Demography India ISSN 0970-454X
Vol. 51, No. 2 (2022), pp.25-44

Some Probability Models for Age at Infant Death and Estimation Procedure

Brijesh P. Singh*[1], Shweta Dixit[2] and Deepak Kumar Maurya[3]

Abstract: Infant Mortality Rate (IMR) is one of the most sensitive and gloomy episodes in reproductive span, the levels and trends of infant mortality are usually taken as indicator of the development of a country, effectiveness of health programs and improvement in medical facilities. In this study, we proposed some probability models for the distribution of infant deaths by age. The suitability of the proposed model is explained through its application on the data of most populous state Uttar Pradesh and India, for all four rounds of National Family Health Survey (NFHS) i.e., NFHS-I to NFHS-IV. The goodness-of-fit and the value of parameters of proposed model are also compared. If the parameter decreases the proportion of neonatal death increases i.e., parameter is inversely proportional to neonatal death.

Introduction

One of the most important determinants of population growth is mortality. Decline in mortality has been quite rapid because of better medical facilities, but in developing countries, the rate of child and infant mortality is still very high in comparison to developed countries. Thus, the primary concern for these countries is to reduce the level of infant as well as child mortality. Infant and child mortality is an excellent measure of the level and quality of health care and other social activities prevailing in a population. A low infant and child mortality reflects availability of good health services and its proper utilization by community. Child loss is measured by the probability of dying between the first and fifth birthday, whereas infant mortality is measured by the probability of death before the first birthday. Indian government has launched several child survival programs over the past decades. Due to these programs, perhaps, initially a rapid decline in infant and child mortality was noted but during the last decade of twentieth century the rate of decline was very slow and at present.

Infant mortality (the number of deaths of children less than one year of age per 1,000 live births) is not only an important factor in population growth but it is also an important measure of economic development. As Infant Mortality Rate (IMR) is a sensitive issue, the levels and trends of infant mortality are often taken as indicators of the development of a country. Thus, among many available indicators of socio-economic development, IMR is used as a sensitive and powerful index of development. Reduction in IMR is also known to reduce the fertility as probability of survivorship of children increases and then fulfill the desire family size. Any research on IMR begins by setting up a model that contains risk factors in three domains: proximate factors, maternal factors, and household/community factors. These three factors are responsible for infant deaths. Proximate factors are those items that involve medical care and non-medical care during the antenatal period, care at birth, and care during the postnatal period. Maternal

[1] Centre of Population Studies, Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi
[2] Centre of Population Studies, Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi
[3] Centre of Population Studies, Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi

factors refer to such things as the age, and birth intervals of the mother and the third factor, household/community factors refer to such things as sanitation, water supply, and household and community cleanliness.

Children are important assets of a nation as well as of a family therefore reduction in infant and child mortality is likely to the most important objective of the Millennium Development Goals (MDG). IMR reflects the level of socio-economic development a country and quality of life and are used for monitoring an evaluating population health programs and policies. Birth and death, registered through vital registration system in the developing countries also suffer from age misreporting, omissions and under count. To overcome this defectiveness in data and to obtain reliable estimates of birth and death rates, India introduced Sample Registration System (SRS) in 1960 but still they suffer from considerable degree of errors. It is observed in retrospective surveys that events are highly misreported due to respondent's ignorance, misreporting due to recall lapse and digit preference of the respondents, which also distorts the distribution of deaths by age through infancy. Thus, the data on deaths collected, suffer from many defects. To overcome this problem and smooth the data, attempts have been made to develop and fit suitable models to data on age distribution of deaths.

During last fifty years, studies on early age death were mostly related to infant mortality. But it has been increasingly realized that child loss needs to be examined in addition to infant mortality. The most common problem in such studies is associated with the data of deaths during infancy and childhood which suffer from substantial degree of errors. In such situations development of probability models is perhaps the most appropriate way to minimize the effects of these errors. Actually, a probability model smoothes the data and provide a reasonable explanation of phenomenon under study. The representation of mortality data via a parametric model has attracted the attention of demographers and statisticians for over a century.

The most popular model was given by Gompertz (1825) to study the mortality, which is still used by demographers today. But very few studies related to the age patterns of mortality have developed models of mortality which are mathematical expressions for graduation of age patterns of mortality. They are mostly suitable to graduate rates by age after thirty. Theile (1972) has proposed a seven-parameter model to capture age pattern of mortality for the whole age range. He has considered three different factors, exponential in nature, representing young, adult and old age mortality pattern. Keyfitz (1977) might be the first person to attempt a study on infant mortality by using hyperbolic function. An attempt to represent mortality across the entire age range was eight parameter non-linear models of Heligman and Pollard (1980) and tested on Australian data at different points of time. They also gave three different factors like Theile (1972).

These models have been used in the past for a wide range of mortality data. Since force of mortality is relatively high at early stages of life than at teenage, so, graduating the survival function at these young ages have been attempted separately also. Later, for study of infant and child mortality, Weibull function was recommended by Choe (1981) and Hartman (1982) proposed a logarithmic approximation. Attracted by the ability of Weibull function in effectively describing the age pattern of mortality at early ages, Krishnamoorthy (1982), Krishnamoorthy and Mathew (1994) and Mathew and Krishnamoorthy (1995) have applied Weibull function in their works. Renshaw (1991) presented generalized linear and nonlinear model approach to mortality graduation and provided arguments in their favour. Krishnan and Jin (1993) fitted the Pareto distribution on national data of Canada for studying distribution of infant deaths. In this paper an attempt has been made to develop a model with an idea of the majority of infant deaths occurs within the first month of their life, which takes into account the nature of declining tempo of deaths by age during the first year of life. The model is used to give a functional shape to the phenomenon of infant deaths distribution at different ages and apply on real data taken from National Family Health Survey. Infant and Child mortality has been a matter of concern for researchers because these are directly related to the fertility and have also indirect relation with use of contraceptives (Kabir et al., 1993). The data on infant and child mortality is observed in retrospective surveys and these types of data are highly misreported due to respondent's ignorance, misreporting due to recall lapse and digit preference of the respondents, which affects the distribution of actual data. Thus, the data on deaths collected, suffering from various defects. Several attempts have been made to develop and fit suitable models to data on age distribution of deaths, to overcome these difficulties. Models smoothes the data and gives the realistic nature of the data.

A few numbers of mortality models are available in literature to study the age pattern of mortality (Gompertz, 1825; Mitra and Denny, 1994; Perks, 1932). Chen (1977) derived an exact confidence interval for parameter C and an exact joint confidence region for its parameters. In spite of Gompertz (1825), various models have been proposed to explain the survival and mortality trend. Bhuyan and Degraties (1999) used the Polya-Aeppli model to explain the pattern of child mortality, by considering the number of child death in the household. Singh et al. (2011) used the Beta-Binomial and Singh et al. (2012) used the inflated Binomial model to study the number of child deaths for a fixed parity. Singh et al. (2015) proposed an inflated geometric model to analyze the infant deaths by age. Singh et al. (2021) developed a single parameter discrete probability model to explain the pattern of child mortality. In previous studies, all these models are used for a wide range of mortality data. The force of mortality is comparatively higher at early stages of life and in old age than the other stages of life, therefore the surviving at the early stages of life have been also study separately. Krishnamoorthy and Ranjna (1999) also examined the suitability of this model and made some modification in it to graduate the survivorship function.

