Analysis of Age-Specific Fertility Rates: A Polynomial Model Approach

***[1]Gowher Ahmad Wani, [2]Peer Javaid Ahmad and [3]Nisar Ahmad Rather**
*[1, 3] Department of Statistics, Govt. Degree College Uri, Baramulla, J&K
[2]Govt. MVM Bhopal MP India
Email: *gowherwani413@gmail.com
[2]syedjavaid111@gmail.com
[3]nisarmphil@gmail.com

Abstract
In the present world fertility has become a long-standing issue among demographers. Although a lot of work has been done on such burning issue but research regarding the fertility measure in Jammu and Kashmir is yet to be appalling. So, in the present study analysis of Age-Specific Fertility Rate (ASFR) of Jammu and Kashmir has been performed by applying Polynomial regression approach. The second degree polynomial has been used to fit the data after observing scatter plot. To test the stability and validity of fitted regression models, cross-validation prediction power, shrinkage, F-test and the coefficient of determination have been estimated. All the regression models suggest that the proposed models are significantly fit to evaluate the patterns of ASFR.
Keywords: Polynomial regression, Fertility, J&K and coefficient of determination

1. Introduction

Fertility is considered as one of the main demographic feature for any population. The development of a human population depends exclusively on human fertility. Based on the total number of live births to a woman, fertility represents the actual level of reproduction of a population. Fertility levels are the determinant of the age structure of population which intern governs the demographic and socioeconomic distinctiveness of the population. To know the levels of fertility various methods are applied either direct or indirect. The mathematical functions (models) are applied to evaluate the patterns of fertility for any population that reflect the true picture of it. The polynomial regression model used in this study takes mid-point of age-group as independent variable and ASFR as outcome variable. Apart from this the inverse polynomial regression model is used to estimate ASFR. Among the two models the best fitted model which gives the better results has been recommended for further analysis.

In demography, fertility is influenced by several factors and therefore its modeling and estimating future fertility is more complicated than that of mortality [1]. In any nation, the population and improving of living standard is closely related to postponement of first childbirth to the later age [2, 3]. The mathematical modeling helps in population projection by analyzing fertility and plays a vital role to measure the strength of population growth. ASFR is an important component required to know the levels and patterns of fertility. It is one of the simplest methods to analyze the patterns of fertility. Methods for analyzing age-specific demographic data is alienated into two categories: one considers age as a discrete variable in real space R^p (p signify the total number of ages) while as the other considers age a continuous variable in function space bounded within a finite interval. An extensive work has been

The Philippine Statistician Vol. 69, No. 1 (2020), pp. 08-14

performed on modeling and predicting age-specific fertility and mortality by taking into consideration age as a discrete variable, [4, 5, 6, 7], where the demographic rates at various ages are measured as discrete data points. On the other hand, the methods that consider age as a continuous variable have pioneered functional data analysis to model and predict age-specific demographic rates at a given year as a constant and smooth function and can later be changed to discrete ages at any sampling interval, [8, 9, 10, 11].

2. Methodology

2.1 Polynomial

A mathematical expression of the form:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n \qquad a_n \neq 0$$

Where, a_i is the coefficient of x^i (i=1, 2, 3 … n) but $a_1, a_2, a_3 \ldots a_n$ are also constants, a_0 is the constant term and n is a positive integer, is called a polynomial of degree n and the symbol x, in this case, is called an indeterminate. If $n = 0$, then it becomes constant function, if $n = 1$, then it is called polynomial of degree 1 i.e. simple linear function, if $n = 2$, then it is called polynomial of degree 2 i.e. quadratic polynomial, etc [12].

To study the patterns of fertility, scatter plot of age-specific fertility rates for state Jammu and Kashmir has been plotted. The scatter plot that has been fitted for the observed data of both the state shows that age-specific fertility rates can be fitted by a polynomial model with respect to age. Therefore, keeping in view the nature of data different models that have been used to predict the age-specific fertility rates for Jammu and Kashmir shown below:

Model I: The third degree polynomial with respect "x"

$$y = a + bx + cx^2 + dx^3$$

Model II: The third degree polynomial with respect "(1/x)"

$$y = a + \frac{b}{x} + \frac{c}{x^2} + \frac{d}{x^3}$$

Where "y" represents the age-specific fertility rates and "x" is the mid-point of age-interval. Here two models have been discussed the Model I have been designed to explain the fertility pattern, in which ASFR acts as dependent variable represented by "y" and the independent variable, mid-point of age-interval is represented by "x". Modification of the model has been performed and the reciprocal of mid-point of age-intervals (1/x) instead of "x" has been used. Here again "y" acts as the dependent variable and the predictor here is reciprocal of mid-point of age-interval (1/x).

