INDIAN JOURNAL OF POPULATION AND DEVELOPMENT
Volume 1(1), January 2021: 137-144

Modelling Seasonality of Deaths in Guwahati City, Assam, India

Kishore K. Das
Sahana Bhattacharjee

Abstract

This paper attempts to model the seasonality in the occurrence of deaths from specific causes of death groups circular statistical data analysis tools. Parameter estimation of future observations based on the past and the current observations have been carried out assuming von Mises distribution. It is observed that for some causes of death groups, the hypothesis of seasonality is accepted whereas it is rejected for the other causes of death groups. Based on the estimated parameters of the von Mises distribution, the seasonality in the occurrence of deaths from different causes of death groups can be predicted.

Introduction

The seasonality in the occurrence of deaths has been known to the world, at least since Hippocrates, more than 2000 years ago. By seasonality, we mean the trend pattern that repeats every 12 months. Irrespective of the reasons behind seasonality, predicting seasonality in the occurrence of death plays an important role in planning and programming public health services. If any disease, and death following the disease, occur seasonally, "an environmental factor has to be considered in the etiology of that disease" (Marrero, 1983). There is already ample evidence to suggest that an enormous diversity of causes of death is related to seasonal incidence: cardiovascular diseases (Yen et al, 2000), asthma (Cadet et al, 1994), infectious diseases (Momiyama, 1987), diarrhoea and cholera (Bouma and Pascual, 2001; Villa, 1999), suicide (Hakko, 2000), and congenital malformations (Elwood and Elwood, 1987; Kanai and Nakamura, 1987) to name only a few. Different analytical approaches have been adopted to identify and test the seasonality in the occurrence of diseases and deaths. These include seasonal mortality ratio, concentration or dissimilarity indexes, time series modelling and the so-called X-11 method. Detailed description of these and other methods is given elsewhere (Rau, 2007). In this paper, we follow a modelling approach to analyse the seasonality in the occurrence of deaths due to selected causes of deaths groups. The approach adopted can also be used to predict seasonality in the occurrence of deaths.

The paper is organised as follows. Section 2 describes the model used to capture the seasonality in the occurrence of deaths. Section 3 describes the data used for the analysis including a description about the data collection methodology. Results of the test of seasonality in the occurrence of deaths due to different causes of death groups are presented and discussed in section 4 of the paper. Section 5 presents results of predictive modelling of seasonality in the occurrence of deaths due to selected causes of death groups. The last section of the paper summarises the findings of the analysis.

Methodology

The underlying assumption of the present analysis is that there is substantial degree of seasonality in the occurrence of deaths, but the magnitude of seasonality remains unpredictable. We, therefore, assume that the seasonality in the occurrence of deaths either from all causes of deaths or from specific causes of death groups is a circular random variable or a random variable on a circle and then apply circular statistical analysis tools and techniques to model the seasonality in the occurrence of deaths. The benefit of treating monthly occurrence of deaths as a circular random variable is the continuity of the curve between the months of December and January.

We analyse seasonality in terms of both month wise occurrence of deaths and season wise occurrence of deaths. For the month wise analysis, the angle of the circle at the centre (360°) is divided into 12 parts in accordance with the length of different months. For season wise analysis, the angle at the centre of the circle is divided into 4 quarters according to the length of the quarter. The seasons used in the present analysis are the same as identified by the Regional Meteorological Centre, Guwahati. The Regional Meteorological Centre classifies 12 months of the year into four seasons as follows:

1. Winter seasons - January, February
2. Pre-monsoon season - March, April, May
3. Monsoon season - June, July, August, September
4. Post-monsoon season - October, November, December

The most fundamental question in the circular data analysis is to test whether the data are uniformly distributed around the circle, or whether it is concentrated around at least one preferred direction. There are different tests available for the purpose including Rayleigh uniformity test, V-test, Watson's test, Kuiper's test, and Rao's spacing test (Landler et al, 2018). We have used the Rayleigh uniformity test in the present analysis. If there is no seasonality in monthly or seasonal occurrence of deaths, then the data can be regarded as drawn from a uniform distribution on the circle. We, therefore, frame our null and alternative hypothesis as:

H_0: The occurrence of deaths does not have any seasonality.
H_1: The occurrence of deaths has seasonality.