Objectives and Methodology

As discussed above that in developing countries that infant deaths registration is a subject to error of misreporting which distorts the distribution of deaths by age during infancy. Therefore, a method for distribution of infant deaths during the first year of life is important to get a clear picture on the number of deaths at various age points under the first year of life. Keeping in view the above discussion an attempt has been made here in this paper is to develop some probability models for distribution of deaths by age within infancy.

Sources of Data

The data used in this paper have been used from National Family Health Survey (NFHS). The National Family Health Survey (NFHS) is a large-scale, multi-round survey conducted in a representative sample of households throughout India. The NFHS is a collaborative project of the International Institute for Population Sciences (IIPS), Mumbai, India. The First National Family Health Survey (NFHS-I) was conducted in 1992-93. While the Second National Family Health Survey (NFHS-II) and Third National Family Health Survey (NFHS-III) was carried out in 1998-99, 2005-2006 respectively and the Fourth round of National Family Health Survey (NFHS-IV) was carried out in 2015-16.

Proposed Probability Model (Model-I& II)

Since we need a concise and clear representation of extensive data sets and as probability models provide that concise representation of data in a better way in recent years increased attention has been paid to the proposition and derivation of probability models for the distribution of infant mortality. Now the questions arise weather to use discrete or continuous probability distribution for the representation of this particular phenomenon. In this study we went for both types of distributions to see whichever gives the best picture of mortality in the desired age group of infancy.

It is worth to mention here that infant mortality ranges over 0 to 1 year, and the Beta distribution is a flexible nature of continuous distributions for the same range i.e., 0 to 1. Thus, we may propose this distribution to explain this phenomenon i.e. infant mortality. Thus let *X* is the age at death during infancy which follows Beta distribution of first kind (Model-I) having density function as follows;

$$f(x) = \frac{1}{B(a,b)}(x)^{a-1}(1-x)^{b-1}; \quad 0 < x < 1 \ \ \& \ \ a > 0, b > 0 \qquad (1)$$

Where *a* and *b* are shape parameters. This distribution is applied at first to explain the data and found adequately good. This distribution is complex and having two parameters, the parameter *b* is less sensitive parameter than the parameter *a*. Also, we are thinking about a distribution which explain the infant mortality as well and simple than the Beta distribution of first kind, keeping this into mind we fix the value of less sensitive parameter b=1 and thus a modified proposed distribution (Model-II) is found which as follows;

$$f(x) = ax^{(a-1)};\ 0 < x < 1\ \ \&\ \ a > 0 \tag{2}$$

This distribution has only one shape parameter a. Here, x represents the age at death of infants. This distribution is also known as Power function distribution. Which is the probability of deaths distributed among total infant deaths by age and is same as proposed distribution by Mukherji and Islam (1983) for reliability analysis and named as finite range model where range is 0 to 1. This model was further used by Chauhan (1997) to graduate the age distribution of early age at deaths in various states of India.

Now, equation (3) below gives cumulative distribution function (cdf). The cdf gives the cumulative proportion of deaths up to a desired age point 'x'

$$F(X < x) = x^a;\ 0<x<1\ \&\ a > 0 \tag{3}$$

Here, one thing should be taken care of while applying the model is that we assume there is no effect of seasons over deaths at different ages i.e., the deaths are independent of seasonal effect during the reference period.

Estimation of the Parameters

Common methods of estimation of the parameters of the beta distribution are maximum likelihood and method of moments. The maximum likelihood equations for the beta distribution have no closed-form solution; estimates may be found through the use of an iterative method. The method of moments estimators is more straight- forward and in closed form. We examine both of these estimators for proposed models here, along with one other proposed estimator for model-II.

Method of moment estimator (Model-I)

The method of moments estimators of the beta distribution parameters involves equating the moments of the beta distribution with the sample mean and variance (Bain and Engelhardt 1991).As the mean and variance of the given density is as follows,

$$Mean = E(x) = \frac{a}{(a+b)} \tag{4}$$

$$Second\ moment = E(x^2) = \frac{(a+1)a}{(a+b+1)(a+b)}$$

$$Variance = E(x^2) - [E(x)]^2 = \frac{(ab)}{(a+b)^2(a+b+1)} \tag{5}$$

Our method of moments estimators is found by setting the sample mean and variance equal to the population mean and variance

$$\bar{X} = \frac{a}{(a+b)} \tag{6}$$

$$S^2 = \frac{(ab)}{(a+b)^2(a+b+1)} \tag{7}$$

To obtain estimator of α and β, we solve the above equation 6 & 7 for a and b in terms of $\bar{X}$ and S^2First, we solve for b (in terms of α);

$$(a+b)\bar{X} = a \;\Rightarrow b\bar{X} = a - a\bar{X} \;\Rightarrow b = \frac{a}{\bar{X}} - a$$

Now solve for a:

$$ab = (a+b)^2(a+b+1)S^2$$

$$\Rightarrow a\left(\frac{a}{\bar{X}} - a\right) = \left(a + \frac{a}{\bar{X}} - a\right)^2\left(a + \frac{a}{\bar{X}} - a + 1\right)S^2$$

$$\Rightarrow a^2\left(\frac{1}{\bar{X}} - 1\right) = \left(\frac{a}{\bar{X}}\right)^2\left(\frac{a}{\bar{X}} + 1\right)S^2 \Rightarrow \left(\frac{1}{\bar{X}} - 1\right)\left(\frac{\bar{X}^2}{S^2}\right) = \left(\frac{a}{\bar{X}} + 1\right)$$

$$\Rightarrow \left(\frac{\bar{X}(1-\bar{X})}{S^2}\right) = \left(\frac{a}{\bar{X}} + 1\right) \Rightarrow a = \bar{X}\left(\frac{\bar{X}(1-\bar{X})}{S^2} - 1\right)$$

Finally, we express β in terms of $\bar{X}$ and S^2,

$$b = \frac{a}{\bar{X}} - a$$

$$\Rightarrow b = a\left(\frac{1-\bar{X}}{\bar{X}}\right) \Rightarrow b = \bar{X}\left(\frac{\bar{X}(1-\bar{X})}{S^2} - 1\right)\left(\frac{1-\bar{X}}{\bar{X}}\right) \Rightarrow b = (1-\bar{X})\left(\frac{\bar{X}(1-\bar{X})}{S^2} - 1\right)$$

