Biological Forum – An International Journal 16(8): 243-248(2024)

ISSN No. (Print): 0975-1130
ISSN No. (Online): 2249-3239

Fitting the Lee-Carter Model: A Statistical Approach to Mortality Forecasting

Bhupendra Meena*, Shivendra Pratap Singh, Anuj Kumar and Rahul Shukla
Department of Statistics,
Babasaheb Bhimrao Ambedkar University (A Central University) Lucknow (Uttar Pradesh), India.

(Corresponding author: Bhupendra Meena)*
(Received: 07 June 2024; Revised: 25 June 2024; Accepted: 19 July 2024; Published: 14 August 2024)
(Published by Research Trend)

ABSTRACT: The study of mortality rates has a long history which has been used in the literature of fitting and forecasting. For Part-1, I have modelled mortality rates for 21 countries in Europe using the models: Lee Carter Model this model estimates parameters which contribute in calculating mortality. The dataset comprises available data from male and female aged 0-99 from the years 1985-2014. The comparison between the fitted mortality rates of the model and the actual mortality formula has been based on total mortality rate for each specific gender. This paper focuses on investigating the evaluation of this model based on different errors.

Keywords: Mortality modelling, statistical analysis, Lee Carter Model, Fitting Errors.

INTRODUCTION

Social networks have been intensively studied by social scientists for several decades in order to understand both local phenomena, such as relationship formation and their dynamics, as well as network-wide processes, such as transmission of information. The recent substantial interest in the structural and functional properties of complex networks has been partially stimulated by attempts to understand the characteristics of social networks. In everyday social life or professional collaborations, people tend to form relationships, the existence of which is a prominent characteristic of social networks. Network formation has been studied in many research fields with their different focuses. Modelling social networks serves at least two purposes. Firstly, it helps us understand how social networks form and evolve. Secondly, in studying network-dependent social processes by simulation, such as diffusion or retrieval of information, successful network models can be used to specify the structure of interaction (Amati *et al.*, 2018).

The field of statistics has seen significant advancement in recent years, particularly in the area of social network analysis through stochastic modelling. Researchers have made considerable progress in creating statistical models that can represent the overall structure and formation of social networks. These networks are primarily shaped by two key elements: the evolving relationships among individuals, and demographic factors like age, gender, population size, and country of origin, which are typically captured by statistical models of mortality (Snijders, 2011).

Death rates play a vital role in shaping population changes and are essential to understand in various disciplines, including economics, population studies, and social sciences. Initially, mortality tables were fixed and did not account for changes over time or differences between individuals. However, as populations age more quickly in recent years, researchers have increasingly focused on developing models to predict future mortality rates. This shift in focus also includes efforts to better understand and quantify the uncertainties associated with these mortality predictions.

Initial efforts to model mortality were limited in scope, as they did not consider potential future improvements in death rates. These early models simply extrapolated from historical and current data, assuming that future mortality patterns would remain unchanged. A significant advancement in the field came with the introduction of age-continuous mortality models, which were based on early mortality laws. These laws involved fitting mathematical formulas to mortality data. The pioneering attempt to represent mortality using a continuous mathematical formula date back to 1725, when Abraham De Moivre proposed his groundbreaking approach

$$l_x = k\left(1 - \frac{x}{86}\right) \text{ for } 12 \le x \le 86$$

where l_x represents the count of surviving individuals at a specific age denoted by x. The letter k functions as a normalizing constant in the equation. A key premise of this model is that the entire population is assumed to have passed away by the time they reach 86 years of age.

Earlier mortality models were largely based on subjective interpretations rather than data extrapolation, reflecting a heavy reliance on the modelers' personal judgments. Obtaining precise mortality rates has proven to be a challenging, if not impossible, task. In recent decades, however, the most effective approach to mortality modelling has been the extrapolative method. This technique capitalizes on the increasing availability of relevant data, allowing for more accurate and data-driven predictions.

