İstatistik Araştırma Dergisi
Journal of Statistical Research
ISSN: 2791-7614

(2023), 13 (1)
Başvuru/Received: 12/11/2022
Kabul/Accepted: 04/07/2023

https://journal.tuik.gov.tr/

ARAŞTIRMA MAKALESİ RESEARCH ARTICLE

A One-Dimensional Biased Probability Model Based on Himanshu Distribution for Vital Events Related to Migration and Mortality Data

Abhishek AGARWAL
D.D.U Gorakhpur University, INDIA
abhishek.agarwal014@gmail.com
Orcid No: 0000-0002-2425-7471

Himanshu PANDEY
D.D.U Gorakhpur University, INDIA
himanshu_pandey62@yahoo.com
Orcid No: 0000-0002-3839-6560

Abstract

In this paper a one dimensional biased probability model based on Himanshu distribution has been obtained by size biasing the Himanshu distribution introduced by Agarwal and Pandey (2022). Its raw moments and central moments has been obtained. Hence expression of coefficient of variation, Index of dispersion, Skewness, Kurtosis have also been given. The parameter involved in the proposed Model has been obtained by the estimation techniques. The suitability of the one dimensional biased probability model tested through the real data sets related to human migration and mortality patterns of different regions.

Keywords: One Dimensional Biasing, Himanshu Distribution, Moments, Estimation of Parameter, Goodness of Fit

Corresponding Author / Sorumlu Yazar: 1-Abhishek AGARWAL, D.D.U Gorakhpur University, Department of Mathematics and Statistics

2-Himanshu PANDEY, D.D.U Gorakhpur University, Department of Mathematics and Statistics

Citation / Atıf: AGARWAL A., PANDEY H. (2023). A One-Dimensional Biased Probability Model Based on Himanshu Distribution for Vital Events Related to Migration and Mortality Data. İstatistik Araştırma Dergisi, 13 (1), 48-60.

Göç ve Mortalite Verilerine İlişkin Yaşamsal Olaylar için Himanshu Dağılımı Temelli Tek Boyutlu Önyargılı Olasılık Modeli

Özet

Bu çalışmada, Agarwal ve Pandey (2022) tarafından tanıtılan Himanshu dağılımına dayalı tek boyutlu önyargılı bir olasılık modeli elde edilmiştir. Ham ve merkezi momentleri çıkarılmıştır. Değişim katsayısı, dağılım indeksi, çarpıklık, basıklık değerleri de verilmiştir. Önerilen modelde yer alan parametre, uygun tahmin yöntemleri ile elde edilmiştir. Tek boyutlu yanlı olasılık modelinin uygunluğu, farklı bölgelerdeki göç ve ölüm örüntülerine ilişkin gerçek veri setleri üzerinden test edilmiştir.

Anahtar kelimeler: Uyum iyiliği, Tek boyutlu yanlılık, Himanshu Dağılımı, Momentler, Parametre.

1. Introduction

Probability model based on distributions play an important role in the various fields of Social Sciences, Medical Sciences, Environmental Sciences etc. Recently Pandey and Jai kishun (2009,2010), Dubey and Pandey (2022) has suggested probability model for vital events.

Himanshu distribution has probability mass function (PMF) as below. It is used by Agarwal and Pandey (2022).

$$P(X=x)=p^n(1-p^n)^x \qquad ; \quad \begin{array}{l} x=0,1,2,\dots \\ 0<p<1 \\ n\in I^+ \end{array} \tag{1.1}$$

for modelling count data from different field of real life problems.

$$\text{with Mean} = \frac{1-p^n}{p^n} \quad \text{and Variance} = \frac{1-p^n}{p^{2n}}$$

One dimensional biased distribution is a special type of weighted distribution. They can be seen while observation from a sample are selected with probabilities proportional to any measures of unit size. It was first given by Fisher (1934). One dimensional biased observation occurs in many research areas related to real life problems includes Environmental Science, Medical Science, Population Science, Psychology, Ecology etc. Van Deusen (1986) has been discussed the application of one dimensional biased distribution theory in fitting distributions of diameter at breast height (DBH) data arising from horizontal point sampling. Lappi and Bailey (1987) have applied one dimensional biased distribution to analyze horizontal point sampling diameter increment data.

Patil and Rao (1977, 1978) has studied detailed statistical application of one dimensional biased distribution to analysis of data relating to human population and ecology.