Thus our method of moments (MOM) estimates of α and β are

$$\hat{a}_{MOM} = \bar{X}\left(\frac{\bar{X}(1-\bar{X})}{S^2} - 1\right) \tag{8}$$

$$\hat{b}_{MOM} = (1-\bar{X})\left(\frac{\bar{X}(1-\bar{X})}{S^2} - 1\right) \tag{9}$$

Method of maximum likelihood estimator (Model-I)

Another well-known method of estimating parameters is the maximum likelihood approach. We define the likelihood function for an iid sample $X_1,...,X_n$ from a population with pdf $f(x_i|\theta_1,...,\theta_k)$ as $L(\theta_1,...,\theta_k|x_1,...,x_n)=\prod_{i=1}^{n} f(x_i|\theta_1,...,\theta_k).$ The maximum likelihood estimator (MLE) is the parameter value for which the observed sample is most likely. Possible MLEs are solutions to $\frac{\partial}{\partial\theta_i}L(\theta|X)=0, i=1,...,k.$

We may verify that the points we find are maxima, as opposed to minima, by checking the second derivative of the likelihood function to make sure it is less than zero. Many times, it is easier to work with the log likelihood function, $LogL(\theta|X)$, as derivatives of sums are more appealing than derivatives of products (Casella and Berger 2002).MLEs are desirable estimators because they are consistent and asymptotically efficient; that is, they converge in probability to the parameter they are estimating and achieve the lower bound on variance.

The likelihood function for the beta distribution is

$$L(a,b|X)=\prod_{i=1}^{n}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x_i^{a-1}(1-x_i)^{b-1}=\left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right)^n\prod_{i=1}^{n}x_i^{a-1}(1-x_i)^{b-1}$$

Yielding the log likelihood

$$LogL(a,b|X)=n\log\left(\Gamma\left(a+b\right)\right)-n\log\left(\Gamma\left(a\right)\right)-n\log\left(\Gamma\left(b\right)\right)+(a-1)\sum_{i=1}^{n}\log(x_i)+(b-1)\sum_{i=1}^{n}\log(1-x_i)$$

To solve for our MLEs of $\alpha$and β we take the derivative of the log likelihood with respect to each parameter, set the partial derivatives equal to zero, and solve for $\hat{a}_{MLE}$ and $\hat{b}_{MLE}$:

$$\left.\begin{aligned}\frac{\partial}{\partial a}\log L(a,b|X)=\frac{n\Gamma'\left(a+b\right)}{\Gamma(a+b)}-\frac{n\Gamma'\left(a\right)}{\Gamma(a)}+\sum_{i=1}^{n}\log(x_i)=0\\ \frac{\partial}{\partial b}\log L(a,b|X)=\frac{n\Gamma'\left(a+b\right)}{\Gamma(a+b)}-\frac{n\Gamma'\left(b\right)}{\Gamma(b)}+\sum_{i=1}^{n}\log(1-x_i)=0\end{aligned}\right\} \qquad (10)$$

There is no closed-form solution to this system of equations, so we will solve for $\hat{a}_{MLE}$ and $\hat{b}_{MLE}$ iteratively, using the Newton-Raphson method, a tangent method for root finding.

Iteration Method or First cell frequency Method (Model-II)

For fitting the proposed model from data on total and neonatal infant deaths, the first necessary thing was estimation of the parameter. So, for that iterative method has been used and the fitting is tested.

Let D be the total number of infant deaths and N be the neonatal deaths reported in a particular area during the reference period (Neonatal mortality: The probability of dying in the first month of life and Infant mortality: The probability of dying before the first birthday). If R denotes the proportion of neonatal deaths among the total infant deaths the proportion can be given as,

$$R = N/D \tag{11}$$

Now here we are using the data available as "reported age of death annually". And it can be computed by Eqn. (2), the cumulative density function of the model. We can estimate the total proportion of deaths under the age points

$$F(X < x) = x^a \text{ ; } 0<x<1 \text{ \& } a > 0 \tag{12}$$

In equation (12) if we put x=1/12, then we get the total proportion of deaths up to first month of age i.e., neonatal deaths to total infant deaths.

$$F\left[X < \frac{1}{12}\right] = \left[\frac{1}{12}\right]^a \text{ ; } 0<x<1 \text{ \& } a > 0 \tag{13}$$

So, from equation (11) and (13), we can write;

$$R = \left[\frac{1}{12}\right]^a$$

Taking logarithmic on both the sides;

$$\log R = a \log\left[\frac{1}{12}\right]$$

solving this we get,

$$\hat{a} = \frac{\log R}{\log\left[\frac{1}{12}\right]} \tag{14}$$

By putting x=1/12,2/12,3/12,……,1 in equation (12) we get the cumulative proportion of deaths up to age of first month, second month etc. up to twelfth month or up to first birthday.

Therefore;

Proportion of deaths during a particular month of age

=proportion of death up to (x+1) - proportion of death up to x

$$= F(X < x+1) - F(X < x); \text{ x>1} \tag{15}$$

Proportion of deaths during first month of age $=F(X<1)$ (16)

By multiplying equation (16) with D we get estimated number of deaths for the first month;

i.e.

$$d_1 = D*F(X<1) \tag{17}$$

Similarly; the rest estimates can be obtained by;

$$d_x = D*\left[F(X<x+1)-F(X<x)\right];\ \text{x>1} \tag{18}$$

Method of moment estimator (Model-II)

As the mean and variance of the given density is as follows,

$$Mean = E(x) = \frac{a}{(a+1)}\ ; Median = M_d = \left(\frac{1}{2}\right)^{1/a}$$

$$Variance = E(x^2)-[E(x)]^2 = \frac{(a)}{(a+2)} - \left[\frac{(a)}{(a+1)}\right]^2 = \frac{(a)}{(a+1)^2(a+2)}$$

The basic principle for method of moments is to equate population moments (i.e., the means, variances etc. of the theoretical model) to corresponding sample moments (i.e., the means, variances etc. of the sample data observed) and solve for the parameter(s). So, the method of moments of the single parameter *a* will be,

$$\hat{a}_{mm} = \frac{E(x)}{1-E(x)} \tag{19}$$

As we have derived the expected number of deaths and observed number of deaths has already known to us, we can use the chi square goodness of fit to test whether the model fits the data.

