A COMPARATIVE STUDY OF MATHEMATICAL MODELS FOR RANDOM DATA ANALYSIS: INSIGHTS FROM CASE STUDIES

T. BHOJADAS
Research Scholar
Shri JJT University
Rajasthan.

DR. PRABHA SUDHIR RASTOGI
Professor
Shri JJT University
Rajasthan.

Dr.TVA PADMANABHA SASTRY
Professor
KMIT - Hyderabad.

ABSTRACT

Mathematical modeling studies are increasingly recognized as an important tool for evidence synthesis and to inform clinical and public health decision-making, particularly when data from systematic reviews of primary studies do not adequately answer a research question. The use of a common terminology for modeling studies across different clinical and epidemiological research fields that span infectious and non-communicable diseases will help systematic reviewers and guideline developers with the understanding, characterization, comparison, and use of mathematical modeling studies. Scientific data is often analyzed in the context of domain-specific problems, for example, failure diagnostics, predictive analysis, and computational estimation. These problems can be solved using approaches such as mathematical models or heuristic methods. In this study we compare a heuristic approach based on mining stored data with a mathematical approach based on applying state-of-the-art formulae to solve an estimation problem. However, systematic reviewers and guideline developers may struggle with using the results of modeling studies, because, at least in part, of the lack of a common understanding of concepts and terminology between evidence synthesis experts and mathematical modellers. The goal is to estimate results of scientific experiments given their input conditions.

Keywords: *Mathematical modeling, scientific data, failure diagnostics, predictive analysis, mathematical modellers, input conditions*.

INTRODUCTION

In particular, models and their projections/forecasts have received unprecedented attention. With a multitude of modeling frameworks, underlying assumptions, available datasets and the region/timeframe being modeled, these projections have varied widely, causing confusion among end-users and consumers. We believe an overview (non-exhaustive) of the current modeling landscape will benefit the readers and also serve as a historical record for future efforts. There is no life without disease. Therefore, over the years, people have spent a lot of effort and financial and moral support to control the spread of diseases or avoid their side effects. The side effects of the spread of diseases vary from health damage to psychological impacts, social impacts, and economic impacts. With the scientific and technological developments of humankind, studies of problems and setting up scenarios for dealing with them have become one of the most important strategies. Accordingly, nations can avoid the serious side effects of such problems. Among the most prominent of these problems is the spread of infectious diseases, whether they are bacterial or viral. The category of epidemic diseases is one of the most difficult and dangerous disease types that civilization has faced. The dynamics of transmission of almost all infectious diseases are known, so the establishments of good mathematical models for the spread of infectious diseases are reachable tasks. Mathematical models are considered among the most major pre-studies that can be used in developing strategies for dealing with epidemics.

LITERATURE REVIEW

M. H. M. Hassan (2022) A local performance of the SIR model on actual data is introduced. A good approximation of the SIR model parameters in Saudi Arabia during a period of 275 days is determined. The parameters are estimated from the recorded data and used to predict the values in the next subsequent period. The performance of the standard fourth order Runge–Kutta method is considered for the classical SIR models over different periods. A comparison of the recorded data and the predicted values during the considered period illustrated the effectiveness of the treatment. The mathematical properties and initial conditions are considered within the estimated parameter values. It is shown that lockdown and social distance attitudes effectively controlled the spread of the disease. The maximum number of daily active infected cases is 63, 026, and occurs in July and this agrees with the calculated values.

M.A. Masud (2022) The COVID-19 pandemic has caused a global crisis with 47, 209, 305 confirmed cases and 1, 209, 505 confirmed deaths worldwide as of November 2, 2020. Forecasting confirmed cases and understanding the virus dynamics is necessary to provide valuable insights into the growth of the outbreak and facilitate policy-making regarding virus containment and utilization of medical resources. In this study, we applied a mathematical epidemic model (MEM), statistical model, and recurrent neural network (RNN) variants to forecast the cumulative confirmed cases. We proposed a reproducible framework for RNN variants that addressed the stochastic nature of RNN variants leveraging z-score outlier detection. We incorporated heterogeneity in susceptibility into the MEM considering lockdowns and the dynamic dependency of the transmission and identification rates which were estimated using Poisson likelihood fitting.