One dimensional biasing is defined in the following way.

Let a random variable X has original probability distribution(PD) $P_o(x;\theta)$; $\begin{array}{l} x=0,1,2,\dots \\ \theta>0 \end{array}$

Suppose the sample units are weighted or selected with probability proportional to some measure. x^{α}. Here α is a positive integer. Then the corresponding one dimensional biased PD of order α can be defined by its PMF.

$$P_1(x;\theta) = \frac{x^{\alpha}.P_0(x;\theta)}{\mu'_{\alpha}} \tag{1.2}$$

$$\text{where } \mu'_{\alpha} = E(X^{\alpha}) = \sum_{x=0}^{\infty} x^{\alpha}\, P_0(x\,;\theta)$$

If $\alpha = 1$, then the distribution is known as one dimensional biased and is applicable for one dimensional biased sampling in sampling theory.

The PMF of the one dimensional biased Himanshu distribution with parameter p can thus be obtained as;

$$P_2(x;\theta) = \frac{x.P_0(x;\theta)}{\mu'_1} = x.p^{2n}(1-p^n)^{\,x-1};\quad \begin{array}{l} x = 1,2,3,\dots \\ 0 < p < 1 \\ n \in I^+ \end{array} \tag{1.3}$$

$$\text{where } \mu'_1 = \frac{1-p^n}{p^n} \text{ is the Mean of Himanshu distribution with p.m.f (1.1)}$$

Mortality is the population's other important occurrence. In the actual world, mortality has emerged as the key concern for hospitals and the insurance sector. Living things have the ability and may even be required to die at some point. Many demographers and social scientists are able to quantify the event mortality, and for the smooth study of mortality pattern they adopted the form of modeling in a probabilistic environment. This is due to the increased interest in recent decades in understanding the mortality pattern and risk. The many models have been suggested in this manner by Pandey and Shukla (2014), Pandey etal. (2015), Agarwal and Pandey (2022) etc.

Migration is the third event of the population change, the other two being Mortality and Fertility. The nature of migration is very complex as a factor affecting population size different from that of mortality and fertility. The Migration affect the Social, Cultural, Economical, Political Characteristic of the society. Any region's population distribution and the expansion of its work force are significantly influenced by migration. Studying the migratory pattern and its effects on society at the micro level may roughly be done using the probability model provided by Aryal (2003, 2011), Singh etal.(2016), Dubey and Pandey (2021), Agarwal and Pandey (2022) etc.

2. Proposed Model

The PMF of the proposed model can be formed by using (1.1) and (1.2) in the following way.

$$P(X=x) = x\,p^{2n}(1-p^n)^{\,x-1}\quad;\qquad \begin{array}{l} x = 1,2,3,\dots \\ 0 < p < 1 \\ n \in I^+ \end{array} \tag{2.1}$$

The Moment Generating Function of (2.1) is given as

$$M_X(t) = \sum_{x=1}^{\infty} e^{tx}\, x p^{2n}(1-p^n)^{x-1}$$

$$M_X(t) = \frac{p^{2n}e^t}{\{1-e^t(1-p^n)\}^2} \tag{2.2}$$

Then the first four moments (about origin) are as follows.

$$\text{Mean} = \mu_1' = \frac{2}{p^n} - 1 \qquad \mu_2' = \frac{p^{2n} - 6p^n + 6}{p^{2n}}$$

$$\mu_3' = \frac{-(p^{3n} - 14p^{2n} + 36p^n - 24)}{p^{3n}}$$

$$\text{and} \quad \mu_4' = \frac{p^{4n} - 30p^{3n} + 150p^{2n} - 240p^n + 120}{p^{4n}} \tag{2.3}$$

and central moments as,

$$\left.\begin{aligned} \mu_2 = 2\left(\frac{1-p^n}{p^{2n}}\right), \mu_3 = \frac{2(1-p^n)(2-p^n)}{p^{3n}} \\ \text{and} \quad \mu_4 = \frac{2(1-p^n)(p^{2n} - 12p^n + 12)}{p^{4n}} \end{aligned}\right\} \tag{2.4}$$

$$\left.\begin{aligned} \gamma_1 = \frac{2-p^n}{\sqrt{2(1-p^n)}} \\ \text{and } \gamma_2 = \frac{p^{2n} - 6p^n + 6}{2(1-p^n)} \end{aligned}\right\} \tag{2.5}$$

$$\left.\begin{aligned} \text{Index of dispersion} = \frac{2}{p^n}\left(\frac{1-p^n}{2-p^n}\right) \\ \text{and C.V} = \frac{\sqrt{2(1-p^n)}}{2-p^n} \end{aligned}\right\} \tag{2.6}$$