Method of maximum likelihood estimator (Model-II)

For the maximum likelihood of the density (2) can be estimated by writing its likelihood equation and then differentiating it with respect to the single parameter α and after that equating it to zero. The density is given by

$$f(x) = (a)x^{(a-1)}\ ;\ 0<\text{x}<1\ \&\ 0<a<1$$

The likelihood function L,

$$L = \prod_{i=1}^{n}(a)x^{(a-1)} \Rightarrow L = (a)^n \prod_{i=1}^{n} x_i^{(a-1)}$$

Taking logarithmic,

$$LogL = n\log(a) + (a-1)\sum_{i=1}^{n}\log x_i = n\log(a) + a\sum_{i=1}^{n}\log(x_i) - \sum_{i=1}^{n}\log(x_i)$$

Then differentiating this with respect of α will be,

$$\frac{\partial L^*}{\partial a} = \frac{n}{a} + \sum_{i=1}^{n} \log(x_i)$$

On simplifying, the maximum likelihood estimate of α will be,

$$\hat{a}_{ml} = -\frac{n}{\sum_{i=1}^{n} \log(x_i)} \tag{20}$$

Goodness of Fit

Goodness of fit test procedures is intended to detect the existence of a significant difference between the observed frequency and the theoretical pattern of infant deaths by age within their first year of life. For testing the goodness of fit of the proposed model we have used Kolmogorov-Smirnov (K-S) (Kolmogorov, 1933; Smirnov, 1933) test for testing the hypothesis:

The Null hypothesis is given by

H_0 : Data follows the considered distribution.

Against the Alternative hypothesis

H_a : Data do not follow the considered distribution.

The Kolmogorov-Smirnov (K-S) test is based on the empirical cumulative distribution function (ECDF). Given N ordered data points $Y_1, Y_2, Y_3, \ldots, Y_n$ the ECDF is defined as:

$$E_N = \frac{n(i)}{N}$$

Where $n(i)$ is the number of points less than Y_i (the i^{th} ordered observation, ordered from smallest to largest). This is a step function that increases by $1/N$ at each ordered data point. The K-S test is based on the maximum distance between the fitted theoretical cumulative distribution function and the ECDF. The Kolmogorov-Smirnov test statistic is defined as:

$$D = \max_{1 \le i \le N} \left| F(Y_i) - \frac{n(i)}{N} \right|$$

Where, $F(Y)$ is the theoretical cumulative distribution. The hypothesis regarding the distributional form is rejected if the test statistic D is greater than the critical value obtained from the table. At5% level of significance the critical value of K-S test for large sample can be approximated by $\frac{1.36}{\sqrt{N}}$, where N is the sample size.

 ISSN 0970-454X

Findings

As discussed in the above section the models that has been validated here was the general beta distribution and proposed distribution and the validity of that model has been tested by applying it to the data of Uttar Pradesh and India as a whole. The findings of first model give us the idea of the inappropriateness of the data due to its lack of fitness and further the need of its modification as the other proposed model here. The validity of the proposed model is first testified here by fitting it at the Uttar Pradesh and India data of infant deaths by age in one-twelfth of a year which is obtained by all round of National Family Health Survey. The model fitted for this data at the given time period by first estimating the parameters and later on estimating the expected deaths has been explained above.

In the table 1-8 the first column is mentioned for the age at which infant died which is measured in years, second column is representing the number of observed deaths. In the last row some characteristics has been given which is the proportion of neonatal deaths among total infant deaths,$\hat{a}$ value and corresponding calculated with the corresponding degree of freedoms. The third, fourth, fifth and sixth column shows the expected number of deaths from different method of estimation for model I and II respectively. In the last row some characteristics has been given i.e.,$\hat{a}$and $\hat{b}$values and corresponding calculated chi-square values with their respective *p*-values.

The proposed model is fitted for some real data taken from all round of NFHS. The observed and estimated values and the estimates of parameters are shown in tables 1-9. Table 1, shows the pattern of infant mortality in Uttar Pradesh during first round of NFHS (1992-93), from this table we observed that the expected frequencies for Model-II based on different estimation procedures such as iteration method, ML estimation and method of moments, are found to be very close to observed frequencies than the expected frequencies obtained through Model-I. Both the Models I and II gives the over estimates in correspondence to observed frequencies, instead of estimates obtained through iterative method for Model-II. ML estimates of Model-II gives the closer fit to the observed one. The value of KS-test statistics is found to be lower for Model-II than the Model-I, which indicates that the Model-II provide the better fit to the observed one which is supported by the *p*-value.

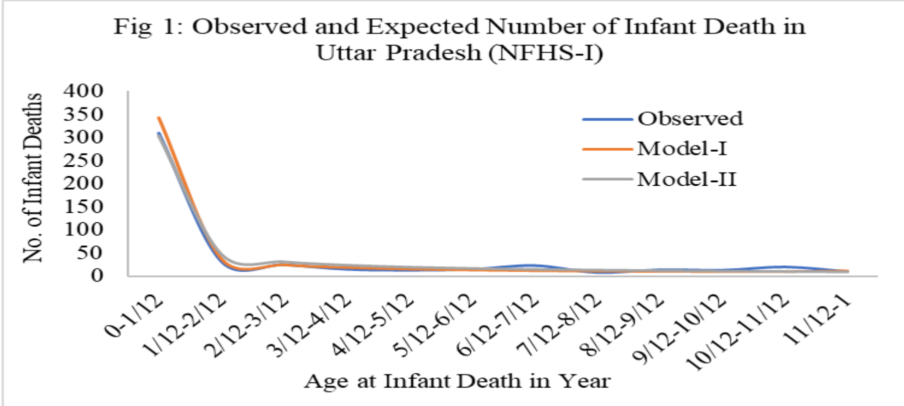


Table 1: Observed and Expected Number of Infant Death in Uttar Pradesh (NFHS-I)

Age at Death (In years)		Observed Number of Infant Death	Expected Number of Infant Death: Model-I MM estimates	Model-II MM estimates	Model-II Iteration estimates	Model-II ML estimates
0-1/12		310	342.68	303.33	310.00	315.36
1/12-2/12		33	36.90	47.12	45.99	45.05
2/12-3/12		26	23.73	30.89	30.00	29.28
3/12-4/12		16	18.01	23.55	22.81	22.21
4/12-5/12		14	14.78	19.26	18.61	18.09
5/12-6/12		16	12.71	16.42	15.84	15.37
6/12-7/12		24	11.30	14.38	13.85	13.43
7/12-8/12		9	10.31	12.83	12.34	11.96
8/12-9/12		15	9.62	11.62	11.16	10.80
9/12-10/12		14	9.21	10.64	10.21	9.87
10/12-11/12		21	9.16	9.83	9.42	9.11
11/12-1		11	10.59	9.14	8.76	8.46
Total		509	509.00	509.00	509.00	509.00
Parameters		$\hat{a}$	0.162	0.208	0.200	0.193
		$\hat{b}$	0.730	-	-	-
K-S test		statistic	0.008	0.050	0.056	0.061
		p-value	0.003	0.149	0.081	0.044

Table 2, shows the pattern of infant mortality in India for NFHS-I (1992-93), here we also observed that the Model-II gives better fit to this data set than the Model-I. From this table it is found that the MM estimates of Model-I and ML and MM estimates of Model-II gives the over estimates for first two cells than the observed one, but we also observed that the Model-I gives over estimates than the Model-II through all estimation procedures. The value of statistics is also lower for the Model-II than the Model-I, means Model-II gives the better fit to the observed than the Model-I and the *p*-value is reveals the same.

