IJSRM
INTERNATIONAL JOURNAL OF SCIENCE AND RESEARCH METHODOLOGY
An Official Publication of Human Journals

Human Journals
Review Article
August 2021 Vol.:19, Issue:2

Analytical Models of Survival Analysis: Concepts and Their Applications

IJSRM
INTERNATIONAL JOURNAL OF SCIENCE AND RESEARCH METHODOLOGY
An Official Publication of Human Journals

Ajay S. Singh*, Sotja G. Dlamini

Department of AEM, University of Eswatini, Eswatini (Swaziland).

Submitted: 23 July 2021
Accepted: 31 July 2021
Published: 30 August 2021

HUMAN JOURNALS

www.ijsrm.humanjournals.com

Keywords: Survival analysis, Cox model, Duration, Kaplan-Meier, Log-rank, Life-table

ABSTRACT

The main objective of this article is to describe the methods and applications of survival analysis. Survival analysis is a subdivision of statistics and is also known as duration modeling in economics. It is an analysis of the anticipated duration of time till one event, such as death, happens in biological organisms. It is also known as reliability analysis in the engineering field, for example, failure in mechanical systems. Generally, survival analysis involves the modeling of time to event data. The main objective of survival analysis or time to event analysis is to estimate the time for an individual or a group of individuals to experience an event (Time estimate) and survival probability is calculated as the number of subjects surviving divided by the number of individuals at risk for each time interval. Common methods of survival analysis are life-table, Kaplan- Meier, log-rank, and Cox model. In this article, we consider the concept and important applications of analytical models of survival analysis for health and economic research. This article also includes analytical principles and methods for estimation. It can be useful for researchers and professionals for better applications of survival analysis.

INTRODUCTION:

The term was invented from a medical background in which it has been used to estimate the survival rate of patients. Survival analysis is a group of statistical procedures for data analysis, for which the outcome variable of interest is time until an event occurs. It is the study of time between entry into observation and a subsequent event. The term survival analysis was derived in the initial researches, where the event of interest was death (no more). Now the scope of the survival analysis has become wide and more popular.

In another word, survival analysis is an analytical technique focused on time-to-event data. Frequently the event is overall survival, but some other events can also be considered, such as progression of the disease, or occurrence of disease. The attraction of survival analysis for economic evaluation is that economic endpoints based on survival curves and the health outcomes are considered longitudinally over time, and not cross-sectionally at a specific point in time.

Survival analysis is also known as reliability analysis in the engineering discipline and event history analysis in sociology. The survival analysis can be easily used for socio-economic research to investigate complex phenomena such as job status, inflation, supply and demand for bank credits, the life expectancy of the product, the producer and consumer, etc. Survival analysis is also adapted in conventional econometric modeling.[1-2].

Advantages of survival analysis are classified as follows:

- Descriptive, for the sample of subjects observed
- Predictive for the representative population of subjects
- As a comparative tool being objective and presenting with high accuracy

Concept of Censoring:

In survival, investigations consider a key, and a simple analytical problem is known as censoring. It happens when some information about individual survival time, but we do not know the exact survival time.

There are three main reasons for censoring mentioned below:

- When an individual has not seen the event before the study ends.
- When an individual is absent for follow-up during the study period.
- When an individual is out from the study due to death or some other causes like as adverse effects of the medicine.

Censoring is of two types, right and left. Generally, the researchers encounter right-censored data. Left-censored data can happen when a person's survival time becomes incomplete on the left side of the follow-up period for the person. For example, follow up a patient for any infectious disease from the time of his being tested positive for the infection. In another word, never know the actual time of exposure [3].

REVIEW AND LITERATURE:

Survival analysis is an approach to answer specific questions, such as what is the ratio of a population which will continue or survive earlier than a definite time? What are the rates they will expire? How do specific circumstances or characteristics increase or decrease the chance of survival? To answer these types of questions, it is very important to define "lifetime". In the case of biological survival, death is unambiguous, in some biological problems; events may have the same ambiguity. Survival analysis is used to estimate the lifespan of a specific population under study.

