CHAPTER 51

STATISTICS

Doctoral Theses

01. AGGARWAL (Sanya)
Robust Statistical Models for Health Economic Evaluation: Analysing the Burden of Covid-19 and HIV/AIDS.
Supervisors: Prof. Gurprit Grover and Dr. Sangeeta Chakravarty
Th 26915

Abstract

The considerable advancements in technologies and treatments available on various disease has increased the survival rates of the population that have shifted the focus from mortality to morbidity. Economic evaluation studies have assumed an eminent role that provides the framework for measuring, valuing, and comparing the costs and benefits of different health care interventions and addressing various aspects of health policy. To model complex issues related to competing interventions, statistical models have become prominent part of health economic evaluation studies. A comprehensive measure that allows to aggregate the health outcomes in terms of total years of life lost in a population is Disability Adjusted Life Years (DALY). It was developed to quantify the burden of disease, can be utilised in cost-effectiveness analysis as well as to set priorities for resource allocation. Despite the advantages of being a uniform measure of population health, the conceptual framework of DALY has many gaps, as it is non-representative of the impact of multiple morbidities and the calculations are perplexing in case of censored observations. Thus, there is an urgent need to develop valid and efficient statistical models for addressing various gaps in DALYs estimation. A key goal of this thesis is to enumerate practical application and development of robust statistical techniques by covering different aspects of health economics. The existing literature on the topic has been utilized and attempts are made to bring forth the application of core theories of sub-discipline. For this purpose, two types of retrospective datasets are utilized in this thesis, viz, population level data related to COVID-19 in India and SAARC nations that is utilized in Chapters 2, 3 and 4. The second is the individual level data based on the medical records of patients diagnosed with HIV/AIDS in Delhi, India. The estimations are done on both types of datasets to highlight the widespread application of statistical techniques, which can also be generalized.

Contents

in HIV-AIDS risk groups 7. Quantile regression approach in analyzing the effect of lifestyle factors on disability adjusted life years of HIV/AIDS infected adults in India 8. Conclusion and future studies. References List of publication.

02. ALI (Danish)
Bayesian Inference for Lifetime Models: Some Contributions.
Supervisor: Prof. Ranjita Pandey
Th 26916

Abstract

Bayesian inference regards unknown parameter as a random variable. The process of fitting a probability model to a collection of data and subsequently synthesizing the results in conjunction with the probability distribution of the model parameters and the observed quantity is main aim of any Bayesian model assessment study. Future outcomes are thus probabilisticaly predicted using Bayes paradigm. The core component of the Bayesian approach is the explicit use of probability to measure uncertainty for unidentified population characteristics. Bayesian data analysis is accomplished using the four idealistic stages that consist of 1.Discovering a concrete probability model that contains a combined probability distribution for all the visible and unobservable variable components of the sample. 2.In the distribution above, a chance distribution is given for the unknown parameter. 3.Computation of the posterior probability distribution, which, given the available data, represents the conditional probability distributions of important but unobserved variables. 4.Evaluating the model`s fit and summarizing the posterior distribution that was computed For some proposed lifetime models we develop distributional aspects and mathematical properties such as statistical moments, Renyi and Shannon entropies, incomplete moments, mean residual life, Bonferroni and Lorenz curves with an emphasis on modelling capability with respect to patterns of the corresponding hazard rate functions. Bayes estimation is undertaken under SELF, LINEX and GELF environments. We use Type II, Generalized Type II Progressive Hybrid, Generalized Type I Progressive Hybrid Censoring plans in addition to parametric estimation under complete sample set up. Real data applications and simulations demonstrate proximity of the model under consideration To compare the new model to the most significant alternative models already in use, we specifically focus on the underlying nature of data from the lifetime process in respect of its symmetricity in order to establish that the suggested model works better than the existing rival models.

Contents

1. Introduction 2. Nakagami distribution Under Generalized Type-II progressive Hybrid Censoring Scheme 3. Gamma Rayleigh distribution type II censoring 4. Sine generated pareto distribution of two parameters 5. Hjorth distribution under generalized type I progressive hybrid censoring scheme 6. Modified alpha power Rayleigh distribution 7. Generalized gamma distribution under generalized type-II progressive hybrid censoring scheme 8. Modified beta distribution of second kind. Bibliography. List of publication and communicated research work. list of papers presented in conferences.

03. JAIN (Neetu)
Some Contributions to Moments of Order Statistics and Their Inferential Aspects.
Supervisor: Prof. Narinder Pushkarna
Th 26917

Abstract

This thesis focuses on the work on the exact expressions with the recurrence relations for single and product moments of ordered random variables such as order statistics, record values, generalized order statistics and dual generalized order statistics etc., from some specific continuous distributions. These recurrence relations are very important in numerical evaluation and interpretation of the moments of ordered random variables in a simple recursive manner. The various statistical properties of the distribution under consideration can be studied with the help of these moments. Conditional moments are also obtained for generalized order statistics as characterizations and special cases for record values and order statistics have been deduced. Estimation of parameters through various methods such as L-moments, TL- moments and maximum likelihood approach are done for some particular continuous distribution. Apart from this, it also deals with stress-strength reliability and its estimation when stress, strength variates are given by two independent random variables following a continuous distribution.