 ISSN 0970-454X

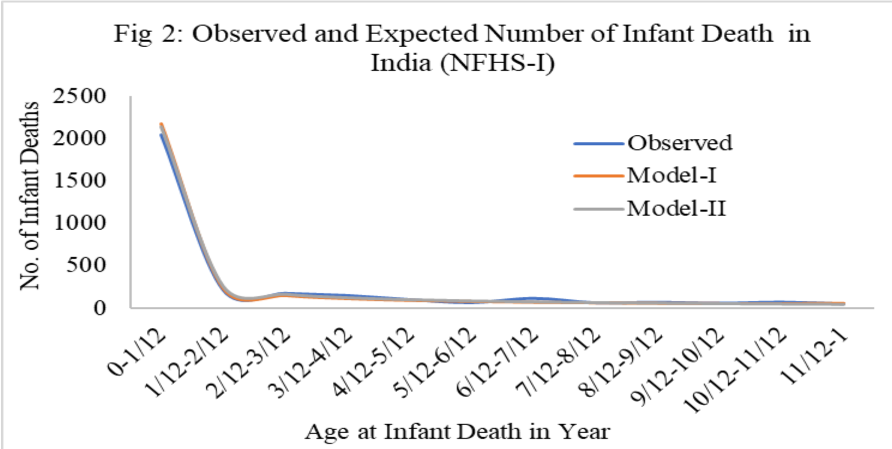


Table 2: Observed and Expected Number of Infant Death in India (NFHS-I)

Age at Death (In years)	**Observed Number of Infant Death**	**Expected Number of Infant Death**			
		Model-I	**Model-II**		
		MM estimates	**MM estimates**	**Iteration estimates**	**ML estimates**
0-1/12	2040	2175.75	2132.41	2040.00	2054.44
1/12-2/12	216	233.06	255.01	272.36	269.71
2/12-3/12	176	149.26	163.08	175.88	173.90
3/12-4/12	151	112.73	122.42	132.84	131.22
4/12-5/12	104	92.00	98.99	107.91	106.51
5/12-6/12	68	78.64	83.60	91.46	90.22
6/12-7/12	118	69.40	72.64	79.71	78.60
7/12-8/12	66	62.76	64.42	70.87	69.85
8/12-9/12	72	57.96	58.00	63.95	63.00
9/12-10/12	60	54.70	52.83	58.37	57.49
10/12-11/12	73	53.18	48.58	53.77	52.94
11/12-1	53	57.57	45.01	49.90	49.12
Total	3197	3197.00	3197.00	3197.00	3197.00
Parameters	$\hat{a}$	0.145	0.163	0.181	0.178
	$\hat{b}$	0.890	-	-	-
K-S test	statistic	0.048	0.041	0.020	0.022
	p-value	0.000	0.000	0.136	0.084

Table 3 and 5, represents the pattern of infant mortality in Uttar Pradesh, and table 4 and 6 represents the pattern in India for NFHS-II & III (1998-99 and 2005-06) respectively, from table 3 and 5, we observed that the estimates of Model-II obtained through iteration method and ML estimation are closer to observed one. While Model-I and Model-II (MM estimates) gives the slightly over estimates than the observed frequencies. We found that the estimates of Model-II are nearer to observed one than that of the Model-I. On the basis of the value of statistics of KS-test, we found that the Model-II provide the closer fit to the observed than the Model-I, which is also obtained from the p-value. From table 4 and 6, we also observed that the estimated frequencies obtained through Model-I is over estimated and Model-

II also gives over estimate through MM estimates just like Model-I but slightly less and closer to observed one. The estimates obtained from Model-II through iterative method and ML estimation are very close to observed frequencies, which is also supported by the value of statistics and p-value.

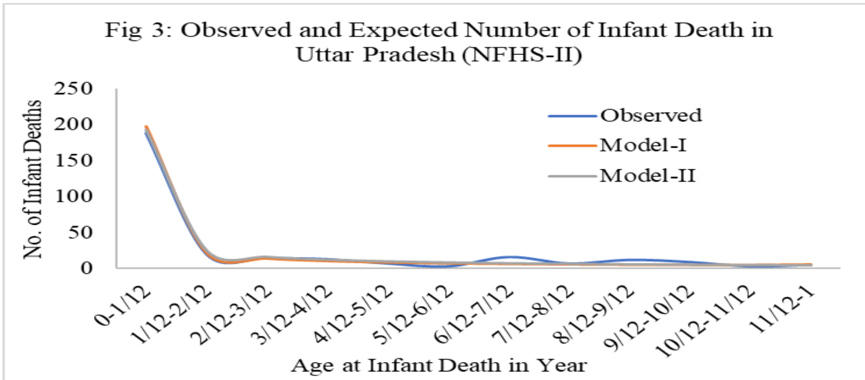
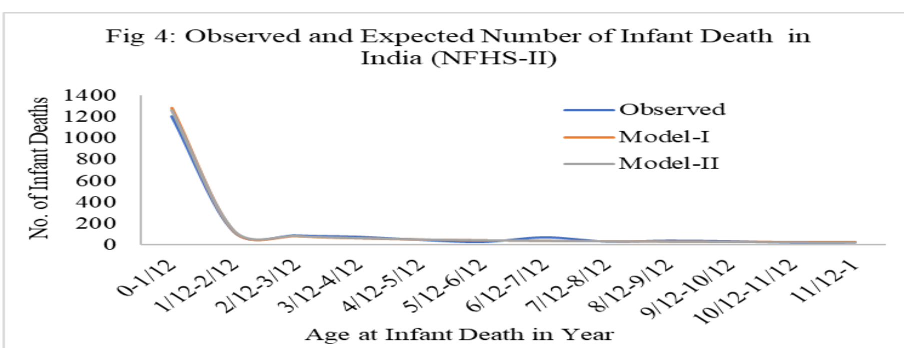


Table 3: Observed and Expected Number of Infant Death in Uttar Pradesh (NFHS-II)

Age at Death (In years)	Observed Number of Infant Death	Expected Number of Infant Death Model-I MM estimates	Model-II MM estimates	Model-II Iteration estimates	Model-II ML estimates
0-1/12	189	198.17	192.71	189.00	188.31
1/12-2/12	20	22.39	25.12	25.80	25.92
2/12-3/12	15	14.44	16.19	16.69	16.79
3/12-4/12	13	10.97	12.21	12.63	12.70
4/12-5/12	7	9.00	9.91	10.27	10.33
5/12-6/12	3	7.74	8.39	8.71	8.77
6/12-7/12	16	6.87	7.31	7.59	7.65
7/12-8/12	7	6.26	6.49	6.75	6.80
8/12-9/12	12	5.83	5.86	6.10	6.14
9/12-10/12	9	5.57	5.34	5.57	5.61
10/12-11/12	3	5.51	4.92	5.13	5.17
11/12-1	5	6.25	4.56	4.76	4.80
Total	299	299.00	299.00	299.00	299.00
Parameters	$\hat{a}$	0.152	0.177	0.185	0.186
	$\hat{b}$	0.859	-	-	-
K-S test	statistic	0.053	0.059	0.054	0.053
	p-value	0.362	0.249	0.343	0.364