3. Parameter Estimation

The parameter p of the Model (2.1) can be estimated using Maximum likelihood estimation in the following way.

$$L = \prod_{i=1}^{k} f(x_i\,; p)$$

$$L = (p^{2n})^k \prod_{i=1}^{k} x_i\, (1-p^n)^{\sum_{i=1}^{k} x_i - k}$$

Taking log and upon differentiating w.r.t *p* and equating to zero we get

$$\hat{p} = \left(\frac{2}{\bar{x}+1}\right)^{\frac{1}{n}}$$

Again the parameter *p* estimated by method of moments after using (2.3) we get

$$\bar{x} = \mu_1'$$
$$\Rightarrow \hat{p} = \left(\frac{2}{\bar{x}+1}\right)^{\frac{1}{n}}$$

4. Application

The proposed probability model at *n=2* is fitted using some real data sets collected from different sample surveys entitled "Demographic survey of Chandauli district (Rural Area- 2001-2002)"-(2015); Rural development and population growth survey 1978-PRC, BHU" (2015); and "the survey collected by the researcher in Varanasi district (2018) for the single adult male migrant of 15 years and above.

Mortality does not depend only on the biological and epidemiological factors. It also depends on some socioeconomic and cultural factors. Prevailing health conditions, medical facilities, environmental conditions also worth mentioning. In developing besides under developed countries the mortality among infants and children is much higher than youngsters. These reasons the high infant mortality has thrown a serious challenge to the doctors and medical personnel. So, it can be seen as one of the sensitive position of existing medical and health facilities in the population. Therefore, we have studied the infant mortality pattern using proposed model. In this respect the real data set of Sri Lanka taken from the survey by Meegama (1980) and the real data set of India is taken from Lal (1955).

Table 1: Observed and Expected number of households(NoH) with at least one male migrant according to the number of male migrants aged 15 years and above (2001 survey)

No of migrants	Observed no. of households	Exp. no. of households
1	97	91.60
2	35	44.56
3	19	16.25
4	6 } 9	7.58
5	3	
Total	**160**	**160**
Mean=1.643 Variance=0.8494 $\hat{p}$ =0.8698	$\chi^2 = 3.08$ (after pooling) p-value=0.2143 $\chi^2_{(2)} = 5.99$ at 5% level of significance	

Table 2: Same data with Table 1 according to the 1978 survey

Number of migrants	Observed no. of households	Expected no. of households
1	375	367.2
2	143	154.9
3	49	49.0
4	17	13.7
5	6	5.2
Total	**590**	**590**
Mean=1.535 Variance=0.6780 $\hat{p}$ =0.8882	$\chi^2 = 1.96$ (after pooling) p-value=0.5807 $\chi^2_{(3)} = 7.815$ at 5% level of significance	

Table 3: Same data with Table 1 according to 1978 survey in three types of households.

	Type of households					
	Semi Urban		Remote		Growth Centre	
	Observed	Expected	Observed	Expected	Observed	Expected
1	95	86.60	176	169.66	154	146.02
2	19	31.25	59	66.74	47	59.31
3	10 } 15		18	19.69	18	18.06
4	2 } 15	11.15	6 } 10		9 } 11	
5	3 } 15		4 } 10	6.90	2 } 11	6.60
Total	**129**	**129**	**263**	**263**	**230**	**230**
Mean=	1.44		1.49		1.51	
Variance=	0.5367		0.6098		0.6397	
χ^2(after pooling) $=$	6.92		2.65		5.91	
d.f=	1		2		2	
p-value=	0.0085		0.2658		0.0520	

Table 4: Observed and Expected NoHs having adult Male Migrants aged 15 years and above.

Number of migrants	Observed no. of households	Expected no. of households
1	97	93.00
2	42	47.19
3	16	17.95
4	7 } 12	8.86
5^+	5 } 12	
Total	**167**	**167**
Mean=1.68 Variance=0.9111 $\hat{p}$ =0.8638	$\chi^2 = 2.07$ (after pooling) p-value=0.3552 $\chi^2_{(2)} = 5.99$ at 5% level of significance	

Table 5: Observed NoHs having adult Male Migrants aged 15 and above in North Eastern Bihar.