Abdu Gumaei (2021) Social media is popular in our society right now. People are using social media platforms to purchase various products. We collected the data from various social media platforms. We analyzed the data for prediction of the consumer behavior on the social media platform. We considered the consumer data from Facebook, Twitter, Linked In and YouTube, Instagram, and Pinterest, etc. There are diverse and high-speed, high volume data which are coming from social media platform, so we used predictive big data analytics. In this paper, we have used the concept of big data technology to process data and analyze it to predict consumer behavior on social media. We have analyzed consumer behavior on social media platforms based on some parameters and criteria. We analyzed the consumer perception, attitude towards the social media platform. To get good quality of result, we pre-process data using various data pre-processing to detect outlier, noises, error, and duplicate record.

Abdullah M. Almeshal (2020) The state of Kuwait is facing a substantial challenge in responding to the spread of the novel coronavirus 2019 (COVID-19). The government's decision to repatriate stranded citizens back to Kuwait from various COVID-19 epicenters has generated a great concern. It has heightened the need for prediction models to estimate the epidemic size.

Mathematical modeling plays a pivotal role in predicting the spread of infectious diseases to enable policymakers to implement various health and safety measures to contain the spread. This research presents a forecast of the COVID-19 epidemic size in Kuwait based on the confirmed data. Deterministic and stochastic modeling approaches were used to estimate the size of COVID-19 spread in Kuwait and determine its ending phase. In addition, various simulation scenarios were conducted to demonstrate the effectiveness of non-pharmaceutical intervention measures, particularly with time-varying infection rates and individual contact numbers.

Anil Vullikanti (2020) COVID-19 pandemic represents an unprecedented global health crisis in the last 100 years. It's economic, social and health impact continues to grow and is likely to end up as one of the worst global disasters since the 1918 pandemic and the World Wars. Mathematical models have played an important role in the ongoing crisis; they have been used to inform public policies and have been instrumental in many of the social distancing measures that were instituted worldwide. In this article, we review some of the important mathematical models used to support the ongoing planning and response efforts. These models differ in their use, their mathematical form and their scope.

Mathematical model

A mathematical model is an abstract description of a concrete system using mathematical concepts and language. The process of developing a mathematical model is termed mathematical modeling. Mathematical models are used in applied mathematics and in the natural sciences (such as physics, biology, earth science, chemistry) and engineering disciplines (such as computer science, electrical engineering), as well as in non-physical systems such as the social sciences (such as economics, psychology, sociology, political science). It can also be taught as a subject in its own right. The use of mathematical models to solve problems in business or military operations is a large part of the field of operations research. Mathematical models are also used in music, linguistics, and philosophy (for example, intensively in analytic philosophy). A model may help to explain a system and to study the effects of different components, and to make predictions about behavior.

Mathematical statistics

Mathematical statistics is the application of probability theory, a branch of mathematics, to statistics, as opposed to techniques for collecting statistical data. Specific mathematical techniques which are used for this include mathematical analysis, linear algebra, stochastic analysis, differential equations, and measure theory. Statistical data collection is concerned with the planning of studies, especially with the design of randomized experiments and with the planning of surveys using random sampling. The initial analysis of the data often follows the study protocol specified prior to the study being conducted. The data from a study can also be analyzed to consider secondary hypotheses inspired by the initial results, or to suggest new studies. A secondary analysis of the data from a planned study uses tools from data analysis, and the process of doing this is mathematical statistics.