Survival analysis in Economics:

The survival analysis can be used for socio-economic research to investigate complex phenomena such as unemployment, employment, inflation, supply and demand for bank loans, the life expectancy of the products, the producer and consumer, etc.

Survival analysis in health care:

Survival analysis is concerned with the time elapsed from a known origin to either an event or a censoring point. It may deal with survival, such as the time from diagnosis of a disease to death, but can refer to any time-dependent phenomenon, such as time in hospital or time until a disease recurs. Major applications of survival analysis in public health include the following: (a) estimation of survival distributions, (b) testing hypotheses of equal survival distributions, and (c) identification of risk or prognostic factors.

Survival analysis in clinical trials:

Many clinical trials involve following patients for a long time. The primary event of interest in those studies is death, relapse, adverse drug reaction, or development of a new disease. The follow-up time for the study may range from few weeks to many years. A different set of statistical procedures are employed to analyze the data, which involves time to the event and analysis. It is a very useful tool in clinical research and provides invaluable information about an intervention. This article introduces the researcher to the different tools of survival analysis.

Survival Analysis in Health and Economics:

In the beginning, the survival analysis was used to study death as an event specific to medical studies [4-5] and demographical studies [6-8], as from the '70s these statistical techniques have been increasingly used in economics and social sciences. Survival data requires a different statistical analysis compared to quantitative data due to their particularities.

Kachman (1999), explained the applications of SA and derive a probability model based on Weibull distribution that animals survive at least until a specific time. The parametric approach is used for estimation because it is more suitable for large complex models [9].

Babucea and Danacica (2007) used survival analysis in the analysis of socio-economic phenomena and concentrated on the Cox regression, and based on the concept of hazard, baseline hazard, hazard rate, hazard rate interpretation. The application of the survival analysis in unemployment is presented [10].

In the area of labor, economics applied SA, for example, durations of employment and treated as survival times and analyzed accordingly Heckman and Singer,1985; Kiefer, 1988; Lancaster, 1990 [11-14]. Recently, survival analysis approaches have been proposed for analyzing medical costs. In the survival analysis approach to cost data, individuals' cumulative costs are treated like survival times and analyzed accordingly Dudley et al., 1983; Fenn et al., 1995, 1996 [15-16].

Ruth et al. (1999) analyzed and explained SA techniques for the estimation of medical care cost [17].

Survival Analysis in Business and Marketing:

Adrian and Kuldeep (2008) derived the statistical business failure prediction models attempt to predict the failure or success of a business. In this study, survival analysis has been used for business failure prediction. Overall, the results suggest that survival analysis techniques provide more information that can be used to further the understanding of the business failure process [18].

Tim, et al. (2001) mentioned customer lifetime analysis uses customer information. like the typical time between purchases to predict the potential value of a customer and conclude that survival analysis is a good, sound, and flexible tool to analyze the length of time-to-next-purchase, cancellation and also output of survival analysis can be fed into various forms of customer lifetime analysis to provide more reliable conclusions and a better understanding of customer lifecycles and value for more efficient use of the available budgets [19].

According to Michael, (2009), the simplest survival-based model captures effects of a changing subscriber mix that might be hard to capture with a traditional, time-series-based approach. Michael, (2009) derived a long-range daily forecast for a subscriber population that is a constantly changing mix of customers segments, each with its survival function [20].

Rashid (2019), proposed a solution that integrates various techniques of customer data analysis, modeling, and mining multiple concept-level associations to form an intuitive and novel approach to gauging customer loyalty and predicting their likelihood of defection [21].

Goetz (1992) derived a selectivity model of household food marketing behavior in Sub-Saharan Africa[22].