There are different approaches of finding the predictive densities in the circular data. These include: i) methods based on conditioning through sufficiency; ii) Bayes predictive densities; and iii) maximum or profile likelihood methods (Bjornstad, 1990; Butler, 1986). Different approaches used to generate models for circular data are discussed elsewhere (Mardia and Jupp, 1999). In this paper, we model the future occurrence of deaths by causes of death groups assuming the classical von Misses model (also known as circular normal distribution) for the past and present observations. The probability density function of the classical von Mises distribution is given by

$$f(\theta, \mu_0, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu_0)}, \qquad 0 < \theta,\ \mu_0 \le 2\pi,\ \kappa > 0 \qquad (1)$$

Then, under the sufficiency approach, the predictive density of θ_{n+1} is givenby

$$g(\theta_{n+1} | \theta_1, \theta_2, \cdots, \theta_n) \propto \frac{1}{\psi_{n+1}(r_{n+1})}, \qquad (2)$$

where$\psi_r = \int_0^\infty J_0(rt) J_0^n(t) dt,\ 0 \le r < n,$ and $J_v(z)$ is the Bessel function of v^{th}order. The predictive density $g(.)$ Is proportional to the von Mises distribution $f(\theta_{n+1}; \bar{\theta}_n, 2\bar{R})$ for large n. (Rao and Sengupta, 2001, pp. 207). When κ=0 in equation (1), the von Mises distribution reduces to the circular uniform distribution. Mathematically, this result can be easily proved (Rao and Sengupta, 2001). Logically also, this is valid because, in the uniform distribution, the observations do not have a preferred direction which means that all directions are equally preferable. This, in turn, means that the concentration about the mean direction is not present in the data, i.e., it is 0. Accordingly, for large n, the distribution will tend to von Mises distribution with parameters $(\hat{\theta}_n, \hat{\kappa})$, where $\hat{\theta}_n$is based on the past n observations and $\hat{\kappa} = \frac{2r_n}{(n+1)}$ is the approximate maximum likelihood estimator for smaller values of κ (Rao and Sengupta, 2001). The value of mean direction μ indicates the direction towards which most of the observations are concentrated, on average. In the present case, the value of μ indicates the corresponding month or season in which most of the observations in the dataset are concentrated, on average. The value of estimated κ gives an idea about the spread or dispersion in the observations. The lower the value, lower is the variance in the data set or higher is the concentration and vice-versa.

Data

Data for the present study have been taken from the research project “Statistical Modelling of Circular Data: An Application to Health Science” which was funded by the University Grants Commission, India (Das, 2015). Under this project, 1371 deaths reported in Guwahati city during different months of the year 2013 and 2014 were covered. The sampling frame of all birth and death registration offices under Guwahati Municipal Corporation, Guwahati, Assam was first prepared and then, taking resort to simple random sampling scheme, three birth and death registration offices, viz. MMCH

PHC, Panbazar, Office of the Joint Director of Health, Uzan Bazar and Baripara PHC, Pandu were selected. Data have been collected for the years 2013 and 2014. These deaths were classified into the following causes of death groups

1. Respiratory diseases
2. Gastro-intestinal diseases
3. Diseases of the urinary tract
4. Cardio-vascular diseases
5. Neurological disorders
6. Accidents and injuries
7. Endocrinal diseases
8. Virological diseases
9. Parasitic diseases
10. All other diseases

Month wise occurrence of deaths from the 10 causes of deaths groups in the year 2013 and 2014 is given in table 1 along with the coefficient of variation (CV). These summary measures suggest that seasonality in the occurrence of deaths is the highest in January but the lowest in December. For all causes of death combined, the average monthly occurrence of deaths ranges from 85 in the month of January to 29 in the month of December while the overall coefficient of variation is 0.304.

Table 1: Reported occurrences of deaths by causes of death groups in Guwahati city, 2013 and 2014

Month	All	Reported deaths by causes of death									
		1	2	3	4	5	6	7	8	9	10
Jan	170	6	11	50	40	17	1	3	9	2	31
Feb	145	9	11	41	27	15	3	3	7	1	28
Mar	146	6	6	42	27	16	2	3	11	2	31
Apr	110	6	9	36	23	12	2	3	3	2	14
May	144	10	13	42	27	9	0	6	11	1	25
Jun	126	4	13	38	21	9	2	1	8	1	29
Jul	129	6	6	33	33	11	0	4	2	2	32
Aug	104	2	17	23	28	5	1	4	6	0	18
Sep	74	3	9	21	21	5	0	1	5	0	9
Oct	77	6	4	22	20	5	3	3	3	0	11
Nov	88	2	5	21	29	5	2	2	5	0	17
Dec	58	3	2	18	20	1	1	0	2	1	10
CV	0.304	0.488	0.496	0.334	0.226	0.559	0.765	0.583	0.541	0.852	0.422