Probability distributions

A probability distribution is a function that assigns a probability to each measurable subset of the possible outcomes of a random experiment, survey, or procedure of statistical inference. Examples are found in experiments whose sample space is non-numerical, where the distribution would be a categorical distribution; experiments whose sample space is encoded by discrete random variables, where the distribution can be specified by a probability mass function; and experiments with sample spaces encoded by continuous random variables, where the distribution can be specified by a probability density function. More complex experiments, such as those involving stochastic processes defined in continuous time, may demand the use of more general probability measures.

Statistical Modelling

In the more modest words, Statistical Modeling is an interpreter, mathematically-prescribed method to approximate truth which is being generated by the data and for making forecasts out of this approximation. For example, depicting a quantity through an average and a standard deviation is the simple form of statistical modelling. And here, the statistical model is the mathematical expression that is being deployed. "Statistical Modelling is simply the method of implementing statistical analysis to a dataset where a Statistical Model is a mathematical representation of observed data."

RESEARCH METHODOLOGY

Aiming at creating conditions for analyzing the association among curricular contents and application of mathematical modeling with the support of technology in the students daily life situations, the first author of this study conducted, during the second semester of 2007, a pedagogical experiment on the Linear Program discipline, which is part of the Information Systems Course at a Private College in Campinas, where the students were encouraged to work on real situations. The project evaluation represents 20% of the final students' grade and, for such evaluation, it's taken into consideration: a) the complexity of the problem chosen; b) the mathematical formulation as well as the justification for the objective function, the restrictions and the variables; c) software simulations for theory checking and sensitivity analysis. In this experience, as previously said, we have tried to emphasize the construction of knowledge and to make the students more critical and with a higher arguing skill. Linear Programming discipline is taught on the third year of Information Systems Course and, at that point, the students usually are already working, not as interns, but rather as regular employees. This way, the time spent with school activities is very low, mainly for mathematical disciplines which are considered as supporting subjects to the ones specific to the students' formation. At such assignments the students (groups of two) chose problems related to what is being taught at the discipline, search for data, model the problems, that is, try to mathematically represent it, solve the problems using software required that are available, analyze and validate, whenever possible, the solution found.

RESULTS AND DISCUSSION

Test for Equality of Proportions for Number of Passed Students

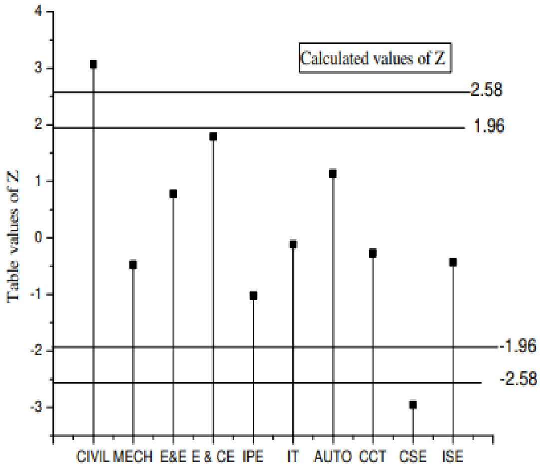
The Z-Test for proportion of number of passed students is calculated and modulus of Z for different branches for the year

2007 under and 2008 [A3.1] under autonomous were tabulated in Table 1. The calculated value of Z in the Civil Engineering branch is 3.0764 and in Computer Science branch is |-2.9537| are greater than the tabulated value of 2.58 at 1% level of significance. So, null hypothesis is rejected. For remaining branches the calculated value of Z is smaller than the required value at 1% level of significance, hence null hypothesis is accepted.