A pivotal development in early survival modelling occurred in 1825 with Gompertz's law, which introduced the concept now known as the "force" of mortality. A significant shift in approach took place in the early 1990s when researchers began using time series analysis to project future mortality trends based on historical data. These models operate on the premise that patterns observed in past data will persist into the future. Among these approaches, the Lee-Carter mortality model (https://en.wikipedia.org/wiki/Lee%E2%80%93Carter_model) stands out as the first and most widely recognized. It employs a one-factor stochastic model to represent the evolution of mortality rates over time (Booth and Tickle 2008). Over the past 10 years, the model and its variants have been used by actuaries for a wide range of purposes, from the forecasting of mortality reduction factors Renshaw and Haberman (2003) to the assessment of retirement income adequacy (Chia and Tsui 2003). Other applications in demographic science include population projections (Booth and Tickle 2003), the forecasting of sex differentials in mortality (Lee and Carter 1992b), and the projection of mortality patterns for the "oldest-old" (Buettner, 2002). Intrinsically, the core assumption of this model is that changes in mortality rates over time are primarily influenced by a single dynamic factor, referred to as the mortality index. To predict future mortality trends, the model extrapolates this index using a suitable linear time-series statistical approach. This method allows for the forecasting of death rates based on the projected evolution of this key parameter Mortality models examine diverse influences on death rates:

- **Historical trends:** These models analyze past mortality patterns and forecast future trends by considering how death rates change in relation to both age and time.
- **Contributing factors**: accountable for differences in mortality that influence the likelihood of death

The field of mortality modelling has seen significant expansion and increased sophistication since the introduction of the initial mortality law. An effective model offers a straightforward yet sufficient mathematical representation of how mortality varies with age and/or time. This paper aims to examine and compare two distinct extrapolative mortality models, focusing on summarizing their key parameters.

FUNDAMENTALS

A. Data Description

Our research utilizes mortality data from 21 European countries spanning three decades, from 1985 to 2014. We sourced this information from two reputable databases: www.mortality.org and the Eurostat database (https://ec.europa.eu/eurostat/data/database). These sources primarily compile data published or distributed by national statistical offices, ensuring reliability. The dataset comprises population and death tables, which are available separately for males and females, covering ages 0 to 99. To better illustrate trends and patterns in our data, we have categorized it into distinct age groups and calculated the total mortality rate for each country.

Table 1: This study incorporates mortality data from 21 European nations, utilizing the country abbreviations as defined by Eurostat.

Countries Abbreviation		
Austria-AT	Finland - FI	Norway – NO
Belarus - BY	France – FR	Poland - PL
Belgium - BE	Germany – DE	Portugal – PT
Bulgaria – BG	Greece – EL	Spain – ES
Czechia – CZ	Italy – IT	Slovakia – SK
Denmark -DK	Lithuania - LT	Sweden - SE
Estonia - EE	Netherlands -NL	Switzerland -CH

B. What is Mortality ?

Mortality rates measure the proportion of deaths in a population, accounting for its size and age structure over a specific period (1985-2014). This metric provides a broad assessment of population health, essentially quantifying the frequency of deaths within a given timeframe.

$$m(x,t) = \frac{c\ (x.t)}{p\ (x,t)} = \frac{\text{Deaths } (x,t)}{\text{average Population } (x,t\ )}$$

B. Age-group Mortality

To analyze mortality patterns across different demographics, we've categorized the data into age-group tables for both female and male populations. This approach allows us to account for varying population structures when comparing mortality rates. We've established five age categories: 0-19, 20-39, 40-59, 60-79, and 80-99 years. Each category summarizes mortality data for the entire study period (1985-2014). Additionally, we've calculated the total mortality for each country by aggregating mortality rates across all age groups. This classification enables us to observe and compare mortality trends between genders and across different life stages within each country.

Table 2 presents gender-specific mortality data for the 21 countries in our study. The mortality rates exhibit significant variability, fluctuating both across different age groups and from one year to the next. For reference, the country abbreviations used in this table can be found in Table 1.

Table 2: Actual Mortality Rate–Age-group.