The present study analyzes the fertility behavior of females of Jammu and Kashmir by using both the models. The estimation of ASFR, as well as the model fittings for the above discussed models, has been obtained with the help of SPSS V.16.0 and MS-Excel 2007.

2.2 Model Validation

It is necessary to check the adequacy of the model to know how accurate the model will perform. The cross validity prediction power (CVPP) is used to verify the stability of the models given by [13]

$$\rho_v^2 = 1 - \frac{(n^2 - 1)(n - 2)}{n(n - p - 1)(n - p - 2)}(1 - r^2)$$

Where "n" is the number of cases, "p" is the number of regressors or predictors and "r" is the correlation between observed and predicted values of the dependent variable.

2.3 Shrinkage

To compensate for the subjective effect of further sampling, the standard adjustment is made in the coefficient of determination; the shrinkage of the model is given as:

$$Shrinkage = |\rho_v^2 - r^2|$$

Where ρ_v^2 is the cross validity prediction power (CVPP) and r^2 is the coefficient of determination of the model. Moreover, the stability of r^2 of the model is equal to (1-shrinkage).

2.4 F-Test

The F-test is applied to check the overall measure of the significance of the model as well as the significance of R^2. The formula for F-test is given by:

$$F = \frac{R^2/(p-1)}{(1-R^2)/(n-p)}$$

Where "n" = the number of cases and "p" = the number of parameters to be estimated, and R^2 is the coefficient of determination [14].

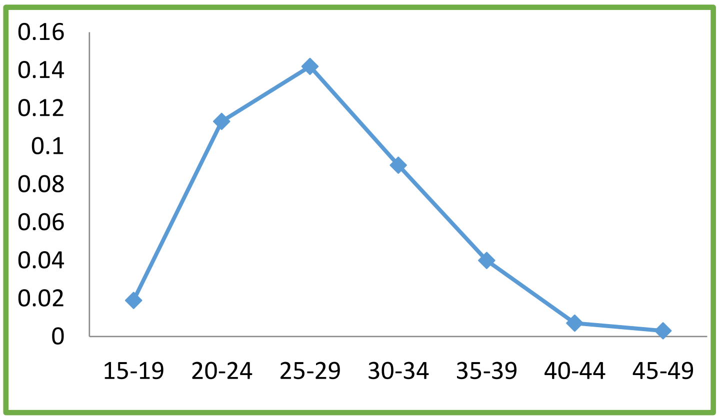
3. Data analysis and interpretation

3.1 Modeling of ASFR for J&K

To analyze the fertility pattern of J&K the estimates of ASFR for the females of Jammu and Kashmir has been taken from NFHS-4 (2015-16). The highest fertility is found in the age group of women (25-29) years in each age-group and the lowest fertility is found in the age group of women (45-49) years in each age-group (Table 1).

Age-group	Total
15-19	0.019
20-24	0.113
25-29	0.142
30-34	0.09
35-39	0.04
40-44	0.007
45-49	0.003

Table 1. ASFR of females of J&K (NFHS-4, 2015-16)



Source: NFHS-4 (2015-16)

Figure 1. Graphical representation of ASFR of J&K

ASFR slowly increases from 15 years age and attain the peak value in the age-group (25-29) years and afterwards it declines sharply and approaches to zero as age of females approaches to 50 years (Figure 1).

3.2 Analysis of data for the state of J&K

For the given data, Model I which is a third degree polynomial has been fitted to the observed data and coefficient of determination was obtained as 0.989. This implies that the Model I explain about 98.8% variation of ASFR of the females of Jammu and Kashmir. Also, the adjusted coefficient of determination was found to be 0.978 and F-value for the same Model was 91.9892.

The equation of the curve is given by:

$$y = -1.441 + 0.149x - 0.004x^2 - 4.27E\text{-}05x^3 \qquad (1)$$

With Adj. R^2=0.978

The scatter plot of curve (1) in which mid-point is shown along x-axis and ASFR along y-axis clearly represents that the third degree polynomial fits the data (Figure 2).