Table 1: Calculated Z value for number of passed students

	VTUJanuary-2007				PDAJanuary-2008							
Sl. No.	Name of the branch	Number of students appeared	Number of passed students	pi=xi/ni	Number of students appeared examination	Number of passed students	pi=xi/ni	pi-pi	p=(x1+x2)/(n1+n2)	q=1 - p	SQRT{p*q(1/n1+1/n2)	Z calculated value for number of passed students |Z|
1	CIVIL	99	28	0.28283	101	50	0.4950	0.21222	0.3900	0.6100	0.0690	3.07649
2	Mech	93	40	0.43011	96	38	0.39583	-0.0343	0.4127	0.5873	0.0716	-0.4785
3	ELECT	41	19	0.46341	40	22	0.5500	0.08659	0.50617	0.49383	0.1111	0.77927
4	E&CE	121	83	0.68595	118	93	0.78814	0.10219	0.7364	0.2636	0.0570	1.79264
5	IPE	39	11	0.28205	43	8	0.1860	-0.0960	0.23171	0.76829	0.0933	-1.0290
6	IT	29	8	0.27586	38	10	0.26316	-0.0127	0.26866	0.73134	0.1093	-0.1162
7	AUTO	16	0	0.00000	26	2	0.07692	0.07692	0.04762	0.95238	0.0677	1.1368
8	CCT	17	3	0.17647	34	5	0.14706	-0.0294	0.15686	0.84314	0.1080	-0.2723
9	CSE	88	55	0.62500	98	40	0.40816	-0.2168	0.51075	0.48925	0.0734	-2.9537
10	ISE	43	21	0.48837	43	19	0.44186	-0.0465	0.46512	0.53488	0.1076	-0.4324

The graphical view of the above Z-calculated values is shown in Graph 1. From this we can say that the results of Civil and Computer Science for VTU affiliated and autonomous status are not equal.



Graph 1: Calculated value of Z for number of passed students

Since Civil has positive value of Z, so the result of January 2008 is better than January 2007. Since Z value of CSc is negative which indicates that the result of January, 2007 is better than that of January.2008. For rest of the branches the results of both the years are equal.

Test for Equality of Variances for Number of Passed Students

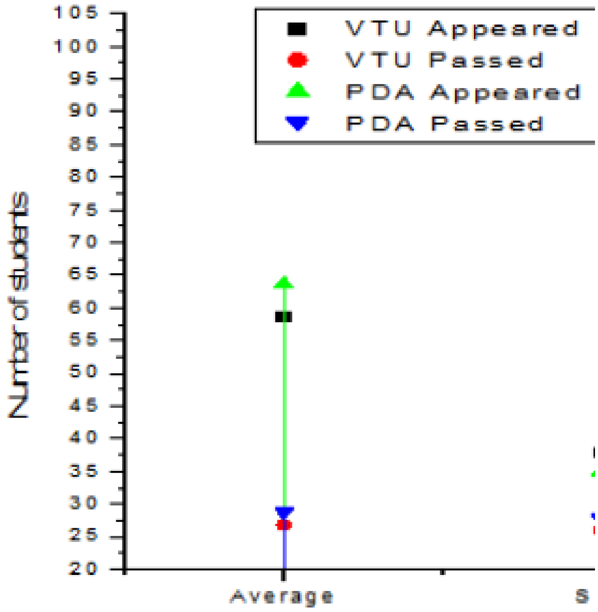
The average number of passed students, its standard deviation and coefficient of variance of the results of VTU affiliated 2007 and autonomous status 2008 is tabulated in Table 2.

Table 2: Variance ratio test results of number of passed students

	VTUJan-		P

	2007			DAJan-2008	
Sl.No.	**Name of the branch**	**Number of students appeared examination**	**Number of passed Students**	**Number of students appeared examination**	**Number of passed students**
1	CIVIL	99	28	101	50
2	Mech	93	40	96	38
3	ELECT	41	19	40	22
4	E&CE	121	83	118	93
5	IPE	39	11	43	8
6	IT	29	8	38	10
7	AUTO	16	0	26	2
8	CCT	17	3	34	5
9	CSE	88	55	98	40
10	ISE	43	21	43	19
	Total	**586**	**268**	**637**	**287**
	Average	58.6	26.8	63.7	28.7
	SD	37.907	26.059	34.868	27.877
	CV	64.688	97.235	54.738	97.132
VARIANCE RATIO TEST					**1.0011**

The coefficient of variance of 2007 and 2008 results are 97.235 and 97.132 respectively. The coefficient of 2008 results is lower than that of 2007 results, hence, results of 2008 is more consistent. The calculated value of F (1.0011) is less than the F – table value (3.1789) at 5 % level of significance for (n1 – 1), (n2 – 1) degrees of freedom i.e. (9, 9) degrees of freedom the table value is 3.1789. Thus the null hypothesis is accepted.