	0-19		20-39		40-59		60-79		80-99		Total	
	F	M	F	M	F	M	F	M	F	M	F	M
AT	0.17	0.26	0.18	0.46	1.07	2.23	7.45	13.81	83.65	100.55	**92.52**	**117.31**
BE	0.17	0.26	0.21	0.48	1.17	2.13	7.33	14.28	77.18	99.19	**86.06**	**116.34**
BG	0.37	0.51	0.3	0.7	1.63	4.09	12.97	20.95	105.12	114.83	**120.39**	**141.08**
BY	0.29	0.45	0.4	1.51	2.05	6.3	12.78	25.29	85.61	101.83	**101.13**	**135.38**
CH	0.16	0.23	0.18	0.44	0.88	1.62	5.88	11.43	74.46	93.95	**81.56**	**107.67**
CZ	0.18	0.27	0.2	0.53	1.39	3.3	11.09	19.84	99.46	116.9	**112.32**	**140.84**
DE	0.16	0.23	0.19	0.42	1.13	2.27	7.84	14.45	80.23	94.29	**89.55**	**111.66**
DK	0.2	0.23	0.3	0.44	1.83	2.28	11.69	15.14	90.88	98.19	**104.9**	**116.28**
EE	0.68	0.81	0.4	1.37	1.77	4.77	11.45	23.96	48.84	122.36	**63.14**	**153.27**
EL	0.18	0.26	0.17	0.47	0.85	1.89	7.48	12.41	75	85.94	**83.68**	**100.97**
ES	0.16	0.23	0.18	0.49	0.83	1.98	5.98	12.17	71.84	88.36	**78.99**	**103.23**
FI	0.14	0.22	0.2	0.59	1.04	2.55	7.44	14.94	81.27	100.12	**90.09**	**118.42**
FR	0.16	0.24	0.22	0.56	1.04	2.45	5.66	12.14	67.15	90.42	**74.23**	**105.81**
IT	0.16	0.23	0.16	0.41	0.88	1.73	6.55	12.54	72.74	92.32	**80.49**	**107.23**
LT	0.3	0.46	0.37	1.44	1.89	5.7	10.2	20.63	84.5	92.46	**97.26**	**120.69**
NL	0.16	0.22	0.18	0.3	1.12	1.71	7.35	14.08	77.9	99.16	**86.71**	**115.47**
NO	0.33	0.23	0.21	0.42	1.24	1.72	10.36	13.05	99.25	97.13	**111.39**	**112.55**
PL	0.28	0.41	0.23	0.77	1.57	4.23	10.5	20.05	87.78	101.23	**100.36**	**126.69**
PT	0.23	0.35	0.24	0.71	1.08	2.49	7.86	14.39	79.16	98.53	**88.57**	**116.47**
SE	0.13	0.17	0.16	0.35	0.96	1.56	6.83	11.93	76.28	98.26	**84.36**	**112.27**
SK	0.26	0.36	0.22	0.64	1.54	4.20	11.53	21.25	98.93	116.2	**112.48**	**142.65**

Analysis of the data reveals consistent patterns across all countries studied: male mortality rates exceed those of females, with the highest mortality observed in the 80-99 age group and the lowest in the 0-39 range, aligning with typical demographic trends.Comparing mortality rates among countries, we find that Bulgaria exhibits the highest female mortality, surpassing Estonia (which has the lowest) by 62.39%. For males, Estonia shows the highest mortality rate, exceeding Greece (the lowest) by 41.14%. To facilitate a more standardized comparison of the mortality data presented in Table 1, we've normalized the total mortality rates to a scale of 0 to 1, as illustrated in Fig. 1. This normalization allows for a clearer visualization of relative differences in mortality rates across countries and genders.

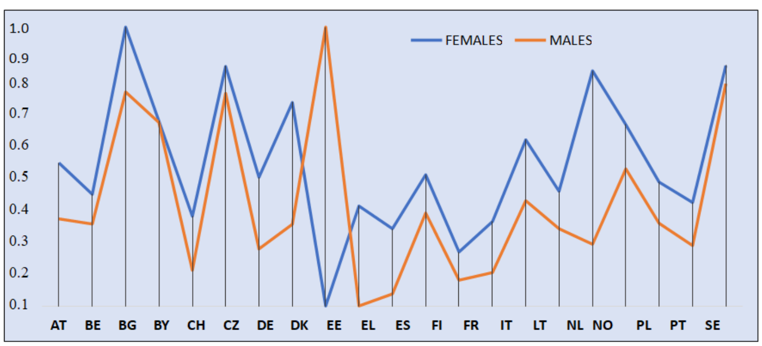

Fig. 1. Normalized Actual Mortality for Females/Males.

FITTING MORTALITY MODEL

A. Lee-Carter Model

The Lee-Carter model, introduced in 1992, pioneered stochastic longevity modelling by analysing historical mortality data and treating time trends as a stochastic process. Its key advantage is its objective approach, relying on data rather than subjective assessments or specific mortality causes. This model has since become the standard against which other stochastic mortality models are evaluated. By incorporating both age-specific and time-dependent mortality dynamics, it bases its projections on observed trends across age groups and time periods (Chavhan and Shinde 2016).

$$\log m_{xt} = a_x + b_x k_t + e_{xt}$$

Typically, x, represents the age at completion, and t, denotes year n in death and population matrices. The Singular Value Decomposition (SVD) method, a factorization technique for real or complex matrices, is employed to estimate the parameters b_x andk_t of the Lee-Carter (LC) model. Lee and Carter aim to summarize an age-period surface of log-mortality rates using $\log m(x,t)$vectors where a capture the overall mortality trend across different ages, and b indicates the rate of change in mortality relative to variations in k_t. The mortality index k_t reflects the period effect, illustrating the relationship between time-dependent events and mortality rates. The error term e_{xt} accounts for random historical fluctuations not explained by the model. This error term is assumed to be an independent and identically distributed Gaussian random variable with a mean of 0 and variance σ2 (Lee and Carter 1992).