Contents

04. KOMAL

Some Estimation Procedures of Stress-Strength Reliability Using Record Values.

Supervisor: Dr. Ashish Kumar Shukla

<u>Th 27115</u>

Abstract

Reliability is usually defined as the probability that a system will perform its intended function for a specified period under a set of specified environmental conditions. Several lifetime distributions are useful in modeling lifetime data. A Lifetime distribution describes, mathematically, the length of the life of a system or a device. Reliability estimation is a common topic of research, and it gained more attention in the last few years. A reliability system containing two random variables as a strength X and stress Y applies to many real-life problems. One key issue is that it is not always possible to monitor the failure times of all test units or objects. Due to the breakage of test units, time constraints, findings, and other factors, the experimenter may have to eliminate some test units or things from the test, either mistakenly or intentionally. In reliability and life testing, the reliability function Under the stress-strength modeling, the performance of a system can be measured as the stress-strength reliability P=P(X >Y) where X represents the strength and Y is the stress of the system. The system will break down when the system's stress exceeds its strength. In Chapter 2, our aim is to obtain Some improved estimators and confidence intervals of the parametric functions are proposed based on record values from the family of lifetime distributions is considered. We Study Preliminary

test estimators (PTEs) for the power of parameter and reliability function specifically R(t) and P for record data. The Preliminary test Confidence interval (PTCI) are also obtained based on ML and UMVU estimator. For PTCI of the parameter we derive the coverage probability and expected length and show that the confidence interval based on MLE are more accurate than the expected length. In Chapter 3, Generalized Inverted Scale family of distribution has been considered. PTEs for power of parameter and two reliability measure R(t) and P are obtained. PTCI are also developed based on UMVUE, MLE. A Study to compare the purposed estimator with ordinary estimator uniformly minimum variance unbiased estimators and maximum likelihood estimator using simulated and real data is performed A new approach for obtaining these estimators is devised, in which power of parameter estimators are obtained first and help to derive R(t) and P. In Chapter 4, Topp-Leone distribution is considered. Two measures of reliability are discussed, namely R(t) and P. Based on records estimation of the parameters, R(t) and P are constructed. Point estimators are constructed for UMVUE and MLE. Asymptotic confidence intervals of the parameters based on MLE and log transformed MLE are constructed. Testing procedures are developed for the hypothesis related to different parametric functions. Real example is used to illustrate the results. Following In Chapter 5, Moore and Bilikam family of lifetime distributions has been considered. Assuming we are aware of some prior guess value of the parametric functions, Shrinkage estimators are developed for the powers of parameter, R(t) and P based on records, which incorporate this prior information. Simulation studies are conducted to judge the performance of these new estimators. Next, In chapter 6, Bayesian estimation procedures based on records for estimation of the parameters, R(t) and P are developed. Two type of loss functions namely Squared error loss function (SELF) and General entropy loss function (GELF) are considered. A new method of obtaining these estimators is utilized which consists the use of estimator of powers of parameter in obtaining the estimators of various reliability measures based on record value. Various risk functions are obtained.

Contents

05. MAGAN (Radhika)
Quantifying the Estimation of quality of Life by Means of Statistical Models.
Supervisor: Prof. Gurprit Grover
Th 26918

Abstract

Health economics in an adaptive area of economics where we study the development of health outcomes. The methodology covered under this area of research includes a set of frameworks and theories of statistics. The healthcare management areas include hospitals, clinic and public as private health care centres which further focus on the application of economic theories in healthcare sector. These theories include consumer- producer, demand-supply, and production with efficiency etc for

a continual procedure to study with respect to health outcome response. Apart from the theories as stated in past literature, we need to recognize and comprehend the behavior of individuals at patient level. This nature of preference of the individuals for a specific disease concern helps in economic evaluations of existing versus new thechnologies in the field. India is a densely populated country with restricted infrastructure for healthcare systems in order to rackle a pandemic. With increased demands of hospital beds, the state and central government are working towards creation of new corona isolation wards, medical equipment like ventilators, testing kits, personal protective quipment (PPE) kits, sanitizers, masks etc. Ranjan (2020) clearly states that the immediate action of lockdown imposed by the Indian government proved to be fruitful in combating the spread of infection.