Number of migrants	Observe NoHs	Expected NoHs
1	95	88.10
2	41	48.98
3	15	20.42
4	12	7.57
5	6	3.93
Total	**169**	**169**
Mean=1.77 Variance=1.066 $\hat{p}$ =0.8497	$\chi^2 = 6.96$ (after pooling) p-value=0.0731 $\chi^2_{(3)} = 7.815$ at 5% level of significance	

Table 6: The number of mothers(NoM) of the Rural Area having at least one live birth and one neonatal death(ND).

Number of NDs	Observed no. of mothers	Expected no. of mothers
1	409	402.67
2	88	97.92
3	19	17.86
4	5 } 6	3.55
5	1	
Total	**522**	**522**
Mean=1.27 Variance=0.3152 $\hat{p}$ =0.9372	$\chi^2 = 2.86$ (after pooling) p-value=0.2393 $\chi^2_{(2)} = 5.99$ at 5% level of significance	

Table 7: The NoMs of the Estate Area having at least one live birth and one ND.

Number of NDs	Observed no. of mothers	Expected no. of mothers
1	71	69.27
2	32	32.38
3	7	11.34
4	5 } 8	5.01
5	3	
Total	**118**	**118**
Mean=1.61 Variance=0.7961 $\hat{p}$ =0.8753	$\chi^2 = 3.48$ (after pooling) p-value=0.1755 $\chi^2_{(2)} = 5.99$ at 5% level of significance	

Table 8: The NoMs of the Urban Area with at least two live births by the number of infant and child deaths.

No. of Infant and child deaths	Observed no. of mothers	Expected no. of mothers
1	176	168.01
2	44	57.19
3	16	14.60
4	6 } 8	4.20
5	2 }	
Total	**244**	**244**
Mean=1.41 Variance=0.4943 $\hat{p}$ =0.9109	$\chi^2 = 6.97$ (after pooling) p-value=0.0306 $\chi^2_{(2)} = 9.21$ at 1% level of significance	

Table 9: The NoM of the completed fertility having experienced at least one child death.

No. of child deaths	Observed NoM	Expected NoM
1	89	79.88
2	25	36.85
3	11	12.75
4	6 }	
5	3 } 10	5.52
6	1 }	
Total	**135**	**135**
Mean=1.60 Variance=0.7797 $\hat{p}$ =0.8770	$\chi^2 = 8.72$ (after pooling) p-value=0.0127 $\chi^2_{(2)} = 9.21$ at 1% level of significance	

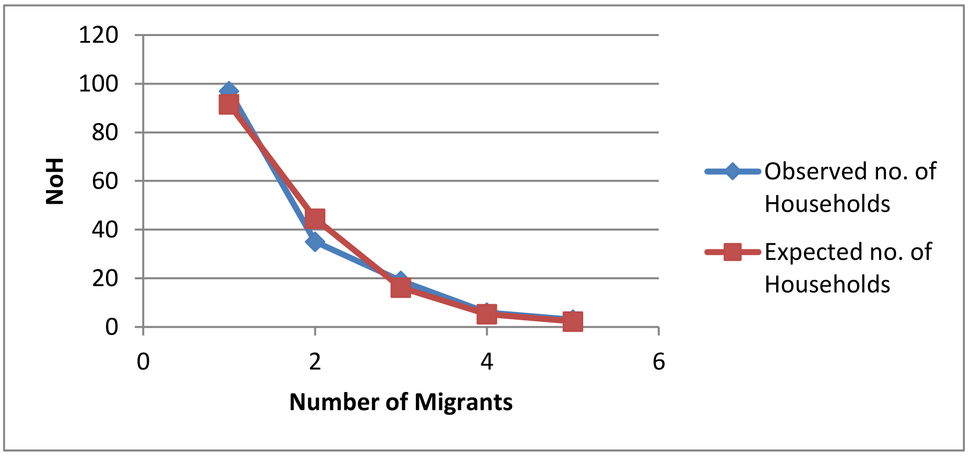

Figure 1: Graphical presentation showing observed and expected NoH aged 15 years and above (survey 2001 data).

