A Review of Survival Analysis Theory and Its Application

Yulu Liu(✉)

Binzhou Medical University, Yantai, China
3189849202@qq.com

Abstract. Marshall McLuhan, a famous Canadian communication scientist, has proposed a far-reaching view of media: media and information. Emphasizing that the media itself is the truly meaningful information, the basic driving force for social development, and the symbol to distinguish between different social forms. With the advent of the Internet era and the development of modern science and technology, people fragmented reading habits, short video as different from text, pictures, audio of a new media form, has become the audience, an important means, is gradually transformed into people's daily life, by the academic and industry, so myopic, frequency news as a new product under the background of the mobile Internet era, the research, has a strong practical significance and theoretical value. Most of the existing research on micro-documentaries focuses on the innovative exploration and narrative strategies in the era of financial media. This paper takes the fourth season of Momstant China as the analysis object, and adopts the sample analysis method and the literature research method to explore how micro-documentaries reflect on the form of the creation or the content of the audience, so as to help the creators achieve the best communication effect in the following works. It introduced the relevant concepts and meanings of the survival analysis method, and uses the hazard rate function, cumulative distribution function, survival function, and cumulative hazard rate function to describe the concept of survival time. Secondly, the application of survival analysis method in various industries at home and abroad is reviewed. Through the review of domestic and foreign applications, it is found that although the survival analysis method was first only applied in the medical field, it has been widely used by more and more scholars in other fields due to its advantages in processing the censored data.

Keywords: survival analysis · survival time · application situation in China · foreign application situation

1 Introduction

With the continuous improvement of research requirements, survival analysis has been gradually cited by various experts and scholars in the medical field, and spread to various fields. It is a basic idea of dealing with deleted data, which has attracted more and more attention in recent years. In the overview of mathematics development in 1986, the US National Scientific Council listed survival analysis as one of the six development directions, which shows its top position in the history of mathematics development.

M. F. b. S. M. Dom et al. (Eds.): CDSD 2022, ASSEHR 739, pp. 477–487, 2023.
https://doi.org/10.2991/978-2-38476-018-3_54

1.1 Main Questions

Based on the background of COVID-19, this paper starts from the innovative working mode of the online office in China, combined with the role of the traditional Chinese culture in the society, and analyzes the many factors and the overall influence of the influence mechanism that affects the employee happiness.

First, the study helps to reveal the positive and negative relationship between online office mode and employee happiness; second, the conclusion will improve the management of employees in the online office mode, and evaluate the development prospects of online office mode; again, the conclusion is based on the primary data obtained by the actual questionnaire, beneficial to make up for the lack of data support in previous literature. Survival analysis refers to a statistical analysis method of combining the endpoint event and the time experienced by this event.

It is the method to study the response time data of survival phenomena and phenomena and their rules. Different from other multifactorial studies, it considers the time of each observed object. The theoretical method is mainly based on the construction of a survival model (survival function, probability density function, etc.), to calculate the risk rate at each time point, that is, the survival time of the object studied, and the possibility of its state change, to analyze the influence of different factors on both, and to conduct factor analysis. In survival analysis, there are mainly three main types of survival models:

1.1.1 Parameter Method

First of all, the observed survival time t is required to meet a specific distribution, and the method of estimation of the parameters in the distribution is used to obtain the estimated survival rate value. The distribution of survival time may be an exponential distribution, an Weibull distribution, etc., and these scores are shown The th curves all have corresponding survival function forms. As long as the estimates of the corresponding parameters are obtained, the survival estimates and the corresponding survival curves can be obtained.

1.1.2 Non-parametric Method

In practical application, most of the survival time distribution does not conform to the distribution described in the above situation, so it is not suitable for analysis with the above method, so the non-parametric method should be used.

The test of such methods assumes the same assumptions as the non-parametric ones, assuming the same overall survival curves for two or more groups, regardless of the form of the overall distribution and parameters.

1.1.3 Semi-parametric Method

It does not limit the time and risk function, but only limits the relationship between the influencing factors and living conditions. This method belongs to the multivariate analysis method, which is mainly used to analyze the factors affecting the survival rate. The typical method is the COX proportional hazard model.

1.2 Literature Review

1.2.1 Survival Analysis of the Current Application Status Abroad

Due to its advantages in data processing, and practical application, survival analysis has been more and more concerned, including Robert J. Kauffman and Angsana A.Techatassanasoontorn research as a typical representative [1].