Survival Analysis in agriculture:

In agricultural research, however, survival analysis has been overlooked, maybe due to the availability of other analysis techniques. Scherm and Ojiambo (2004) reviewed survival analysis and its application in plant pathology and further illustrated the methodology in longitudinal data on the timing of defoliation of blueberry leaves[23]. Hay, et al. (2014) derived modeling seed germination in response to continuous variables and probit analysis and alternative approaches[24]. Isabel and Cathrine (2016) used Survival Analysis of Contract Farming Participation in Northern Ghana and calculated hazard rates for entry and exit in contract farming schemes [25]. Załuski et al. (2018), analyzed the survival of willow plants in a long-term experiment[26]. Recently, Onofri et al. (2019) reviewed analyzing censored data in agricultural research with available software[27]. Rathod, et al. (2020) discussed the applications of survival analysis in agricultural research and illustrated them with clinical data[28]. Benedikt, et al. (2019) used survival analysis for the adjustment phase following investment in Swiss Dairy Sheds[29].

ANALYTICAL CONCEPT OF SURVIVAL MODEL:

The process of survival analytics is different when compared to predictive and descriptive analytics. In this case, the time component is an important factor which efficiently determines the success or failure of a model. Cox (2001), classified survival analytics in four groups [30]. Different kinds of functions are easy-going with different models based on the metric to be used with time as a component for the event or cases to occur. The main target is to determine the right model to be selected for the observed survival analytic data. In parametric model analysis, the survival curve depends only on the shape of the model with its function value. The shape of the model can be estimated concerning the characteristics of a nonparametric model. As an outcome, the shape of the hazard function also varies concerning time.

Important Terminology:

Survival Time: Survival time refers to a variable that measures the time from a starting time (e.g., time initiated the treatment) to an endpoint of interest (e.g., attaining certain functional abilities).

Survival Rate: Survival rate is defined as the percent of observations/people who survive a disease such as cancer for a specified amount of time.

Time-to-event: It is the time until the event occurs. Time-to-event is a positive random variable.

Hazard: The instantaneous rate at which a randomly selected individual known to be alive at the time (t - 1) will die at time t is called the conditional failure rate instantaneous hazard.

Survival function:

The survival function is fundamental to the study of survival analysis. Survival function, S (t) indicates the chance that a person survives longer than some specified time ***t***. It gives the chance that the random variable **T** exceeds the specified time **t**. The survivor function is frequently expressed as a Kaplan-Meier curve and a vertical drop in a curve indicates an event.

Probabilistic Approach:

According to Bayes theorem, the conditional probability of two joint events A and B is defined as;

$$P(A \mid B) = \{P(B \mid A) / P(B)\} * P(A)$$

Or equivalent to

$$P(A) = \{P(A \mid B) / P(B \mid A)\} * P(B)$$

In this approach, define A as surviving through time t, and B as surviving through time t-1. Then P(A│B) is the probability of a participant surviving through time t given that individual participant has survived through all preceding times (t-1), (t-2),………, (1). P(B│A) is the probability of surviving through t-1 given that the individual survived through t. Probability p=1, therefore to calculate the conditional probability of surviving through time t and also two information are required. First, the probability of surviving is through time t given that the individual participant survived the previous time and second is the probability of surviving the previous interval.

Hazard function:

The hazard function λ (t, X_i) gives the instant potential per unit time for the event to happen, given the individual (Xi) has survived/reached up to time *t*. It is the probability of failure in a small period between t and t + Δt given that the subject has survived up till time t.

The probability model of any individual in cohort X_i survives at least until time t. Its risk function is defined as

$$F(t) = P(T_i \leq t) = \int_0^t F(u)du$$

$$S(t) = 1 - F(t) = P(T_i > t) = \int_0^\infty f(u)du$$

For an individual X_i

$$S(t, X_i) = P(T_i \geq t) = 1 - F(t, X_i) \quad \text{Where } T_i \text{ is time of failure.}$$

Risk of failure of an individual at t

$$\lambda(t, X_i) = \lim_{\Delta t \to 0} \Pr \left\{ \frac{(Ti \leq t + \Delta t \mid Ti \geq t)}{\Delta t} \right\}$$

$$= f(t, X_i)/S(t, X_i)$$

$$S(t, X_i) = P(T_i \geq t) = 1 - F(t, X_i)$$

Where T_i = Time of failure. $F(t, X_i)$ is Cumulative distribution function for T_i and $f(t, X_i)$ is density function for T_i.