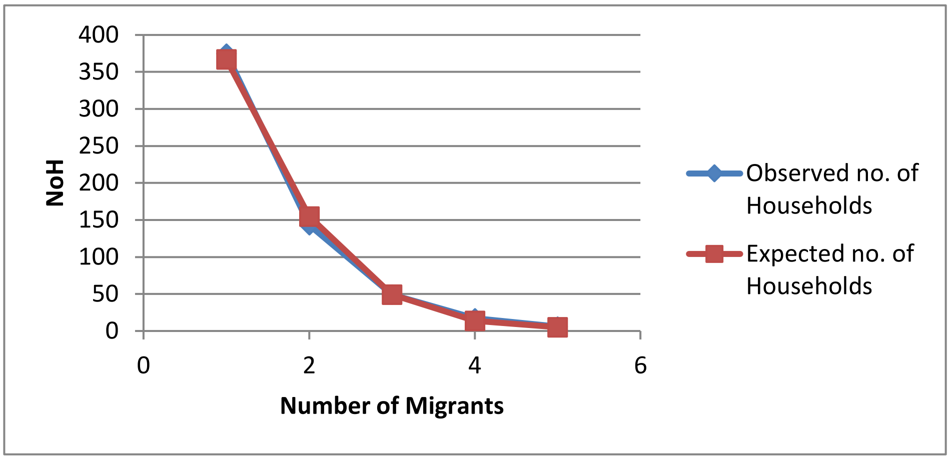

Figure 2: Graphical presentation showing observed and expected NoH with at least one male migrant according to the number of male migrants aged 15 years and above (survey 1978 data).

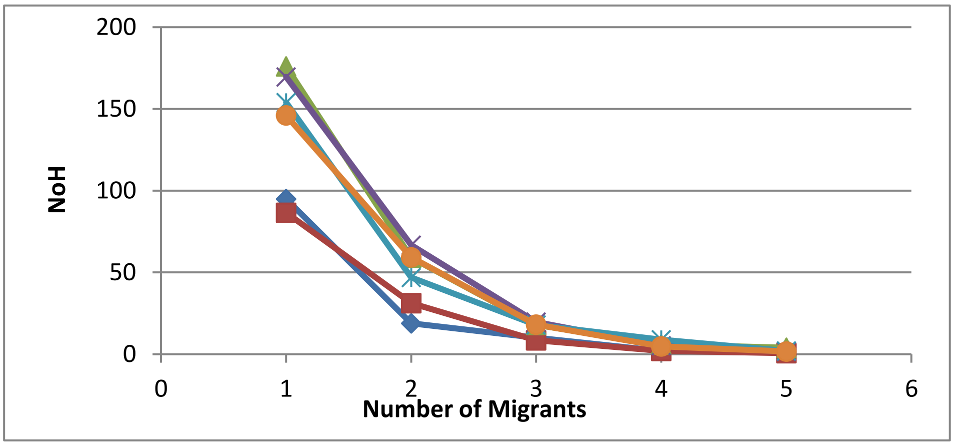

Figure 3: Graphical presentation showing observed and expected NoHs aged 15 years and above (survey 1978 data) in three types of households.

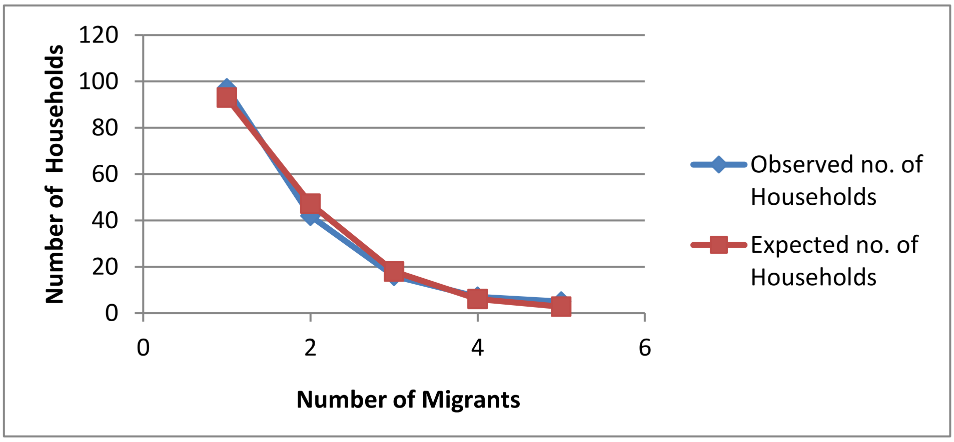

Figure 4: Graphical presentation showing observed and expected number of households having adult Male Migrants aged 15 years and above.