Robert J. Kauffman studied the impact of network externality on network adoption by using survival analysis [2]. The research shows that the earlier the banks that can form a more effective network scale and a higher level of externality tend to adopt ATM, and the size of the bank branch network affects the probability of its adoption. It shows that the external characteristics of the network will affect the adoption degree of the network.

With the integration of the Internet, e-commerce and wireless technologies, Robert J. Kauffman and Techatassanasoontorn AA believe that studying the diffusion of digital mobile devices needed for mobile commerce activities is an effective way to understand the rapid global scale expansion of mobile e-commerce [3]. They adopted the joint risk method in the study of international digital mobile communication diffusion proposed a new perspective, in the whole diffusion process, although known knowledge to dominate the influence factors, different countries is still incomplete, using joint risk model to test national characteristics, digital and analog mobile phone industry characteristics, digital mobile phone regulatory policy to explain diffusion, using nonparametric and parametric survival model analysis, the results of early to local diffusion has a broader impact than early diffusion.

Robert J. Kauffman and Bin Wang analyzed the business models of 130 Internet companies using non-parametric and semparametric Cox proportional hazards models of survival analysis to explore how business models affect the survival of companies [4]. The results show that smaller Internet companies greatly reduce the possibility of interaction between customers, suppliers and trading brokers, and relying on advertising as a major source of revenue.In addition, interactive platforms for individuals and businesses have failed to adversely affect them. With the development of society and the increasing maturity of the Internet, the size of large companies will decrease and the weak will leave the market. The study also pointed out that the business model is the main dimension affecting the survival and development of Internet companies.

In a 2009 study, Robert J. Kauffman and Techatassanasoontorn AA mentioned that although innovative diffusion theory believes that there are five types of adopters, this method lacks a theoretical basis and, more importantly, the key assumption of the normal distribution of the adopters needs to be actually verified [5]. The study used data from different purchaser types and factors influencing decisions on adopting digital wireless mobile phones in 46 developed and developing countries between 1992 and 2002. It uses a two-step analysis method.

The first step is the distribution of users was estimated using a diffusion model. The second step uses an iterative survival analysis model to examine factors influencing adoption behaviour, adding 1% more cumulative penetration as the target event when estimating the survival model. The results of the proposed diffusion model show that the digital wireless telephone diffusion model does not follow a normal distribution and does not completely map to the Rogers' five types of adopters.

An iterative survival analysis shows the results of 30% of the four types of buyer early adopters (innovators, early adopters, breakthrough adopters, and mainstream users). The results of the study can provide telecom operators with development strategies to attract different adopters, and it can also design an effective regulatory framework for policymakers. Angsana A. Techatassanasoontorn uses the method of survival analysis in the development research of digital wireless phone technology, and, with technology advantages and product life cycle theory, to build a model consisting of standards, market competition, technology cost and technology alternative interpretation diffusion and subsequent growth in market penetration [6].

Through data from 41 countries, the results show that the existence and influence of standards in the process of diffusion and market penetration growth plays wireless phones familiarity and analog mobile phone technology can also explain the spread, with non-price factors playing an important role in driving the growth of market penetration. The paper points out that the research can provide reference for the management policies, innovative policies and competition policies of digital wireless phones.

1.2.2 Survival Analysis of the Domestic Application Status

Yang Huan by using the survival analysis of half parametric method Cox proportional risk model for empirical analysis, study the construction enterprise mortality density dependence process and influencing factors, in 1989–2007 in Jiangsu province construction enterprises based on the research, the results show that the death of construction enterprises has a significant impact on the evolution of the industrial structure level [7]. The influence of population density on the mortality rate of construction enterprises is mainly manifested as increasing with the increase of the population density of enterprises, but this rising process is also restricted by the combination of organizational age. In addition, the size and external environment of the construction enterprise and environmental factors also affect the enterprise mortality rate.

Yang Huan analyzed the factors affecting the flow of rural workers, combining statistics from Jilin Province, using female workers' survival analysis, etc. [8]. Lai discussed the application of survival analysis in the loss of information customers, Studying the characteristics of information services [9]. The algorithm of survival time and customer loss definition conditions are studied, and using the user data of the National Science and Technology Books and Literature Center, the results show that the survival analysis method can well explain the living conditions of information customers and the factors affecting their survival, has a guiding role in customer information service management.

Tian Xinyuan collects data from bank credit card customers, The Kaplan-Meier method and Weibull distribution method were used to estimate the life cycle from commercial bank card to level of credit card customers [10]. The conclusion is that the card opening period for more than the general users is 30 months.