To ensure unique parameter estimates in the Lee-Carter model, two constraints are applied to b_x and k_t. This approach addresses the identification problem that occurs when multiple solutions exist for parameter estimates. Specifically, the model requires that b_x components add up to 1 (unity), while the sum of k_t components equal 0. These constraints guarantee that the model produces a single, distinct set of parameter values.

The constraint placed on the mortality index implies a balanced distribution of the time-dependent parameter. This distribution is structured in such a way that the positive and negative values offset each other, resulting in a net sum of zero when all time-dependent parameters are added together. This approach ensures that the model captures relative changes in mortality over time, rather than absolute levels.

MORTALITY MODELLING RESULTS

A. Lee-Carter

(i) Estimating Parameters. This part of the study presents the findings from the Lee-Carter model parameter estimation. Table 3 shows the calculated values of the age-dependent parameter a_x, which covers ages 0 to 99. The estimated parameter vector a_x is calculated by taking the average of the logarithm of mortality rates across all time periods considered in the study.

$$\sum_x b_x = 1 \text{ and } \sum_t k_t = 0$$

Table 3: Age-dependent parameter ax.

	ax			ax			ax	
	F	M		F	M		F	M
AT	-609.77	-548.76	**DK**	-563.80	-550.81	**LT**	-562.43	-481.68
BE	-604.20	-547.19	**EE**	-554.87	-469.88	**NL**	-611.14	-565.65
BG	-558.38	-500.26	**EL**	-617.60	-555.68	**NO**	-581.27	-559.83
BY	-554.00	-473.10	**ES**	-622.16	-555.33	**PL**	-582.86	-508.65
CH	-622.34	-564.52	**FI**	-609.55	-540.07	**PT**	-595.60	-528.39
CZ	-589.05	-524.97	**FR**	-614.38	-546.28	**SE**	-621.70	-572.40
DE	-607.35	-551.65	**IT**	-623.07	-562.24	**SK**	-580.37	-510.25

(ii) Mortality Rate. Table 4 displays the aggregate mortality results generated by the Lee-Carter (LC) model, encompassing all ages from 0 to 99 and covering the years 1985 to 2014. When compared to the actual mortality data presented in Table 4, the LC model's output shows mortality trends that closely align with the observed historical patterns.

The graph displays red circles scattered across it, representing the observed mortality rates. These points are used to illustrate the discrepancy between the actual data and the predictions made by the LC model.

Table 4: LC mortality rate.

	F	M		F	M		F	M
AT	92.36	116.81	**DK**	104.00	115.68	**LT**	96.74	119.78
BE	85.93	115.90	**EE**	63.14	149.11	**NL**	86.62	115.26
BG	119.12	139.31	**EL**	83.39	100.72	**NO**	110.73	112.01
BY	100.45	134.58	**ES**	78.91	103.08	**PL**	100.20	126.39
CH	81.43	107.28	**FI**	89.77	117.42	**PT**	88.40	116.07
CZ	112.04	139.91	**FR**	74.15	105.69	**SE**	84.27	112.01
DE	89.38	110.96	**IT**	80.37	107.06	**SK**	111.92	140.43

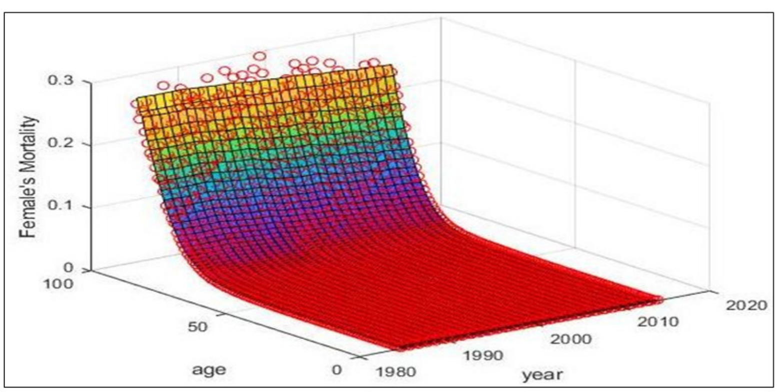

Fig. 2. Female's LC Mortality – Spain.