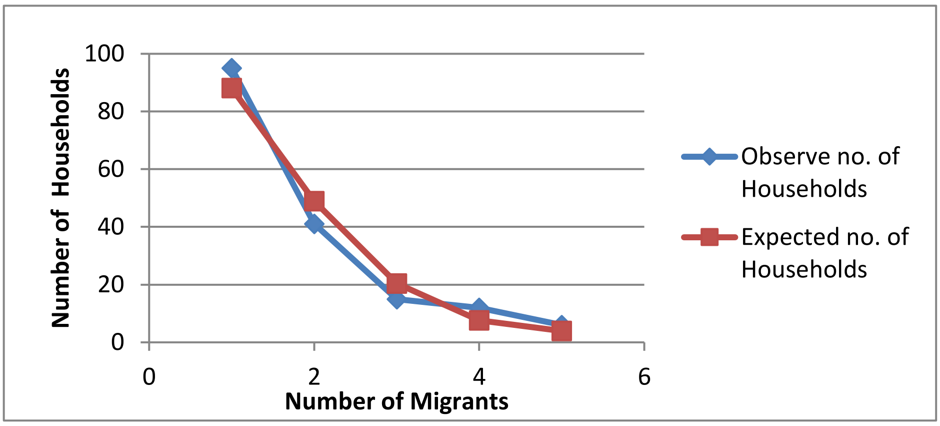

Figure 5: Graphical presentation showing observed and expected NoHs having adult Male Migrants aged 15 and above in North Eastern Bihar.

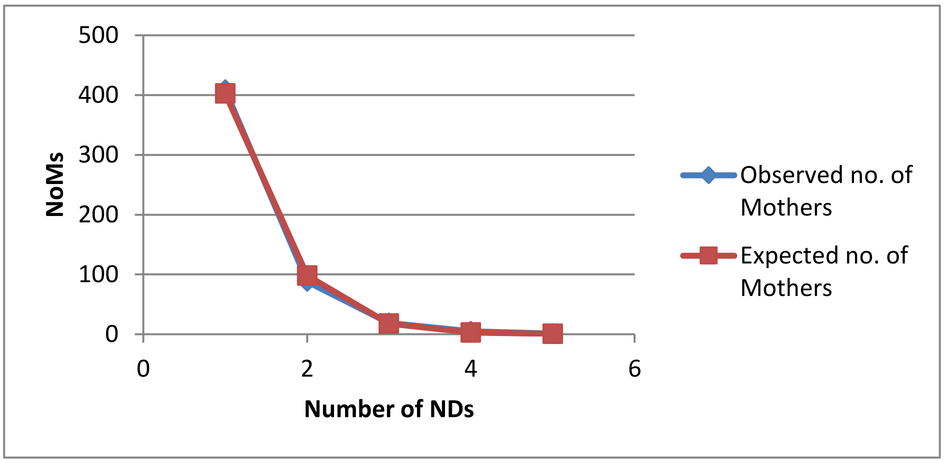

Figure 6: Graphical presentation showing observed and expected no. of mothers of the Rural Area having at least one live birth and one ND.

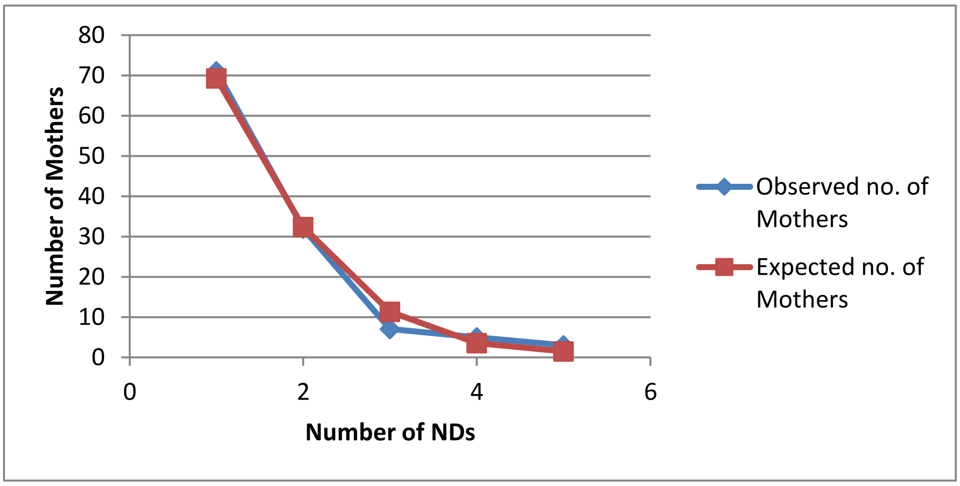

Figure 7: Graphical presentation showing observed and expected no. of mothers of the Estate Area having at least one live birth and one ND.