ISSN 0970-454X

Table 4: Observed and Expected Number of Infant Death in India (NFHS-II)

Age at Death (In years)	Observed Number of Infant Death	Expected Number of Infant Death			
		Model-I	Model-II		
		MM estimates	MM estimates	Iteration estimates	ML estimates
0-1/12	1203	1279.91	1264.67	1203.00	1199.51
1/12-2/12	124	123.67	131.29	143.53	144.20
2/12-3/12	90	78.33	83.03	91.77	92.26
3/12-4/12	77	58.64	61.89	68.88	69.28
4/12-5/12	49	47.46	49.78	55.69	56.03
5/12-6/12	29	40.24	41.87	47.03	47.32
6/12-7/12	72	35.20	36.26	40.87	41.13
7/12-8/12	33	31.53	32.06	36.24	36.48
8/12-9/12	41	28.78	28.79	32.62	32.84
9/12-10/12	35	26.77	26.16	29.72	29.92
10/12-11/12	24	25.46	24.01	27.32	27.52
11/12-1	25	26.00	22.20	25.32	25.50
Total	1802	1802.00	1802.00	1802.00	1802.00
Parameters	$\hat{a}$	0.132	0.142	0.163	0.164
	$\hat{b}$	0.929	-	-	-
K-S test	statistic	0.043	0.038	0.021	0.020
	p-value	0.002	0.010	0.399	0.443

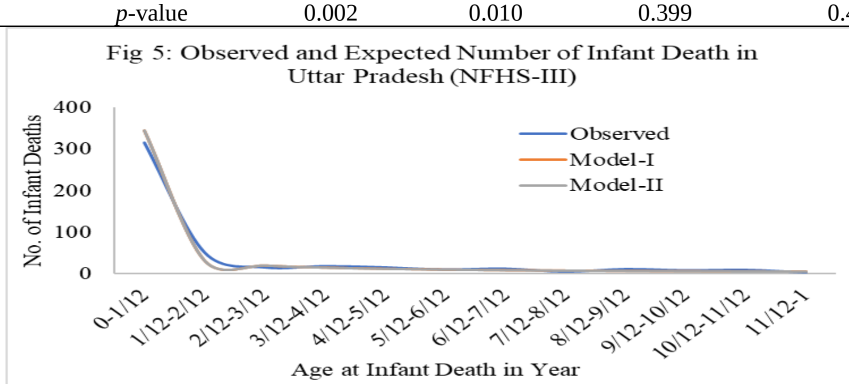


Table 5: Observed and Expected Number of Infant Death in Uttar Pradesh (NFHS-III)

Age at Death (In years)	Observed Number of Infant Death	Expected Number of Infant Death			
		Model-I	Model-II		
		MM estimates	MM estimates	Iteration estimates	ML estimates
0-1/12	315	345.29	344.89	315.00	315.27
1/12-2/12	51	31.29	31.55	37.62	37.56
2/12-3/12	15	19.62	19.77	24.05	24.02
3/12-4/12	18	14.56	14.66	18.05	18.02
4/12-5/12	15	11.67	11.74	14.60	14.57
5/12-6/12	10	9.80	9.84	12.33	12.30
6/12-7/12	12	8.48	8.50	10.71	10.69
7/12-8/12	5	7.49	7.50	9.50	9.48
8/12-9/12	11	6.73	6.72	8.55	8.53
9/12-10/12	8	6.13	6.10	7.79	7.77
10/12-11/12	9	5.65	5.58	7.16	7.15
11/12-1	3	5.29	5.16	6.64	6.62
Total	472	472.00	472.00	472.00	472.00
Parameters	$\hat{a}$	0.125	0.126	0.163	0.162
	$\hat{b}$	0.989	-	-	-
K-S test	statistic	0.064	0.063	0.028	0.028
	p-value	0.040	0.043	0.837	0.851

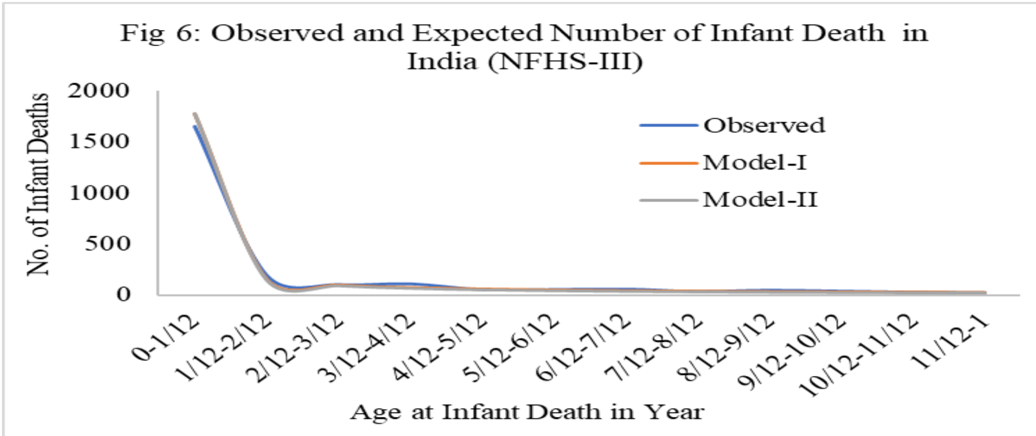


Table 6: Observed and Expected Number of Infant Death in India (NFHS-III)

Age at Death (In years)	Observed Number of Infant Death	Expected Number of Infant Death			
		Model-I	Model-II		
		MM estimates	MM estimates	Iteration estimates	ML estimates
0-1/12	1654	1777.95	1783.32	1654.00	1660.58
1/12-2/12	193	155.25	152.90	179.92	178.60
2/12-3/12	102	96.88	95.45	114.19	113.25
3/12-4/12	110	71.55	70.56	85.31	84.56
4/12-5/12	53	57.12	56.42	68.73	68.10
5/12-6/12	54	47.72	47.22	57.88	57.33
6/12-7/12	58	41.07	40.73	50.18	49.69
7/12-8/12	36	36.08	35.89	44.41	43.96
8/12-9/12	47	32.19	32.13	39.91	39.50
9/12-10/12	38	29.02	29.12	36.30	35.92
10/12-11/12	28	26.36	26.66	33.33	32.98
11/12-1	22	23.80	24.60	30.84	30.52
Total	2395	2395.00	2395.00	2395.00	2395.00
Parameters	$\hat{a}$	0.121	0.119	0.149	0.147
	$\hat{b}$	1.016	-	-	-
K-S test	statistic	0.052	0.054	0.011	0.009
	p-value	0.000	0.000	0.946	0.987