The combination of parametric analysis opens up new ideas for domestic credit card customers. It studies the influencing factors of enterprise ownership, industry type, scale, organization, structure, competition level, industry status and growth stage, and reveals the adoption trends of different types of enterprises [11]. It is also useful to extend the conclusions of typical information technologies to the adoption of other information technologies.

1.3 Research Methods and Approach of the Paper

Survival analysis is not only about describing the survival time distribution of the study subjects, but also about studying the "risk factors" and "protective factors" that affect the survival time.

2 Ease of Use

2.1 Review of the Theoretical Basis Studies of Survival Analysis

Survival theory is the frontier of using differential inclusion to study the state evolution of uncertain systems under a variety of constraints. It is a mathematical method to study the state evolution of uncertainty systems under various constraints.

2.2 Theoretical Basis and Tools of Survival Analysis

2.2.1 Survival Time

Survival time is also called survival time, failure time, etc., which refers to the time experienced from a certain point to the endpoint event of the observed object. Survival time is a broad concept, such as the time that credit card users open the card to the level. The description of the hazard rate function, cumulative distribution function, survival function, cumulative hazard rate function, etc.:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr[t \le T \le t + \Delta t]}{\Delta t} = F(t)/_{St} = -\frac{d \ln(s(t))}{dt} \tag{1}$$

2.2.2 Cumulative Distribution Function

The cumulative distribution function is expressed by F (t), referring to the probability of the survival time being less than or equal to:

$$\mathrm{F}(t) = \mathrm{P}[T \le t] = \int_0^t f(s)ds \tag{2}$$

survival function. Survival function is generally expressed as S (t), which refers to the probability of survival time T being greater than t, corresponding to the cumulative distribution function, and is expressed by the formula:

$$\mathrm{S}(t) = \mathrm{P}[T \ge t] = 1 - F(t) \tag{3}$$

2.2.3 Cumulative Hazard Rate Function

The cumulative hazard rate function is an integral of the hazard rate function that means the cutoff to time t, the probability of a state change. It is expressed as follows:

$$\Lambda(t) = \int_0^t \lambda(s)ds = -\ln S(t) \tag{4}$$

2.3 Main Indicators of the Theoretical Basis of Survival Analysis

The so-called deletion of survival data means that we do not see the state of the individual change during the observation period, in other words, we cannot determine the specific survival time of the individual. The deletion can be divided into right deletion, left deletion and interval deletion:

1) right deletion: for an individual, if only the survival time X C is known at the observation point, then the individual is deleted right at C. In practical problems, most of the deleted data is right deleted.
2) left deletion: refers to the individual survival time X C, then the individual left deletion at C. If the studied individual has a left deletion, it usually also exists, which is called the double deletion of survival time.
3) interval deletion: the survival time of an individual can only be determined in a certain interval (L, R], L represents the left observation moment, R indicates the right observation moment, and can only determine the transition in the ontogenesis state within the interval, but the transition moment is unknown.

In actual research, there is often a large number of deleted data deleted, if the traditional research methods are used, it may lead to a large deviation and affect the scientific nature of its conclusions.

Survival analysis method solves the situation of data deletion, making the conclusion more scientific, rigorous and valuable, which is also one of the main reasons why survival analysis method is widely used in research in various fields (Fig. 1).

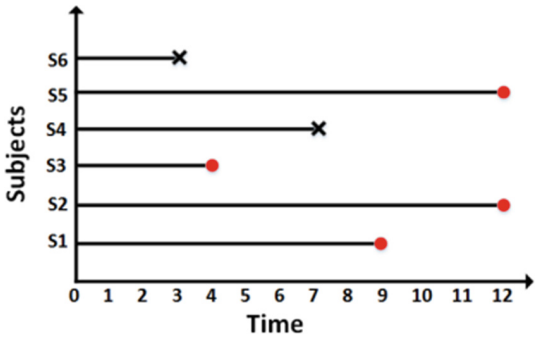

Fig. 1. An illustration demonstrating the survival analysis problem. *(Self-drawn)*

3 Survival Analysis Theory and Methods

3.1 Theoretical Model of Survival Analysis

Among all functions, the survival function or its graphical presentation is the most widely used one. This method is the most widely used one for estimating survival function. Let T1 < T2 < ··· < TK be a set of distinct ordered event times observed for N(K ≤ N)instances.