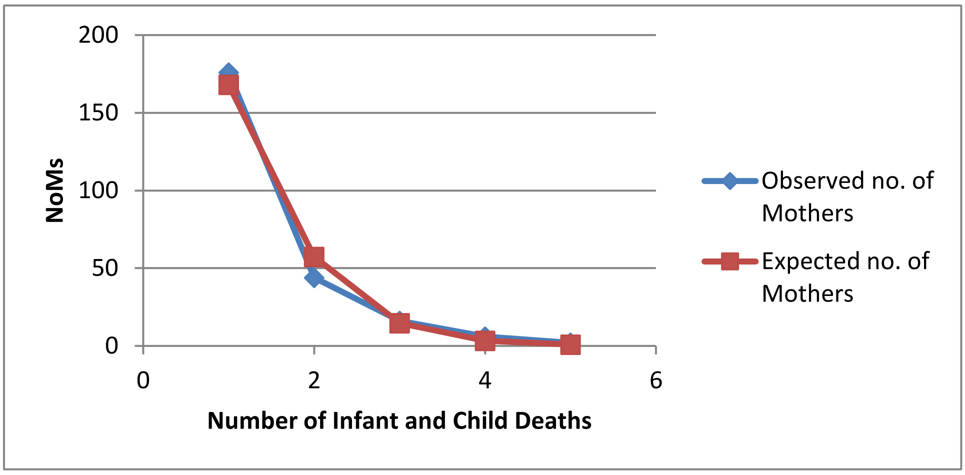

Figure 8: Graphical presentation showing observed and expected no. of mothers of the Urban Area with at least two live births by the number of infant and child deaths.

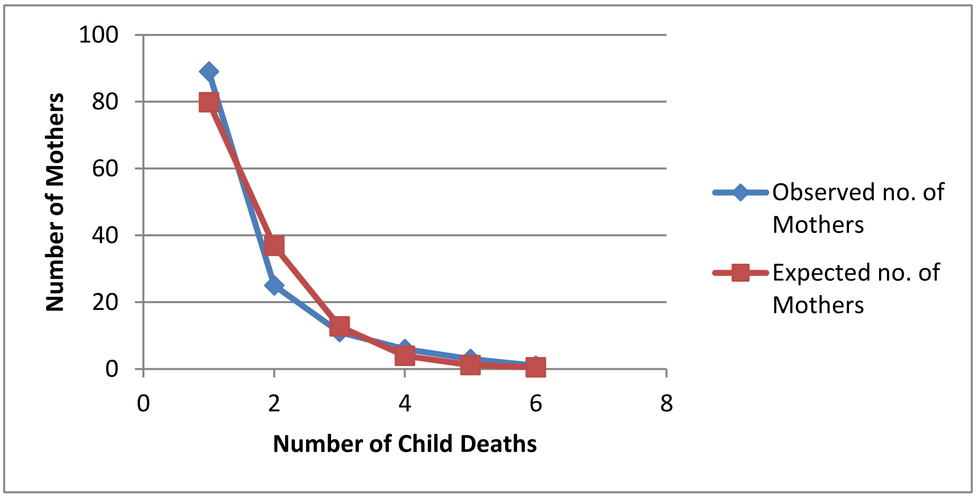

Figure 9: Graphical presentation showing observed and expected no. of mothers of the completed fertility having experienced at least one child death

5. Conclusion

A chi-square goodness of fit test determines if sample data matches a population. A **low value** for chi-square means there is a **high correlation** between your two sets of data. From tables(1 to 9), it clearly indicates calculated χ^2 is less than the critical χ^2 value at 1% and 5% level of significance, Hence we conclude there is no significant difference between the observed and expected value of the given data set. For a Chi-square test, a **p-value** that is less than or equal to your significance level indicates there is sufficient evidence to conclude that the observed distribution is not the same as the expected distribution. According to the value of χ^2 and p-value from (tables 1,2,3,4,5,6,7,8,9) and graphical representation between O_i and E_i, the nature and behavior of proposed one dimensional biased probability model found suitable for the migration pattern as well as infant mortality pattern of different regions. The overall studies shows that the proposed one dimensional biased probability model could also be helpful in policy making, Rural development, Fresh Environment, Medical Facilities for the betterment of the society.