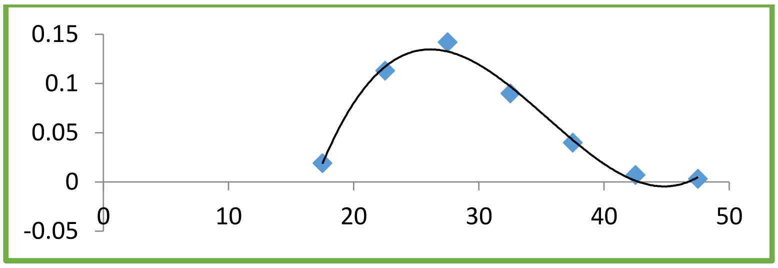


Figure 2. Scatter plot between ASFR and mid-point for J&K, NFHS-4

Again we will consider a third degree polynomial, but here instead of variable "x" which is defined in Model 1, the reciprocal of the variable "x" i.e. (1/x) is considered. In this model, predicting function is a third degree polynomial of inverse of the mid-point of age-group (1/x) and the outcome variable is age-specific fertility rate and is shown by "y". After applying Model IV on the data the coefficient of determination comes out to be 0.938 which means that Model II explains 93.8% variation of ASFR. Also the adjusted coefficient of determination was found to be 0.877 with F-value 15.2928.

The equation of the curve is given by:

$$y = -0.339 + 17.316\left(\frac{1}{x}\right) + 13.012\left(\frac{1}{x^2}\right) - 316.419\left(\frac{1}{x^3}\right) \qquad (2)$$

With Adj. R^2=0.877

The scatter plot of model (2) clearly shows that the inverse third degree polynomial fits the data (Figure 3).

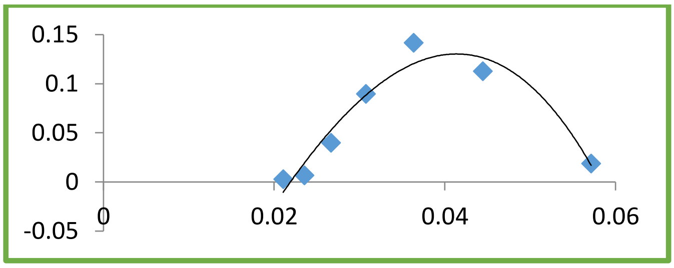


Figure 3. Scatter plot between ASFR and inverse mid-point for J&K, NFHS-4

The predicted values of ASFR shows that in age-group (25-29) years, highest fertility has been obtained by using Model I and if the Model II is used the highest values of ASFR has been obtained in the age-group (20-24) years. Also, both the models explain the lowest value of ASFR in the age-group (40-44) years. The ASFR for the age-group (40-44) years is not explained if Model II is applied (Table 2).

			Model I: $y=a+bx+cx^2+dx^3$	**Model II: $y=a+(b/x)+(c/x^2)+(d/x^3)$**
Age-interval (in years)	**Mid-Point**	**Observed ASFR**	**Predicted ASFR**	**Predicted ASFR**
15-19	17.5	0.019	0.0185	0.0169
20-24	22.5	0.113	0.117	0.1264
25-29	27.5	0.142	0.1325	0.1205
30-34	32.5	0.09	0.097	0.0885
35-39	37.5	0.04	0.0426	0.0527
40-44	42.5	0.007	0.0012	0.0192
45-49	47.5	0.003	0.0049	_

Table 2. Predicted values of ASFR using Model I and Model II for the state of J&K

The observed curve and the curve obtained by applying Model I attain highest peak in the age-interval (25-29) years. While as the curve obtained by applying Model II attains highest peak in the age-interval (20-24) years. The predicted values obtained by Model I are more accurate than the values obtained by Model II (Figure 4).

