rafiqul (2004). Indirect Estimation of fertility Parameters of Bangladesh, Journal of Indian Anthrop. Soc. Vol.39, No. 2, Page 195-202

[6]   Islam, Md. Rafiqul, Islam, Md. Nurul, Ali, M. Korban & Mondal, Md. Nazrul Islam(2005). Indirect Estimation and Mathematical Modeling of Some Demographic Parameters of Bangladesh, The Oriental Anthropologist, Vol. 5(2), Page 163 - 171

[7]   Islam, Md. Rafiqul (2007). Mathematical Modeling of Age Specific Marital Fertility Rates (ASMFRs) in Rural Area, Bangladesh, Middle East J. of Age and Ageing, Vol. 4(4)

[8]   Stevens, J. (1996). Applied Multivariate Statistics for the Social Sciences, Third Edition, Lawrence Erlbaum Associates, Inc., Publishers, New Jersey

[9]   Velleman, P. F. (1980). Definition and Comparison of Robust Nonlinear Data Smoothing Algorithms, Journal of the American Statistical Association, Volume 75. Number 371, Page 609-615