Probability is used in Bayesian analysis for both data and hypotheses. It pertains to a subjective assessment of the veracity of an occurrence. A different approach to traditional statistics is provided by Bayesian statistics. It stands out for its capacity to characterize uncertain values using probability distributions, which leads to elegant solutions to several challenging statistical problems and is extensively useful in the fields of demography, medicine, and insurance. Now that these viewpoints and the work of Rao and Pandey (2020, 2021) have been taken into consideration, it is possible to employ the Bayesian Analysis of the suggested model by figuring out various loss functions.

Acknowledgment

The authors are very thankful to referees for their valuable suggestions.

References

Aryal T.R. (2011). Inflated Geometric distribution to study the distribution of rural out migrants, *Journal of Institute of Engineering*, 8(1&2), 266-268.

Agarwal A., Pandey H., Himanshu distribution and its applications, *Bulletin of Mathematics and Statistics Research*.2022;10(4):24-35.

Agarwal, A., & Pandey, H. (2022). Probability Model Based on Zero Truncation of Himanshu Distribution and Their Applications. *Asian Journal of Probability and Statistics*, 20(4), 57-67.

Agarwal, A., & Pandey, H. (2022). An Inflated Probability Model for the Adult Male Migrants. *International Journal of Statistics and applied Mathematics*, 7(6), 48-52.

Dubey A., Pandey H. (2021). An inflated probability models for Commutation pattern. *Bulletin of Mathematics and Statistics Research*, 8(4),101-108.

Dubey, A., & Pandey, H. (2022). An inflated probability models for adult out migration. *International Journal of Statistics and Applied Mathematics*, 7(5), 80-82.

Fisher, R. A. (1934). The effect of methods of ascertainment upon the estimation of frequencies. *Annals of eugenics*, 6(1), 13-25.

Jai Kishun, Pandey H. (2009). Inflated Probability Model for Risk of vulnerability to HIV/AIDS infection among female Migrants. *Journal of Reliability and Statistical study*, 2(2), 77-84.

Lal DN (1955). Patna in 1955, A Demographic Sample Survey, *Demographic Research Centre*, Department of Statistics, Patna University, India.

Meegama SA (1980). Socio-economic determinants of infant and child mortality in Sri Lanka, An analysis of post war experience, *International Statistical Institute* (World Fertility Survey), Netherland.

Lappi, J. and bailey, R.L. (1987): Estimation of diameter increment function or other tree relations using angle-count samples, *Forest science*, 33, 725 – 739.

Pandey A, Pandey H, Shukla V.K. (2015). An inflated Probability for rural out migration, *Journal of Computer and Mathematical Sciences*, 6(12), 702-711.

Pandey H., Jai kishun, (2010) A probability model for the child mortality in a family, *Revista Colombiana de Estadistica*, 33(1), 1-11.

Patil, G.P. and Rao, C.R. (1977): The Weighted distributions: A survey and their applications. In applications of Statistics, Ed P.R. Krishnaiah, 383 – 405, *North Holland Publications Co*., Amsterdam.

Patil, G.P. and Rao, C.R. (1978): Weighted distributions and size-biased sampling with applications to wild-life populations and human families, *Biometrics,* 34, 179 – 189.

Singh, B. P., & Singh, N. K. (2016). On the distribution of risk of migration and its estimation. *International Journal of Statistical Distributions and Applications*, 2(4), 67-71.

Sahu K.K., Pandey A., Pandey H. (2015). A displaced probability model for child mortality for fixed parity, *International Journal of Science and Research,* 4(12), 1681-1685.

Van Deusen, P.C. (1986): Fitting assumed distributions to horizontal point sample diameters, For. Sci., 32, 146 -148.

Rao A.K, Pandey H. (2020). Parameter estimation of Area biased Ailamujia distribution. *International Journal of Physics and Mathematical Sciences*, 10(1), 21-28.

Rao A.K, Pandey H. (2021). Bayesian estimation for the parameter of Gamma Lomax distribution under different loss function. *International Journal of Scientific Research in Mathematical and Statistical Sciences*, 8(1), 42-51.