Graph 2: Average standard deviation and coefficient of variance of number of passed students

CONCLUSION

The SIR mathematical models, despite their simplicity, are still effective models that give acceptable results within limited time periods. Determination of the parameters of the models must be determined with more reliable techniques and should be updated continuously. The integration of machine learning, Bayesian methods, blockchain, quantum computing, and ethical considerations reflects the evolving landscape of data science. Interdisciplinary collaboration and the development of models with real-time analytics capabilities emphasize the need for adaptability and responsiveness to changing data patterns. Additionally, the emphasis on ethical considerations underscores the importance of responsible data handling and analysis. In essence, the approach towards mathematical and statistical methods for random data computing is dynamic, continually evolving to meet the challenges posed by modern datasets. Researchers and practitioners in this field must stay abreast of emerging technologies and methodologies to contribute to the ongoing progress in data science and its applications across various domains. Lockdown is the safest strategy despite its side effects in case of the absence of medical solutions. The introduced local forms of the SIR models are a remedy for the short time scale of the epidemic mathematical models and can be applied to other models effectively. Our calculated results are in high agreement with the recorded reports due to our local (short time) treatment.

REFERENCE

1. *M. H. M. Hassan (2022), "A Comparative Study for Some Mathematical Models of Epidemic Diseases with Application to Strategic Management", Appl.Sci.,ISSNno:2076-3417,Vol.12(24),Pages.12639.https://doi.org/10.3390/ app122412639*
2. *Abdullah M. Almeshal (2020), "Forecasting the Spread of COVID-19 in Kuwait Using Compartmental and Logistic Regression Models",Applied Sciences,ISSNno:2076-3417,Vol.10(10),Pages.3402.DOI:10.3390/app101034 02*
3. *M.A. Masud (2022), "Comparative study of a mathematical epidemic model, statistical modeling, and deep learning for COVID-19 forecasting and management",Socio-Economic Planning Sciences,ISSNno:0038-0121,Vol.80, Pages.101249.https://doi.org/10.1016/j.seps.2022.101249*
4. *Anil Vullikanti (2020), "Mathematical Models for COVID-19 Pandemic: A Comparative Analysis",Journal of the Indian Institute of Science,ISSNno:00 19-4964,Vol.100,Pages.793-807.https://doi.org/10.1007/s41745-020-00200-6*
5. *Abdu Gumaei (2021), "Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics",Journal*

of Big Data,ISSNno:2196-1115,Vol.8,https://doi.org/10.1186/s40537-021-004 66-2

6. *Dr Rohini Jh (2020) "Review of Data Mining and Data Warehousing Implementation in E-Governance", International Journal of Innovative Science and Research Technology,ISSNno:2456-2165,Vol.5,Issue.10.*
7. *Zaher Al-Sai (2017) "Big Data and E-government: A review", Conference: 2017 8th International Conference on Information Technology (ICIT)At: Jordan – Amman,Page:580–587.*
8. *Ohood M (2021) "A Survey of Blockchain and E-governance applications: Security and Privacy issues",Turkish Journal of Computer and Mathematics Education,Vol.12,No.10,Pages.3126-3134.*
9. *Hariom Kumar Solanki (2016) "Road Transport in Urban India: Its Implications on Health",Indian J Community Med.Vol.41(1),Pages.16–22. doi:10.4103/0970-0218.170959*
10. *Dinesh Sethi (2007) "Preventing the leading cause of death in young people in Europe",J Epidemiol Community Health,Vol.61(10),Pages.842–843.doi:10. 1136/jech.2007.063081*