In this situation, the Hazard is the measure of risk: The greater the hazard between time t1 and t2, the higher the risk of failure in this time interval T_i.

Importance of Hazard function:

Hazard function provides an insight into the conditional failure rates. The hazard function is also used to identify a specific model form by using mathematical modeling of survival data.

Hazard ratio(HR):

The hazard ratio is an estimate of the ratio of the hazard rate in the treated versus the control group. Hazard ratio (HR) is similar to relative risk. HR is described as the outcome of trials where the problem is, to what level can treatment reduce the duration of an illness [31].

Cox Proportional Hazards Model:

The Cox proportional hazards model has been the most widely used in clinical trials because of its applicability to a wide variety of clinical studies [32]. The Cox model was introduced in 1972 by Cox. This model is useful for the analysis of survival data with and without censoring, for identifying differences in survival due to treatment and prognostic factors in clinical trials.

The Cox model is a regression method for survival data. It provides an estimate of the hazard ratio and its confidence interval. Cox regression is considered a ‘semi-parametric procedure because the baseline hazard function, $h_0(t)$, does not have to be specified.

Assumptions of Cox proportional hazard model:

The hazard ratios of two individuals are independent of time and are valid only for time-independent covariates. This means that the hazard functions for any two individuals at any point in time are proportional.

There are two types of survival function analysis. The first important one is Parametric and the second is Nonparametric analysis.

Life Table Analysis:

According to Booth and Hirschl (2005), in longitudinal studies, it is of the main interest to estimate a ‘survival’ curve for the population. What proportion of the population survives beyond a specified time interval without a particular event happening?[32] The life table is a simple technique to describe the survival in the sample. The life table technique is one of the oldest methods for analyzing survival data. The distribution of survival times is divided into a certain number of intervals. For each interval we can then compute the number and proportion of cases that entered the respective interval ‘alive,’ the number and proportion of cases that failed in the respective interval, and the number of cases that were lost or censored in the respective interval. Based on those numbers and proportions, several additional statistics can be calculated, such as the number of individuals at risk, proportion failing, proportion surviving, survival function, hazard rate, and median survival time. The life table technique is generally used for larger samples where the time intervals are large enough to be divided into smaller units.

Survival rates are used extensively in demographic projection techniques. Survival rates are derived from life tables or census data and are used to calculate the number of people that will be alive in the future. Survival rates are used to calculate the number of people that will be alive at a future date in time. Initially, survival rates are derived from life tables. Life table provides the idea to calculate rates for ages birth to 85 plus. If survival rates or life tables are not available, the rates may be computed from a model life table or census data.

Life tables are used to measure mortality, survivorship, and the life expectancy of a population at changing ages.

There are several types of life tables. A generation or cohort life table is a life history of the mortality experiences of an actual cohort of individuals. The cohort begins at birth and their mortality experiences are recorded through the death of the last member of that cohort.

Life Table:

A life table is a statistical tool that summarizes the mortality experience of a population and yields information about longevity and life expectation. Although it is generally used for studying mortality, the life table format can be used to summarize any duration variable, such as duration of the marriage, duration of contraceptive use, etc.

Construction of Life Table:

There are several methods available to construct an ordinary life table using data on age-specific death rates. Namboodiri and Suchindran (1987) discussed the most common methods for the construction of a life table[33].

Current or period life tables:

Period life tables are based on the mortality experience of a hypothetical cohort of new-borns babies, usually, 100,000 new-borns, who are subject to the age-specific mortality rates on which the table is based. It traces the cohort of new-borns babies throughout their lifetime under the assumption that they are subject to the age-specific mortality rates of a region or country.

There are two types of current life tables:

- Unabridged, for single years of life
- Abridged, for 5-year cohorts of life

Fergany (1971) also suggested the construction of an ordinary life table based on age-specific death rates [34]. In this method, the age-specific death rate will be converted into the proportion of deaths in the age interval by using a simple formula.

Life Table Column Notation and interpretation

A simple ordinary life table contains several columns with a unique interpretation (Table-1).