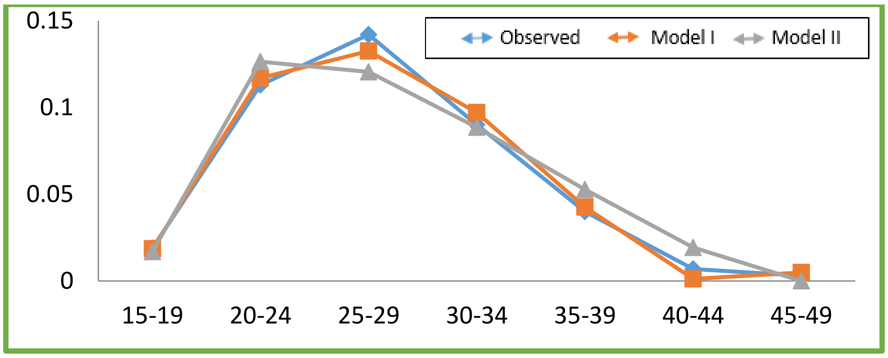


Figure 4. Plot of ASFR for J&K and Curve Fitting of Model I and Model II

Models For		**n**	p	**r^2**	**ρ_v^2**	**Shrinkage**
ASFR	Model I	7	3	0.9892	0.9382	0.051
	Model II	7	3	0.9386	0.6491	0.2895

Table 3. Estimated cross validity prediction power (CVPP) ρ_v^2 and shrinkage of the predicted models

4. Conclusion

The findings of the study revealed that both polynomial and inverse polynomial models comes out to be significant in J&K state. The model having less shrinkage gives the estimated values close to the observed values. Therefore the model with minimum shrinkage is recommended to carry out the fertility changes over age of females.

5. References:

1. Simpach, O., & Langhamrová, J. (2014). Development of Socio-Economic Indicators and Mortality Rates during Ten Years of the CR Membership in the EU. *ICEI 2014*, 675-683.
2. BÍLKOVÁ, D. (2009). Pareto Distribution and Wage Models. Bratislava 03.02. 2009–06.02. 2009. *Aplimat 2009*, 723-732.
3. Simpach, O., & Pechrová, M. (2013). Assessing the impact of standard of living on the life expectancy at birth using Vector Autoregressive Model. *Proceedings of Mathematical Methods in Economics*, *2013*, 921-926.
4. Bell, W. (1992). ARIMA and principal component models in forecasting age-specific fertility.
5. Lee, R. D., & Carter, L. R. (1992). Modelling and forecasting US mortality. *Journal of the American statistical association*, *87*(419), 659-671.
6. Lee, R. D. (1993). Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level. International Journal of Forecasting, 9(2), 187-202.
7. Li, N., & Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. Demography, 42(3), 575-594.
8. Veeranki S. R, Varshney M. (2022). Intelligent Techniques and Comparative Performance Analysis of Liver Disease Prediction, International Journal of Mechanical Engineering, 7(1), 489-503. https://kalaharijournals.com/ijme-vol7-issue-jan2022part2.php,
9. Dr. Ritika Malik, Dr. Aarushi Kataria and Dr. Naveen Nandal, Analysis of Digital Wallets for Sustainability: A Comparative Analysis between Retailers and Customers, International Journal of Management, 11(7), 2020, pp. 358-370.
10. Hersh F. Mahmood, Hooshang Dabbagh, Azad A. Mohammed, Comparative study on using chemical and natural admixtures (grape and mulberry extracts) for concrete, Case Studies in Construction Materials, Volume 15, 2021.
11. Masood Abu-Bakr, Hersh F. Mahmood, Azad A. Mohammed, Investigation of metakaolin and steel fiber addition on some mechanical and durability properties of roller compacted concrete,
12. Case Studies in Construction Materials, Volume 16,2022, e01136.
13. Hyndman, R. J., & Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, *51*(10), 4942-4956.
14. D'Amato, V., Piscopo, G., & Russolillo, M. (2011). The mortality of the italian population: Smoothing techniques on the lee—carter model. *The Annals of Applied Statistics*, 705-724.
15. Hyndman, R. J., Booth, H., & Yasmeen, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, *50*(1), 261-283.
16. Shang, H. L. (2016). Mortality and life expectancy forecasting for a group of populations in developed countries: a multilevel functional data method. *The Annals of Applied Statistics*, *10*(3), 1639-1672.

17. Spiegel, M. R., & Constable, R. L. (1992). *Schaum's outline of theory and problems of statistics* (p. 504). McGraw-Hill.
18. Herzberg, P. A. (1969). The Parameters of Cross-Validation (Psychometric Monograph No. 16). Richmond, VA: Psychometric Society.
19. Gujarati, D. N. (1998). Basic Econometrics, Third Edition, *McGraw Hill, Inc.*, *New York*.