The Lee-Carter (LC) model predicts a slightly lower overall mortality rate for females, with a 0.1% reduction compared to actual data. Fig. 2 reveals that the LC model's peak mortality rate occurs in 1986 for

99-year-olds, while actual data shows the highest mortality in 1999 for the same age group. Conversely, the LC model indicates the lowest mortality rate in 2014 for 8-year-olds, whereas actual data places this minimum in 2010 for the same age.

The LC model underestimates total male mortality by 0.15% compared to actual data. According to the figure, the LC model predicts peak mortality for males in 2014 at age 99, whereas actual data shows it occurred in 1999 at the same age. Both the LC model and actual data indicate the lowest mortality rate in 2014, but at different ages - age 11 for the LC model and age 9 for actual data.

Fig. 4 illustrates the standardized disparities in mortality rates between the Lee-Carter model predictions and the observed mortality data for males and females.

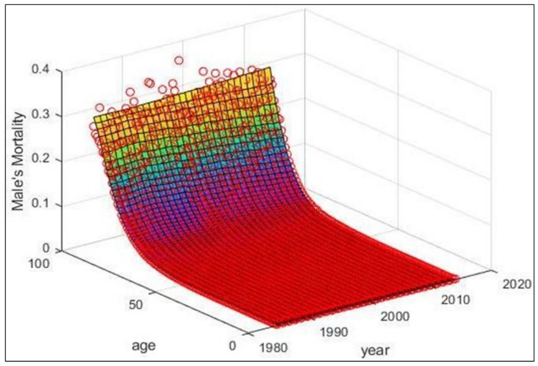

Fig. 3. Male's LC Mortality – Spain**.**

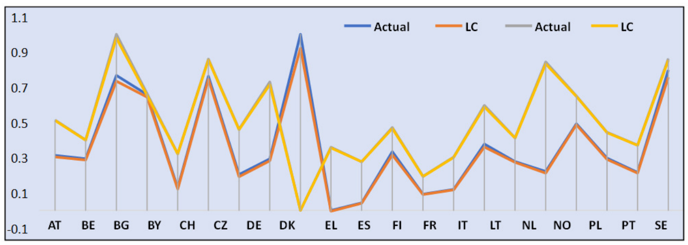

Fig. 4. Comparison between LC & Actual mortality- Female/Male (Normalized).

MEASURING FITTING ERRORS

Calculating errors is crucial for assessing model accuracy. To identify which model performs best, we evaluate four distinct error types. The initial step in error estimation involves computing the discrepancy between the mortality rates predicted by the models and the observed mortality rates. This difference represents the deviation of estimated values from actual values.

Error = Models Mortality $(\widehat{m_i})$ **−Actual Mortality** (m_i)

A. Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) measures the average magnitude of prediction errors, calculated as the standard deviation of the residuals. Residuals represent the distance between actual data points and the regression line. Table 5 presents the RMSE values for each model, broken down by gender and country, allowing for a comparison of model accuracy across different demographics.

$$RMSE = \sqrt{\sum_{I=1}^{N} \frac{(\widehat{m_i} - m_i)^2}{N}}$$

where, N = number of data points, 3000(100 ages × 30 years)

Table 5: RMSE Fitting error.

	LC	
	F	**M**
AT	0.148	0.245
BE	0.134	0.218
BG	0.389	0.385
BY	0.259	0.199
CH	0.148	0.226
CZ	0.173	0.293
DE	0.178	0.275
DK	0.336	0.26
EE	0.442	0.576
EL	0.2	0.164
ES	0.109	0.13
FI	0.208	0.327
FR	0.098	0.107
IT	0.114	0.119
LT	0.251	0.259
NL	0.109	0.148
NO	0.31	0.244
PL	0.141	0.164
PT	0.16	0.204
SE	0.109	0.179
SK	0.254	0.444

The table reveals that Estonia (EE) exhibits the highest RMSE error for the model across both genders. In contrast, France (FR) demonstrates the lowest error rates for both females and males when using the Lee-Carter (LC) model.