Table-1

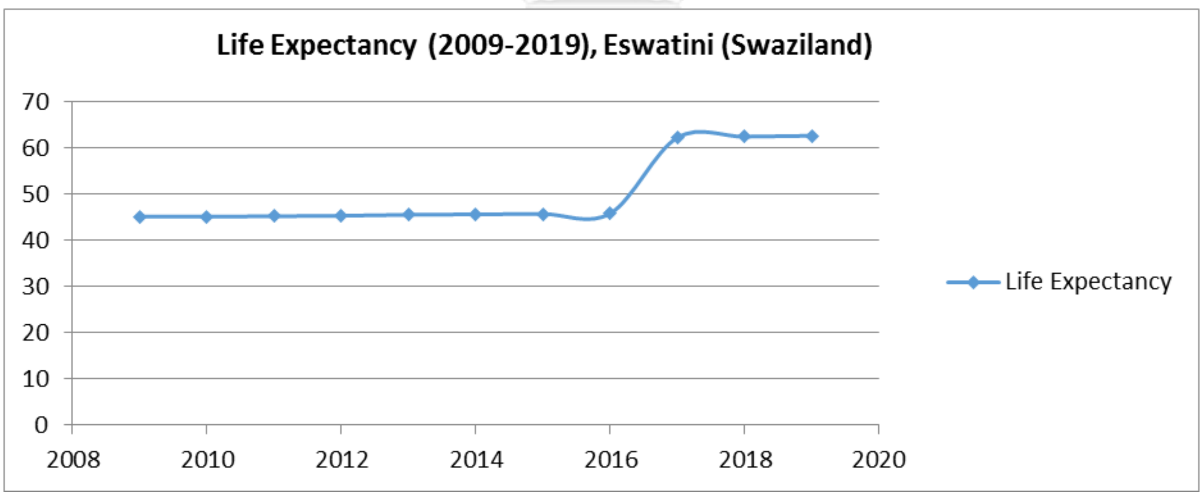
Column	Notation	Interpretation
1	$(x, x+n)$	Age interval. The period between two exact ages mentioned in years
2	${}_nd_x$	The number of persons in the cohort who die in the age interval $(x, x+n)$ *or* The number of persons dying during the age interval
3	${}_nq_x$	The proportion of persons alive/surviving at the beginning of the age interval who die during the age interval. The ${}_nq_x$ column has a probabilistic interpretation; the probability that a person of age x will die in the age interval (x, x + n).
4	l_x	The number of persons surviving to the beginning of the age interval or the number living at the beginning of the age interval.
5	${}_nL_x$	The number of years of life lived by the cohort within the given age interval $(x, x+n)$.
6	T_x	It is the cumulative sum of the ${}_n\mathrm{L}_\mathrm{x}$ values. The total number of person-years that would be lived for a particular age cohort if the cohort were to progress through the remainder of the life table.
7	e_x^0	The average number of years of life remaining for a person alive at the beginning of age interval x

Life Expectancy in Eswatini (Swaziland), SADC Report, 2020:

Worldwide information on life expectancy does appear to be strongly correlated with economic development and employment. Improvements in economic conditions are an important force behind mortality decline. Life expectancy is also one of the good indicators of any developing country. It is indirectly indicating good health infrastructure. SADC Report, September 2020 (Table-2) indicated the life expectancy in the Eswatini [35].Table-2shows a significant increase in the last five years, and only on this indicator, point towards significant improvement in economic and health conditions.

Table-2

Year	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
Life Expectancy	45.1	45.1	45.2	45.3	45.5	45.6
Year	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
Life Expectancy	45.7	45.8	62.2	62.4	62.6	-



Kaplan–Meier Survival function [4]:

The Kaplan-Meier test is a statistical method. Kaplan-Meier method is a non-parametric representation of survival throughout the collected data and it is also allowing for incomplete records. Generally, it is used to analyze death as an outcome, in applied statistics/demography but in recent years these techniques have also gained popularity in the social sciences or industrial statistics or an engineer might measure the time until failure of machine parts) and also economists might measure the length of time people remain unemployed after a job loss.