Table 7 and 8, shows the pattern of infant mortality in Uttar Pradesh and India for NFHS-IV (2015-16) respectively. From table 7, we observed that the estimated frequencies obtained from Model-II through iterative method and ML estimation are closer to observed frequencies and ML estimates are under estimated whereas the estimates of Model-I are slightly less to estimates of Model-II obtained from method of moments but both are over estimates. The value of statistics for Model-I is greater than the Model-II for all the estimation procedures is used here, which indicates that the Model-II is gives better fit than the Model-I and p-value also provide the same. And from table 8, we see that the estimates of Model-I and Model-II (MM estimates) are over estimated than the observed one and it is found that the estimates of Model-I is slightly lower than the Model-II (MM estimates). Estimates of Model-I through iterative method and ML estimation are closer to observed frequencies and ML estimation gives the underestimates than the observed one. Here, we observed that the value of statistics for Model-I is lower than that of the Model-II for method of moment estimation procedure and greater than for the iterative method and ML estimation procedures, which conclude that the Model-II provide the closer fit to the observed one for iterative and ML estimation procedures from Model-I and the same result is also observed from the *p*-value.

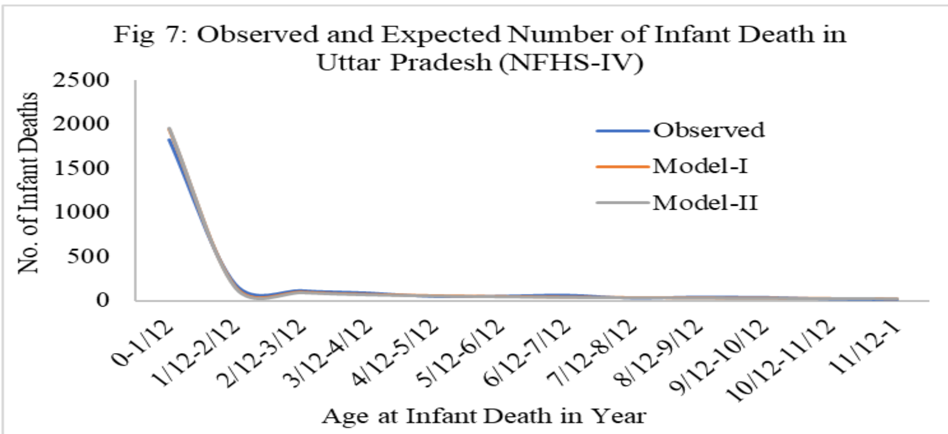


Table 7: Observed and Expected Number of Infant Death Uttar Pradesh (NFHS-IV)

Age at Death (In years)	Observed Number of Infant Death	Expected Number of Infant Death			
		Model-I	Model-II		
		MM estimates	MM estimates	Iteration estimates	ML estimates
0-1/12	1820	1942.65	1956.45	1820.00	1809.65
1/12-2/12	187	154.19	146.61	176.24	178.40
2/12-3/12	116	95.24	90.81	110.90	112.40
3/12-4/12	90	69.73	66.80	82.40	83.58
4/12-5/12	50	55.20	53.21	66.13	67.12
5/12-6/12	53	45.71	44.40	55.51	56.37
6/12-7/12	65	38.96	38.21	48.00	48.76
7/12-8/12	33	33.86	33.60	42.38	43.07
8/12-9/12	44	29.81	30.02	38.02	38.64
9/12-10/12	40	26.43	27.17	34.52	35.09
10/12-11/12	22	23.40	24.84	31.65	32.18
11/12-1	15	19.82	22.89	29.24	29.74
Total	2535	2535.00	2535.00	2535.00	2535.00
Parameters	$\hat{a}$	0.111	0.104	0.133	0.136
	$\hat{b}$	1.063	-	-	-
K-S test	statistic	0.048	0.054	0.009	0.011
	p-value	0.000	0.000	0.977	0.894

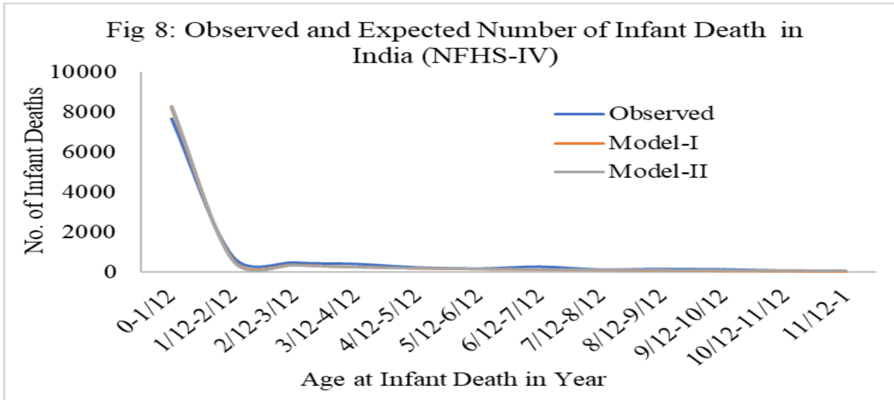


Fig 8: Observed and Expected Number of Infant Death in India (NFHS-IV)

Table 8: Observed and Expected Number of Infant Death in India (NFHS-IV)

Age at Death (In years)	Observed Number of Infant Death	Expected Number of Infant Death			
		Model-I	Model-II		
		MM estimates	MM estimates	Iteration estimates	ML estimates
0-1/12	7679	8204.17	8312.49	7679.00	7633.02
1/12-2/12	771	645.71	583.45	722.72	732.40
2/12-3/12	479	396.00	360.10	453.90	460.58
3/12-4/12	411	287.65	264.30	336.84	342.09
4/12-5/12	240	225.73	210.19	270.07	274.45
5/12-6/12	178	185.07	175.18	226.55	230.33
6/12-7/12	275	155.91	150.57	195.77	199.12
7/12-8/12	133	133.61	132.28	172.78	175.79
8/12-9/12	161	115.56	118.12	154.91	157.66
9/12-10/12	140	100.05	106.83	140.59	143.13
10/12-11/12	75	85.41	97.59	128.84	131.20
11/12-1	59	66.13	89.89	119.02	121.23
Total	10601	10601.00	10601.00	10601.00	10601.00
Parameters	$\hat{a}$	0.111	0.098	0.130	0.132
	$\hat{b}$	1.133	-	-	-
K-S test	statistic	0.050	0.060	0.014	0.016
	p-value	0.000	0.000	0.032	0.007

Table 9, represents the estimates of parameters of Model-I and Model-II, for Uttar Pradesh and India according all round of National Family Health Survey (NFHS). From the estimates of parameters of Model-I, we are unable to draw any meaningful conclusion. Whereas estimates of parameter of Model-II gives the idea about the proportion of neo-natal deaths (death of infants within one month of their birth).