$$p(T_j) = \frac{r_j - d_j}{r_j} \tag{5}$$

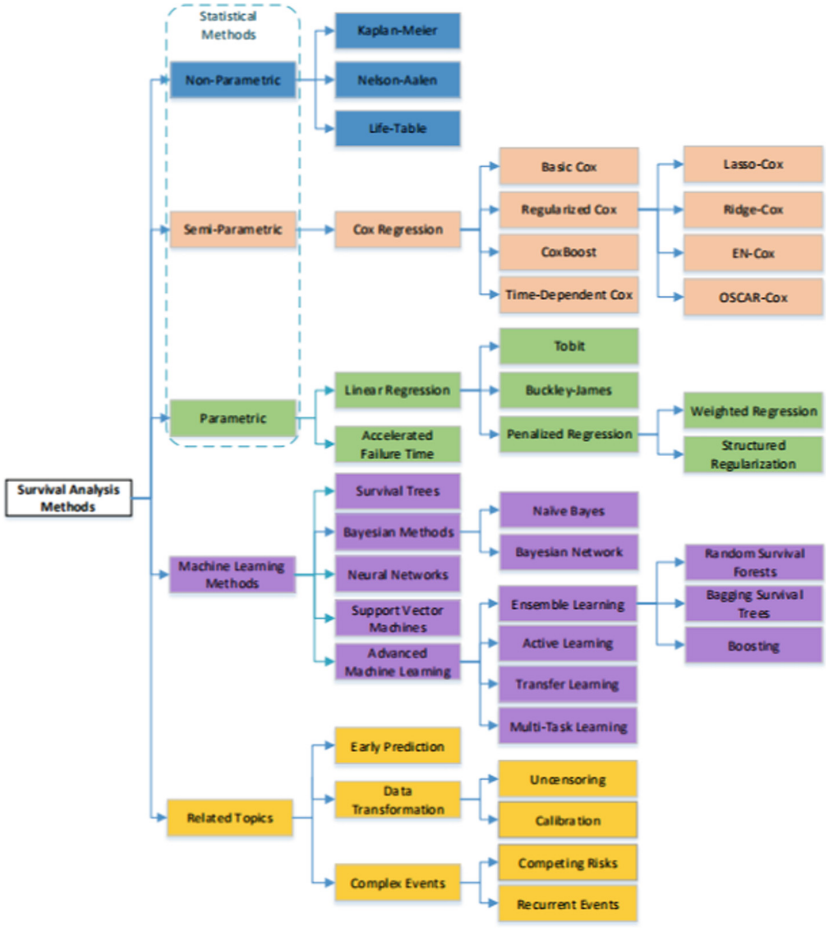

Fig. 2. Taxonomy of the methods developed for survival analysis. Table Type Styles. *(Self-drawn)*

3.2 Survival Analysis and Algorithms

Figure 2 shows both the advantages and disadvantages of each type of methods based on theoretical and experimental analysis and lists the specific methods in each type.

4 Application of the Theory of Survival Analysis

4.1 Definition of the Survival Analysis

Survival analysis refers to a statistical analysis method of combining the endpoint event and the time experienced by this event. It is the method to study the response time data of survival phenomena and phenomena and their rules. Different from other multifactorial studies, it considers the time of each observed object.

The theoretical method is mainly based on the construction of a survival model (survival function, probability density function, etc.), to calculate the risk rate at each time point, that is, the survival time of the object studied, and the possibility of its state change, to analyze the influence of different factors on both, and to conduct factor analysis.

4.1.1 Definition: Survival Time

Survival time also called survival time, failure time, etc., refers to the time experienced from a certain point to the endpoint event of the observed object. Survival time is a broad concept, such as the time that credit card users open the card to the level. Survival time, the common hazard rate function, cumulative risk rate function, cumulative distribution function, cumulative risk rate function and so on description:

4.1.2 Hazard Rate Function

The probability of an individual having a state transition at time t is a hazard rate function;

4.1.3 Cumulative Distribution Function

The cumulative distribution function is expressed by F(t), referring to the probability of the survival time.

4.1.4 Survival Function

The survival function is generally expressed as S(t), which refers to the probability that the survival time T is greater than t, corresponding to the cumulative distribution function;

4.1.5 The Cumulative Hazard Rate Function

The cumulative hazard rate function is an integral of the hazard rate function that means the cutoff to time t, the probability of a state change.

4.2 The Application Value of the Survival Analysis Theory

Survival theory employs differential inclusion to study the state evolution of uncertain systems under constraints, bringing new vitality to the study of uncertain systems.