Kaplan-Meier estimate of the survival function is a series of horizontal steps of declining magnitude which, when a large enough sample is taken, approaches the true survival function for that population. The value of the survival function between successive separate sampled observations is assumed to be constant. The application of KM estimator can be used by applying statistical tests such as hypothesis testing, Wilcoxon test, and likelihood ratio test.

This test especially determines the patient's survival time between two groups. For clinical drug examination, a successful test indicates that the group of people taking the new drug has a shorter time to improvement or death than the group of people taking a place. It also really works well for effective cancer treatments and follow-up.

An important advantage of the Kaplan-Meier curve is that the method can take into account "censored" data. When no truncation or censoring occurs, the Kaplan-Meier curve is equivalent to the empirical distribution. Kaplan Meier's method presupposes a greater reduction in calculus volume than the actuarial method because survival is estimated whenever the pre-established event for a subject occurs.

The main difference is the time intervals, i.e., with the actuarial lifetable approach is consider equally spaced intervals, while the Kaplan-Meier approach uses observed event times and censoring times. The calculations of the survival probabilities are detailed in the first few rows of the table.

Steps of Kaplan-Meier method:

- Listing the time when the pre-established event occurs, since subject's involvement in the survey or participation time.
- Finding for every participation time the number of subjects that continue to participate in the survey – those who did not achieve the pre-established event.
- establishing the number of subjects who achieved the pre-established event within a time interval.

- The calculus of the probability of occurrence of the pre-established event, for each
- participation interval.
- The calculus of survival probabilities for x duration is;$p_x=1-qx$, and the cumulated survival probability is $Px=p_x*(p_{x-1})*(p_{x-2})*$ ....

The Kaplan-Meier technique is the only useful method for initial evaluation since it is purely a descriptive method for the evaluation of one variable. The survival curve of this method is scalar form because the proportion of subjects who have the chance to continue observation without the occurrence of the pre-established event changes exactly at the moments when the pre-established event is achieved. The survival level is 100% from the curve origin until the moment of the first occurrence of the event (employment in our case), where it drops to the newly calculated value, which constitutes a new level during which survival is constant until the next event achieved. Therefore, every step corresponds to the occurrence of one or several pre-established events.

Hypothetical study using Kaplan-Meier analytics:

Consider there are about 500 records of patients which have been followed over 5 years (Table-3).

Table-3

Time Interval	Cohort or On Risk	Unavailable for follow up	No. of Died	Survived	KM Survival Prob.
Year 1	500	15	25	475	475/500= 0.95
Year 2	460	15	48	412	0.95 * (412/460)= 0.85
Year 3	397	15	75	322	0.95* 0.85 * (322/397)= 0.66
Year 4	307	15	99	208	0.95*0.85*0.66*(208/307)=0.36
Year 5	223	15	111	112	0.95*0.85*0.66*0.36*(112/223)= 0.10

Cox Regression [4]:

Cox regression can be used to determine whether a characteristic of subjects affecting the survival and, if so, how much and in what direction (to increase or decrease). Survival prediction can be difficult if not taking into account all factors that influence it. It is therefore necessary to identify those variables that affect survival and that can be used in the calculation of a predictive indicator of survival. A method to determine such an indicator, and associated survival curve, is called Cox proportional-hazard regression. Cox proportional-hazard model is a semi-parametric method that enables the determination of the effect of different variables on the hazard. Assuming that we have "*n*" units' individual under observation, then the model has the form.

Model:

$$\lambda_i (t, X_i) = e^{\beta X_i} .\lambda_0 (t) \; ; i=1, 2, \ldots\ldots\ldots\ldots n$$

$\lambda_i(t, X_i)$ is the hazard calculated for each individual and λ_0 (t) is the baseline hazard.

Where $X_i = (X_1, X_2, \ldots\ldots\ldots\ldots\ldots X_n)$ is the vector variable factor

and $\beta = (\beta_1, \beta_2, \ldots\ldots\ldots\ldots \beta_n)$ is vector regression coefficient.