B. Mean Relative Error (MRE)

The Mean Relative Error (MRE) is a measure that accounts for the scale of the variable being assessed, making it useful for comparing accuracy across different magnitudes. Table 6 presents the MRE values for each model, categorized by country and gender, allowing for a comparison of model performance that takes into account the relative size of the mortality rates being predicted.

$$MRE = \frac{1}{N}\sum_{i=1}^{N}\frac{(\widehat{m_i} - m_i)}{m_i}$$

Table 6: MRE Fitting error.

	LC	
	F	M
AT	0.025	0.015
BE	0.017	0.012
BG	0.015	0.011
BY	0.016	0.016
CH	0.036	0.02
CZ	0.017	0.011
DE	0.003	0.004
DK	0.022	0.029
EE	0.053	0.028
EL	0.025	0.014
ES	0.006	0.006
FI	0.039	0.031
FR	0.003	0.003
IT	0.005	0.005
LT	0.031	0.021
NL	0.011	0.007
NO	0.031	0.031
PL	0.004	0.003
PT	0.015	0.01
SE	0.029	0.021
SK	0.026	0.018

Table 6 indicates that for females, the Lee-Carter (LC) model performs best in France (FR), showing the lowest Mean Relative Error (MRE). Conversely, the same model exhibits the highest MRE for females in Estonia (EE), suggesting less accurate predictions for this country.

CONCLUSIONS

In recent years, significant research has focused on understanding complex systems, particularly social networks. These networks exhibit intricate community structures where individuals typically belong to groups or communities characterized by dense internal connections and loose external links. This arrangement creates a hierarchy of nested social ties.A key feature of statistical models used to analyze social networks is their capacity to directly represent the underlying mechanisms that generate dependencies between network connections. Mortality statistics serve as a valuable tool for assessing social relationships, with age composition playing a crucial role in shaping various social networks, including family, work, and friendship circles. Mortality significantly influences population dynamics and holds great importance in fields such as economics, demography, and social sciences. This thesis delves into various mortality models, examining their contributions to accurately fitting predicted values to observed mortality rates.

Among the various stochastic mortality models available, the Lee-Carter model, introduced in 1992, stands out as the most widely adopted. Its popularity stems from two key advantages: firstly, it employs a relatively small number of parameters compared to alternative models, and secondly, it demonstrates notable robustness. These features contribute to its frequent use in mortality analysis and forecasting.

Lee-Carter model remains a cornerstone in mortality forecasting due to its balance of simplicity and effectiveness. While it is not without limitations, its widespread use and continuous refinement attest to its enduring value. As demographic patterns continue to evolve, the Lee-Carter model and its extensions will likely play a crucial role in understanding and predicting mortality trends, thereby informing critical decisions in both public policy and private sector planning.

Acknowledgement. I am also grateful to all contributors for their assistance with data collection, analysis, and interpretation, which made the study possible

Conflict of interest. None.

REFERENCES

Amati, V., Lomi, A., & Mira, A. (2018). Social network modeling. *Annual Review of Statistics and Its Application*, *5*(1), 343-369.

Booth, H., & Tickle, L. (2003). The future aged: new projections of Australia.'s elderly population. *Australasian Journal on Ageing*, *22*(4), 196-202.

Booth, H., & Tickle, L. (2008). Mortality modeling and forecasting: A review of methods. *Annals of actuarial science*, *3*(1-2), 3-43.

Buettner, T. (2002). Approaches and experiences in projecting mortality patterns for the oldest-old. *North American Actuarial Journal*, *6*(3), 14-29.

Chavhan, R. N., & Shinde, R. L. (2016). Modelling and forecasting mortality using the Lee-Carter model for Indian population based on decade-wise data. *Sri Lankan Journal of Applied Statistics*, *17*(1).

Chia, N. C., & Tsui, A. K. (2003). Life annuities of compulsory savings and income adequacy of the elderly in Singapore. *Journal of Pension Economics & Finance*, *2*(1), 41-65.

https://en.wikipedia.org/wiki/Lee%E2%80%93Carter_model.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American statistical association*, *87*(419), 659-671.

Renshaw, A. E., & Haberman, S. (2003). On the forecasting of mortality reduction factors. *Insurance: Mathematics and Economics*, *32*(3), 379-401.

Snijders, T. A. (2011). Statistical models for social networks. *Annual review of sociology*, *37*(1), 131-153.

How to cite this article: Bhupendra Meena, Shivendra Pratap Singh, Anuj Kumar and Rahul Shukla (2024). Fitting the Lee-Carter Model: A Statistical Approach to Mortality Forecasting. *Biological Forum – An International Journal, 16*(8): 243-248.