Source: Authors
Remarks: CV is the coefficient of variation

Test of Seasonality in the Occurrence of deaths

The test statistic of the Rayleigh Uniformity test and the decision regarding rejection or acceptance of the hypothesis of seasonality in the occurrence of deaths due to the

10 causes of death groups covered in the present analysis is given in table 2. The null hypothesis was accepted in case of 7 of the 10 causes of deaths groups but in three causes of death groups, the null hypothesis was rejected. The table suggests that there is seasonality in the occurrence of deaths related to urinary tract infections, neurological disorders, and other causes of death but not in case deaths from respiratory disease, gastro-intestinal diseases, cardio-vascular diseases, accidents and injuries, endocrine diseases, viral infections, and parasitic diseases. Table 2 suggests the need of exploring the causes of seasonality in the occurrence of deaths due to urinary tract infections and neurological disorders. There may be environmental factors that may be responsible for the seasonality in the occurrence of deaths that need to be explored through public health perspective.

Table 2: Rayleigh test for seasonality in the season-wise occurrence of deaths from different causes of death groups.

Causes of death group		Test statistic	Tabulated value	Decision
1	Respiratory diseases	6.03903	9.21	Accept
2	Gastro-intestinal diseases	7.01272	9.21	Accept
3	Disease of urinary tract	26.07613	9.21	Reject
4	Cardio-vascular diseases	0.80810	9.21	Accept
5	Neurological disorders	17.57457	9.21	Reject
6	Accidents and injuries	1.78778	9.21	Accept
7	Endocrine diseases	2.48050	9.21	Accept
8	Viral diseases	5.55901	9.21	Accept
9	Parasitic diseases	4.64273	9.21	Accept
10	Other diseases	14.85330	9.21	Reject

Source: Authors

Prediction the Occurrence of Deaths

The occurrence of deaths for those causes of death groups in which seasonality in the occurrence of deaths is not found as confirmed by Rayleigh uniformity test follows circular uniform distribution. Table 3 enlists parameters of the predicted von Mises density of the month-wise occurrence of deaths for those causes of death groups for which no seasonality in the occurrence of deaths was found. Here, n represents the number of deaths due to the causes of death group. These parameters can be used to derive the future month-wise occurrence of deaths due to specific causes of death groups by using equation (2). We see that in case of deaths from respiratory diseases and deaths from viral diseases, most of the deaths, on average, occur in the month of April (corresponding to the circular variable group 1.57 to 2.09 radians). On the other hand, most of the deaths from gastro-intestinal diseases and endocrine diseases, on average, occur in the month of May (corresponding to the circular variable group 1.57 to 2.09 radians). In case of deaths from cardio-vascular diseases, most of the deaths occur, on average, in the month of April (corresponding to the circular variable group

0.52 to 1.04 radians). Smaller values of the concentration parameter κ indicate that the month-wise deaths are not dispersed. Since the values of κ are closer to 0, they support the hypothesis of that the occurrence of deaths are circular uniformly distributed.

Table 3: Estimated parameters of the von Mises density of the month-wise occurrence of deaths for groups having uniformly distributed occurrence of deaths for which *n* is large

Groups	n	$\hat{\mu}$	$\hat{\hat{\kappa}}$
Respiratory diseases	63	1.63585	0.00684145
Gastro-intestinal diseases	106	2.74274	0.00339955
Cardio-vascular diseases	316	0.87082	0.00022560
Endocrine diseases	33	2.45545	0.01140377
Viral diseases	72	1.64273	0.00538300

Source: Authors

Table 4 gives the estimated values of the parameters of the von Mises density function based on the deaths registered in the selected areas of Guwahati city during the years 2013 and 2014 for six causes of death groups. In these causes of death groups, no seasonality in the occurrence of deaths was detected based on the Rayleigh uniformity test. The parameters given in table 4 can be used to predict month-wise occurrence of deaths due to the specific causes by using equation (2).

Table 4: Estimated parameters of the von Mises density of the season-wise occurrence of deaths for the uniformly distributed groups of causes of deaths for which *n* is large.