5 Application of Survival Analysis Theory in China

5.1 Historical Origin of the Survival Theory

Cao Yu et al. (2010;2011;2012) based on the data of newly established enterprises in Hunan Province from 1980 to 2007, used the life table method, Cox proportional risk model and accelerated failure model to investigate the quantile of enterprise survival time, average remaining life and average survival time, and studied the living conditions and influencing factors of different types of enterprises from various angles.

5.2 The Development Course and Characteristics of Survival Theory

5.2.1 Development History

Survival analysis takes the “Time——event (Time-to-Event)"as the research object, the author intends to summarize the research progress of the survival analysis method in the field of economic management, especially in China, and finally explore the expansion research of this method.

5.2.2 Development Characteristics

In 1992, Narain innovatively applied the relevant theory of survival analysis to the credit evaluation system by constructing the accelerated failure model of personal loans. Survival analysis reduces the second place more effectively when the proportion of failure events in the sample is relatively large.

5.3 The Enlightenment of Survival Analysis Theory to China

Survival analysis gradually developed based on population vital statistics has been widely used in medicine, biology, demography and other fields. Although the application of maturity is relatively low, application category is also limited, but still made certain achievements, concentrated in the following aspects: given the survival analysis ability to accurately estimate the future survival time, the enterprise skillfully apply this method in financial risk prediction, strengthen prior preventive measures to minimize enterprise loss, prevent in the bud. Taking financial difficulties as a special event, “survival time” is the time when the company is founded to financial difficulties.

6 Conclusion

Previous empirical studies require accurate data, and the emergence of survival analysis method just makes up for the lack of this aspect. With its advantages in processing deleted data, it is quickly and widely applied to various fields except medicine. Through the inductive analysis of the research status of survival analysis at home and abroad, we can publish it. Now, the domestic survival analysis method still starts later than abroad, but the spread speed is very fast.

Most of them focus on COX proportional risk model, which is less involved in other studies. Robert J. The external factors mentioned in the study of the Kauffman in 2000 have a positive correlation on the research subjects, while domestic scholars have few external factors when selecting the influencing factors, which is a blank place in domestic research.

Acknowledgment. This paper is my personal research work and research results under the guidance of my supervisor. The paper does not include the research results published or written by other people or other institutions, except where specially marked and thanked. Other comrades have made clear statements and expressed their gratitude for their inspiration and contributions to the study.

References

1. Jeffrey R Marks, David P Winchester, and David G Bostwick.1997. Artificial neural networks improve the accuracy of cancer survival prediction. Cancer 79, 4(1997), 857–862.
2. Ching-Fan Chung, Peter Schmidt, and Ana D Witte.1991. Survival analysis: A survey. Journal of Quantitative Criminology 7,1(1991),59–98.
3. A.Ciampi, R.S.Bush, M.Gospodarowicz, and J.E.Till.1981.An approach to classifying prognostic factors related to survival experience for non-Hodgkin's lymphoma patients: Based on a series of 982 patients:1967–1975.Cancer 47,3(1981),621–627.
4. A.Ciampi, C-H Chang, S.Hogg, and S.McKinney.1987.Recursive partition: a versatile method for exploratory-data analysis in biostatistics. In Biostatistics. Springer, 23–50.
5. Antonio Ciampi, Johanne Thiffault, Jean-Pierre Nakache, and Bernard Asselain. 1986. Stratification by stepwise regression, correspondence analysis and recursive partition: a comparison of three methods of analysis for survival data with covariates. Computational statistics & data analysis 4,3(1986),185–204.
6. Joseph A Cruz and David S Wishart. 2006. Applications of machine learning in cancer prediction and prognosis. Cancer informatics 2(2006).
7. Sidney J Cutler and Fred Ederer.1958.Maximum utilization of the life table method in analyzing survival. Journal of chronic diseases 8,6(1958),699–712.
8. Cox R David.1972.Regression models and life tables. Journal of the Royal Statistical Society 34,2(1972),187–220.
9. Cox R David.1975.Partial likelihood. Biometrika 62,2(1975),269–276.
10. Robert J. Kauffman, James McAndrews, Yu -Ming Wang. Opening the "Black Box" of Network Externalities in Network Adoption[J]. Information Systems Research,2000,3;61–82.
11. Robert J. Kauffman a, Angsana A. Techatassanasoontorn. International Diffusion of Digital Mobile Technology: A Coupled -Hazard State -Based Approach [J]. Information Technology and Management,2005,6;253–292.