Contents

06. PANT (Seema)
Some Contribution to Cure Rate Estimation Under Various Conditions.
Supervisors: Prof. Gurprit Grover and Dr. Manoj Kumar Varshney
Th 26919

Abstract

In the first chapter we have introduced cure fraction models with related basic concepts. In the second chapter we have introduced Power Gompertz Distribution as the baseline survival distribution and have estimated the cure fraction using cure rate models and have also identified the significant predictors. The mixture cure rate model assumes that a single cause is responsible for occurrence of an event .This assumption is not always true as the event of interest may happen due to different latent competing causes during the course of clinical trials. It further assumes the presence of a proportional hazard structure in uncured patients and not in cure patients. To overcome these shortcomings an alternative model called the Promotional Time Cure Rate Model has been utilized to estimate the cure fraction in chapter 3. In chapter 4, we have used pairwise correlation to reduce the number of predictors when two or more variables are correlated/ associated with each other and have studied its effect on cure fraction estimation In chapter 5, we have developed a non-mixture cure rate model with Weibull distribution while incorporating canonical correlation technique to reduce the number of prognostic factors .The developed model is found to be a better model for the estimation of the cure fraction as compared to the one with all predictors. The diabetic patients can be divided into two categories: those who develop nephropathy due to long term diabetes are classified as susceptible and those who do not develop nephropathy are classified as immunes. Assuming the diabetic population to be composed of cured (immunes) and uncured (susceptible), we have estimated the cure fraction of diabetic patients in chapter 6.

Contents

1. Introduction 2. On the estimation of Cure fraction using power gompertz distribution under Bayesian approach 3. On the estimation of Cure rate in the presence of prognostic factors using various discrete count distribution 4. On the estimation of Cure rate after reducing the number of prognostic factors using pairwise correlation 5. On the estimation of Cure rate after reducing the number of prognostic factors using canonical correlation technique 6. On the estimation of Cure fraction of diabetic nephropathy patients 8. Conclusion and future perspective. References List of publication.

07. SAINI (Shubham)
Some Contributions to Estimation Procedures for Lifetime Models Using Censored Data.
Supervisor: Dr. Sachin Tomer
Th 27116

Abstract

Reliability is usually defined as the probability that a system will perform its intended function for a specified period under a set of specified environmental conditions. Several lifetime distributions are useful in modeling lifetime data. A Lifetime distribution describes, mathematically, the length of the life of a system or a device. Reliability estimation is a common topic of research, and it gained more attention in the last few years. A reliability system containing two random variables as a strength X and stress Y applies to many real-life problems. One key issue is that it is not always possible to monitor the failure times of all test units or objects. Due to the breakage of test units, time constraints, findings, and other factors, the experimenter may have to eliminate some test units or things from the test, either mistakenly or intentionally. As a result, censored samples are preferable to entire samples in life testing investigations. In reliability and life testing, the reliability function $R(t) = P(X > t)$ is defined as the probability of failure-free operation until time t, where X denotes the lifetime of an item. Under the stress-strength modeling, the performance of a system can be measured as the stressstrength reliability 8 = $P(X > Y)$ where X represents the strength and Y is the stress of the system. The system will break down when the system's stress exceeds its strength. In the first chapter, we provide an introductory overview of the Reliability theory and Censoring.

Contents

1. Introduction 2. Statistical for the reliability characteristics of kumaraswamy-G family of distributions using progressively first-failure censored data 3. Statistical inference for the reliability characteristics of generalized Maxwell lifetime model using progressively first-failure censored data 4. Classical and Bayesian estimation of multicomponent stress-strength reliability foe top-leone distribution using progressively censored data 5. Classical and Bayesian estimation of multicomponents stress-strength reliability for burr XII lifetime model using progressively first failure censored data 6. Classical and Bayesian estimation of multicomponents stress-strength reliability for kumaraswamy-G family of distributions using progressively first-failure censored data. Bibliography.

08. SHARMA (Ankita)
On the Development of Models for Studying Progression of Type 2 Diabetes
Supervisors: Prof. Gurprit Grover and Dr. Varshney
Th 27117

Abstract

This thesis is focused on the development of Markov and Survival models concerning the diagnosis, progression, survival and identification of prognostic factors for Type 2 diabetes. Markov models are developed to estimate the transition probabilities, mean sojourn time for each state and to estimate the effect of hidden factors on transition probabilities. In addition to this, survival models are developed to study the effect of prognostic factors on the progression of Type 2 diabetes. This chapter provides a brief introduction to types and early symptoms of Diabetes. This chapter also presents a review of literature and methods used in this thesis This chapter gives the information about the future prospective and modification in the proposed models. Research is a continuous and endless process beyond the limits for the progress and development. To conduct this research, retrospective data collected from All India Institute of Medical Science (AIIMS), during the period 2000 to 2006 and North Delhi Diabetic Centre, Delhi, India; during the period of 2005 to 2018. The data collected from AIIMS is used in Chapter 2 and 3 while the data collected from North Diabetic Centre, Delhi is used in Chapter 4, 5 and 6. The patients for both the datasets were classified as per WHO and ADA criteria on the basis of HbA1c.