Causes of death group	n	$\hat{\mu}$	$\hat{\hat{\kappa}}$
Respiratory diseases	63	1.383007	0.007589
Cardio-vascular diseases	316	1.703929	0.000183
Endocrine diseases	33	2.455454	0.013384
Viral diseases	72	1.642727	0.006373
Gastro-intestinal diseases	255	2.081425	0.001059

Source: Authors

Our analysis suggests that for all the causes of death groups where the seasonality in the occurrence of deaths, most of the deaths, on average, occur in the pre-monsoon season or during the months of March, April, and May. The value of the concentration parameter κ is also found to be small which indicates that the season-wise dispersion of deaths is not large. It may also be observed that values of κ are close to 0, This confirms that there is no seasonality in the occurrence of deaths which implies that the occurrence of deaths from these causes of death groups is circular uniformly distributed.

Summary and Conclusion

In this paper we have attempted to model the seasonality in the occurrence of death from specific causes of death groups in Guwahati city of India. Using the circular data analysis tools, we found that there is seasonality in deaths from urinary tract infections, and deaths due to neurological disorders. There is no seasonality in deaths from other causes of death groups. In these causes of death groups in which there is no seasonality in the occurrence of deaths, the seasonal distribution of deaths may be characterised through the circular uniform distribution which means that occurrence of deaths from these causes of death groups is a special case of von Mises distribution with $\kappa = 0$. The future occurrence of deaths due to these causes of death groups can be modelled with the help of the estimated parameters obtained by fitting the von Mises distribution to the reported number of deaths in the past. Our analysis suggests that there is a surge in the deaths from these causes of death groups during the pre-monsoon season or during the months of April-May are seen to be having a surge of deaths from these causes of death groups. due to these causes, both month-wise and season-wise, having very little variation. The findings of the present analysis may help in planning and programming public health activities to minimize the diseases and death burden of the people.

References

Bjornstad BJF (1990) Predictive likelihood: A review, *Statistical. Sciences*, 5: 285-492.

Butler R (1986) Predictive likelihood inference with applications (with discussion), *Journal of Royal Statistical. Society*, 48: 1-38.

Das KK (2015) Statistical Modeling of Circular Data: An Application to Health Science, UGC sponsored Major Research Report, Guwahati, Gauhati University.

Bouma MJ, Pascual M (2001) Seasonal and inter annual cycles of endemic cholera in Bengal 1891–1940 in relation to climate and geography. *Hydrobiologica* 460:147–156, 2001.

Elwood JH, Elwood JM (1987) Seasonal variation in the prevalence at birth of Anencephalus. In T Miura (Ed) *Seasonal Effects on Reproduction, Infection and Psychoses (Progress in Biometeorology, Vol. 5)*: 111–122. The Hague, SPB Academic Publishing, 1.

Cadet B, Robine JM, Leibovici D (1994) Dynamique de la mortalit´easthmatiqueenfrance: fluctuations saisonni`eres et crise de mortalit´een 1985–87. *Rev. Epid´em. et Sant´e Pub* 42:103–118.

Hakko H (2000) Seasonal variation of suicides and homicides in Finland. With special Attention to Statistical Techniques Used in Seasonality Studies. PhD thesis, Faculty of Medicine, University of Oulu, Finland.

Landler L, Ruxton GD, Malkemper EP (2018) Circular data in biology: advice for effectively implementing statistical procedures. *Behavioral Ecology and Sociobiology* 72: 128

Rao JS, Sengupta A (2001). Topics in Circular Statistics. World Scientific Publishing Co. Pte. Ltd., Singapore.

Marrero M (1983) The performance of several statistical tests for seasonality in monthly data. *Journal of Computational Statistics and Simulation* 17:275–296.

Momiyama MS (1987) Changes in seasonality of human deaths from infectious diseases. In T Miura (Ed) *Seasonal Effects on Reproduction, Infection and Psychoses (Progress in Biometeorology, Vol. 5)*: 111–122. The Hague, SPB Academic Publishing.

Yen C-J, Chao C-L, Sung F-C, Chen W-J, Liau C-S, Lee, Y-T (2000) Seasonal effects on cardiovascular mortality in older patients. *Age & Ageing* 29:186–187, 2000.

Villa S, Guisecafr´e H, Martinez H,Mun˜o O (1999) Seasonal diarrhoeal mortality among Mexican children. *Bulletin of the World Health Organization 77*(5), 375.