The coefficients in a Cox regression relate to hazard; a positive coefficient indicates a worse prognosis and a negative coefficient indicates a protective effect of the variable with which it is related.

Logistic Regression:

Logistic regression is applied when the investigators examine the relationship between risk factors and various diseases or events. The ability to consider the time element of event occurrences by proportional hazards models has meant that logistic regression has played a less important role in the analysis of survival data [4][36]. The Cox model is preferred over the logistic model, which ignores survival time and censoring data[4][37]. Given a Cox model and the coefficients are subsequently estimate as to the baseline hazard function and the survival curves.

Log Rank Test:

The test is more likely to identify the difference between groups when the risk of an event is consistently more for one group than another. It is unlikely to investigate a difference when survival curves cross. It calculates a test statistic for testing a null hypothesis that the survival curves are the same for all groups. The number of expected is calculated as the proportion of subjects who are at risk at a given time point multiplied by the total number of events at that point**.** For each time point the observed number of deaths in each group and the number expected if there has been no difference, are calculated.

The log-rank test is also known as the Mantel Haenszel log-rank test. The Cox Mantel log-rank test is a form of Chi-square test [38].

The log-rank test is based on the same assumptions as of the hazard ratio that the survival chances are the same for individuals early and late in the study, and the events occur at the specified time.

It is useful to draw survival curves when analyzing survival observations. Under the null hypothesis, the log-rank statistic is approximately chi-square with one degree of freedom, and the *p*-value for the log-rank test is determined from the chi-square distribution table.

There are other tests for survival data. One of the important ones is the 'Peto test'. It is an alternative to the log-rank test. In contrast to the log-rank test, the Pito test uses a weighted average of the observed minus expected score. It places more emphasis on the information at the beginning of the survival curve where the number at risk is large.

Parametric representations of survival defined using some statistical distributions, such as Weibull, Gompertz, or Exponential, allow for survival to be extrapolated beyond measured and important for economic modeling.

Exponential function:

The exponential distribution is also known as the negative exponential distribution. The exponential distribution is defined as a process in which events occur independently and continuously at a constant average time.

Now density function $f(t) = \lambda e^{-\lambda t}$

and Survival function $S(t) = e^{-\lambda t}$

Where **the** risk of failure (λ) and constant over a while.

Gompertz distribution:

It is also continuous probability distribution. The Gompertz distribution is frequently applied to describe the distribution of adult lifespans by a statistician. Gompertz distribution is also considered in biological sciences for the analysis of survival. It has been used as an individual-level simulation for customer lifetime value modeling in business management. Recently, computer scientists have also started to model the failure rates of computer code by the Gompertz distribution and network theory, the walk length of a random self-avoiding walk (SAW) is distributed according to the Gompertz distribution.

Weibull distribution:

Weibull distribution is a continuous probability distribution. The Weibull distribution is the maximum entropy distribution for a non-negative real random variate with a fixed expected value. Weibull Model has the following Hazard and Survival functions.

$$\lambda(t) = \rho\,\lambda\,(\lambda\,t)^{\rho-1}$$

and

$$S(t) = e^{-[(\lambda\,t)]\rho}$$

This model is for both increasing and decreasing Hazardous and if $\rho = 1$, the Weibull distribution converted into exponential distribution.

In a simple situation, the marginal likelihood of survival time can be obtained by integrating over random effects. The basic parametric approach involves getting joint likelihood of survival times and the random effects. The parametric approach is more suitable for estimation in the large complex model.

SUMMARY:

Survival analysis is a very good tool when a researcher takes into account the time till an event occurs and the censored data. There are some common errors performed by researchers when applying tools of survival analysis for their research [39]. The first being, only data related to an event of interest occurring is reported. The *time* of the event is not mentioned. How long patients were observed with no events occurring is not considered. Events would be observed more frequently in patients with longer follow-up times than in patients with a short follow-up. Evaluation of raw event frequencies without mention of time will produce biased results. Similarly, when we get biased results, no distinction is made as to whether a patient suffered an event or was censored. The third error is not including the censored data in the analysis. If we take a specific proportion of events from both the groups, without taking into account the censoring, a different method of statistics should be employed, and not the survival analysis technique.