Contents

09. SRIVASTAVA (Pulkit)

Bayesian Parametric Estimation of Log-logistic and Related Models Under Different Gensoring Schemes.

Supervisor: Prof. Pandey Ranjita

Th 27118

Abstract

The basic purpose of statistical inference is to extract information from data or observations of a stochastic phenomenon. Bayesian and classical inferences are the two broad classifications of statistical inference. When an experiment is carried out numerous times, classical inference perceives probability as the upper bound of the relative rate of recurrence of an event, without taking into consideration the experiment's background information. Additionally, the classical approach regards parameters as fixed. While in Bayesian reasoning, the probability indicates one's level of trust in the likelihood of an event and is therefore made up of recent evidence (expressed by a likelihood function) and prior knowledge (represents degree of belied). Parameters ate viewed as random variables in the Bayesian approach. When used for data analysis, the Bayesian approach yields more significant outcome than its classical counterpart, especially for small data sets. Consequently, probabilities are not sole characteristics of random variables, according to Bayesian theory, but rather a quantifiable coding of one's level of knowledge.

Contents

10. TOLANI (Himabshu)
Assessment of Event Dynamics Through Bayesian Models.
Supervisor: Prof. Ranjita Pandey
Th 26920

Abstract

The fundamental objective of statistical inference is to draw conclusions about a stochastic phenomenon from data or observations. Statistical inference can be broadly divided into two types: Bayesian and classical. Without taking into account the background information of the experiment, classical inference interprets probability as the upper bound on the relative rate of recurrence of an event when an experiment is conducted repeatedly. Furthermore, the traditional paradigm sees unknown population parameters as fixed. While in Bayesian reasoning, the probability denotes one's level of confidence in an event's possibility and is therefore made up of current data (represented by a likelihood function) and prior knowledge (represents degree of belief). The Bayesian method treats unknown population parameters as random variables. When used to data analysis, the Bayesian methodology produces results that are considerably more significant than those produced by the traditional method. Therefore, according to Bayesian theory, probabilities are not properties of random variables but rather a measurable coding of one's level of knowledge. Present work comprises of assessment of some events observed in Delhi and in India through statistical models fitted under Bayesian paradigm. The motive behind conducting this research was to analyse the considered events and explain their dynamics through covariates which have been less explored in context of India through Bayesian lens. We have majorly under took four different analysis consisting of events like “property crime in districts of Delhi, dynamics explained through spatio-temporal effects”, “respiratory disorders (RD) in districts of Delhi during winters, dynamics explained through intensity of pollutants and spatio-temporal effects”, “time-varying reproduction number for Covid-19, dynamics explained through Bayesian semi-parametric splines regression” and “batting ability of some batsman of India before and during ICC ODI world cup 2019”, dynamics explained through Bayesian assessment of playing ability.

Contents

11. VINIT (Parmeet Kumar)
On the Development of Some Insurance Plans for HIV/AIDS and Cancer Patients.
Supervisor: Prof. Gurprit Grover
Th 27119

Abstract

An initiative, to fill the void between insurance sector and patients of deadly diseases like HIV/AIDS and Cancer (Acute Lymphoblastic Leukemia (ALL)), is taken here. To achieve this the survival is estimated in different situations. Later, prognostic factors are included by incorporating the Cox-PH model and hence the premium cost is estimated. Single decrement and multiple decrements are devised and deterministic model is developed to estimate the premium cost under various conditions. Significant prognostic factors are identified. The premium cost varies from Rs. 4.16 to Rs. 19.18 for HIV/AIDS patients and Rs. 4.7 to Rs. 19.6 for ALL patients. Finally, Cure fraction model is introduced in Bayesian framework. The premium cost does not change much over the years in Bayesian setup. For a sum of Rs. 100 insured for a year the lowest premium cost payable is Rs. 1.45 in case of HIV/AIDS patients and Rs. 4.21 for ALL patients. Overall, it is observed that as the survival rate decreases, the premium cost increases. The study emphasizes for inclusion of HIV/AIDS and cancer patients in the insurance sector. These premium plans are practical for insurers and economical for the insured due to the greater survival rate and low premium cost. On rigorous work carried out to bridge the two inseparable entities, the insurance and the patients of deadly disease is now accomplishable. The survival estimates are clubbed with deterministic actuarial model to formulate the insurance plan. The premium cost for the insurance plan for HIV/AIDS and ALL patients is estimated in classical and Bayesian framework. The premium cost in both the scenarios are realistic. The developed models can be improved further by incorporating better biomarkers. Mode loading can be included for better workable plans. It can also be extended for other diseases. In future, advancements in medical science may be able to improve the survival further and hence lowering the premium cost.

Contents