Table 9: Summary of the Estimates of the Parameters of the Models

NFHS Survey		Proportion of neonatal deaths	Model-I		Model-II
			$\hat{a}$	$\hat{b}$	$\hat{a}$
India	**I**	63.8	0.145	0.890	0.181
	II	66.8	0.132	0.929	0.142
	III	69.1	0.121	1.016	0.119
	IV	72.4	0.111	1.133	0.098
Uttar Pradesh	**I**	60.9	0.162	0.730	0.208
	II	63.2	0.152	0.859	0.177
	III	66.7	0.125	0.989	0.126
	IV	71.8	0.111	1.063	0.104

Conclusion

From the above result, we found that the proposed Model-II is simple and may be considered as a competitive model of Model-I (Beta-First Kind distribution) to describe the pattern of infant mortality. The main advantage of proposed Model-II is that it is single parameter model and mathematically easy to get the required statistical constants and it also provide the proportion of neo-natal deaths through the value of its parameter, whereas the Model-I has two parameters and doesn't provide any meaningful information through the value of parameters. We observed that the value of parameter of Model-II is decreases as the proportion of neo-natal deaths increases i.e., parameter of Model-II is inversely proportional to neonatal deaths. We also observed that as the value of parameter is increasing the proposed Model-II having a flat tail and for very low value of parameter, the model is being an early failure model and maximum probability lies in the first cell which indicates the neonatal mortality. As far as concern the infant mortality the maximum number of deaths lies in the first month of their life. Thus, this distribution is an appropriate choice for modeling the infant mortality. The model proposed in the present study to graduate infant deaths for its efficiency in exploring the age pattern of infant deaths with its cumulative probabilities. It can be said that model is quite satisfactory in every situation that mentioned. Under-reporting and mis-reporting of infant deaths in both India and Uttar Pradesh have been observed. There may be several reasons for this high under-reporting in data. The reasons may be due to the lower status of socio-economic conditions, lack of knowledge and particular way of functioning of government machinery responsible for survey or data collection regarding infant deaths, cultural factors affecting the reporting of infant death and the response bias for such sensitive data sets. Improved training and data collection facilities may be provided to the functionaries involved in the collection of vital statistics.

References

Bain, L. J., & Engelhardt, M. (1991). Reliability test plans for one-shot devices based on repeated samples. Journal of quality technology, 23(4), 304-311.

Bhuyan, K. C. & Deogratìes, R. (1999). On the probability model of child mortality pattern in North Eastern Libya. The Turkish Journal of Population Studies, 21: 33.37.

Casella, G., & Berger, R. L. (2002). Statistical inference second edition. Duxbur y advanced series.

Chauhan, R. K., 1997, Graduation of infant death by age. Demography India, 26: 261-274.

Chen, Z. (1997). Parameter estimation of the Gompertz population. Biometrical journal, 39(1), 117-124.

Cheo, M. K. (1981). Fitting of Age pattern of Infant and Child Mortality with the Weibull Survival Function. Asianand Pacific Census Forum, 7: 10.13.

Gompertz, B. (1825). XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c. Philosophical transactions of the Royal Society of London, (115), 513-583.

Hartman, M. (1982). Infant and Child Mortality. Dept. of Statistics, University of Lund. Sweden.

Heligman, L., & Pollard, J. H. (1980). The age pattern of mortality. Journal of the Institute of Actuaries, 107(1), 49-80.

Kabir, M.,Amin, R. and Choudhury, J. (1993). Determinants of Infant and Child Mortality in Bangladesh. Journal of Statistical studies, 13: 54-64.

Keyfitz, N. (1997). Introduction to the Mathematics of the Population with Revisions. Addition-Wesley Publishing Company, Reading, Msssachusetts.

Krishnan, P. and Jin, Y. (1993). A Statistical Model for Infant Mortality. Paper presented at the IUSSP Meetings, Montreal, Canada, August 25-September 1, 1993.

Krishnamoorthy, S. (1982, February). On graduating infant and childhood mortality with mathematical curves. In Asian and Pacific census forum (Vol. 8, No. 3, pp. 14-15).

Krishnamoorty, S., & Mathew, A. C. (1994). Estimating the person years lived in the infant and childhood ages. Janasamkhya, 12(1-2), 151-157.

Krishnamoorthy, S., &Rajna, P. N. (1999). Graduating Childhood Deaths and Survivorship Function with the Model by Mukherjee and Islam. Demography India, 28, 13-22.

Kolmogorov A.N. (1933). Sulla determinazioneempirica di unalegge di distribuzione. Giornale dell' IstitutoItalianodegliAttuari 4: 83–91.

Mathew, A. C., & Krishnamoorthy, S. (1995). Refined Estimates of the Person Years at Infant and Childhood Ages of Model Life Tables. Contributions to Applied Mathematical Statistics, 1, 67-76.

Mitra, S., & Denny, C. (1994). On the Application of a Model of Mortality. Canadian Studies in Population [ARCHIVES], 117-132.

Mukherjee, S. P. and Islam, A. (1983). A Finite Range Distribution of Failure Time. Naval Research Logistic Quarterly, 30: 487.491.

Perks, W. (1932). On some experiments in the graduation of mortality statistics. Journal of the Institute of Actuaries, 63(1), 12-57.

Renshaw, A. E., 1991, Actuarial graduation practice and generalized linear and nonlinear models. J. Inst. Act. 118: 295-312.

Singh, Brijesh P., Neha S, Tapan KR, Gunjan S (2012). On the pattern of child loss in Madhya Pradesh. Mortality, Health and Development. Editors Alok Ranjan Chaurasia, MLC Foundation; Shyam Institute, Bhopal.

Singh, Brijesh P., Dixit, S., & Roy, T. K. (2015). Stochastic Analysis of Infant Deaths by Age and Estimation of Parameters. Journal of Institute of Science and Technology, 20(1), 1-5.

Singh, Brijesh P., Das, U. D., Singh, S., & Singh, G. (2021). Development of a discrete probability distribution and its application to the pattern of child deaths. Journal of Scientific Research, 65(5), 186-190.

Singh, K. K., Singh, Brijesh P., & Singh, N. (2011). A probability model for number of child death for fixed parity. Demography India, 40(2), 55-68.

Smirnov NV (1933). Estimate of deviation between empirical distribution functions in two independent samples. Bulletin Moscow University 2: 3–16.

Theile, P.N. (1972). On a mathematical formula to express the rate of mortality throughout the whole life. Journal of Institute of Actuaries, Vol. 16: 313-329.