CONCLUSION:

Survival analysis can be used to determine not only the survival or probability of failure of apparatus based on the life or hours of actions but also to differentiate the difference between operating/actions environments. There are three main objectives of survival analysis, to estimate and interpret survival from the observed survival data; to compare survival functions, and to assess the relationship of survival timed and explanatory variables. Survival analysis provides a great tool for evaluating the time to an event type of data, which is very common in health and economic investigations. Researchers are not using it frequently because they are not confident in the theory of its application and its interpretation. There are some books and research available that provide basic knowledge on survival analysis. This article may help in the understanding and applications of analytical models of survival analysis simply without any common mistakes while users applying these tools to their survival observation.

ACKNOWLEDGEMENT

The authors are grateful to the late Prof. Micah B. Masuku, Department of AEM, Faculty of Agriculture, University of Eswatini, Eswatini (Swaziland) for his support and valuable guidance. "*May God rest his soul*."

REFERENCES:

[1] Kiefer, N. M. (1988). Economic Duration Data and Hazard Functions, Journal of EconomicLiterature26:646-679.

[2] Moffitt, Robert A. (1999). New developments in econometric methods for labor market analysis. In: O. Ashenfelter, and D. Card (eds), "Handbook of Labor Economics".

[3] Prinja S, Gupta N, Varma R.(2010). Censoring in clinical trials: Review of survival analysis techniques. Indian Jr. Community Med., 35,217–221.

[4] Armitage, P. B. G. (1959).Statistical Methods in Medical Research. Blackwell.

[5] Narendranathan, W. and M. Stewart (1993). Modeling the Probability of Leaving Unemployment: Competing Risks Models with Flexible Baseline Hazards, Journal of the Royal Statistical Society, Series C, Applied Statistics, 42(1), pp. 63-83.

[6] Balakrishnan, N. (1991). Handbook of the Logistic Distribution, Marcel Dekker, Inc.

[7] Carroll Nick (2005). Explaining Unemployment Duration in Australia.

[8] Foley, M.C. (1997). Determinants of Unemployment Duration in Russia, Yale Economic Growth Centre Discussion Paper 779:39.

[9] KachmannS. D. (1999). Applications in Survival Analysis, Jr. Animal Sc., 77:147-153.

[10] Danica, D. E. and A. G. Babucea (2007). Modeling Time of Unemployment – A Cox Analysis Approach, SOR'7 International Conference, Nova Gorica, Slovenia, 273-279.

[11] Greene, William H. (2003). Econometric Analysis, New York: Prentice-Hall.

[12] Heckman, J. J. and D. D. Singer (1985). Longitudinal Analysis of Labor Market Data. Cambridge UK, Econometric Society Monographs 10, Cambridge Univ. Press, Cambridge, UK.

[13] Kiefer, N.M., 1988. Economic duration data and hazard functions, Journal of Economic Literature 26, 646–679.

[14] Lancaster, T.(1990). The econometric analysis of transition data. Econometric Society Monographs

[15] Dudley, R.A., Harrell, F.E., Smith, L.R., Mark, D.B., Califf, R.M., Pryor, D.B., Glower, D., Lipscomb, J. and Hlatky, M.(1983). Comparison of analytic models for estimating the effect of a clinical factor on the cost of coronary artery bypasses graft surgery. Journal of Clinical Epidemiology 46, 171–261.

[16] Fenn, P., McGuire, A., Phillips, V., Backhouse, M., and Jones, D.(1995). The analysis of censored treatment cost data in economic evaluation. Medical Care 33, 851–863.

[17] Ruth D. Etzioni, Eric, J. Feuer, Sean D. Sullivan, Danyu, Lin., Chengcheng, Hu. & Scott D. Ramsey (1999). On the use of survival analysis techniques to estimate medical care costs, Journal of Health Economics, 18, 365–380.

[18] Adrian Gepp Facul and Kuldeep Kumar (2008).The Role of Survival Analysis in Financial Distress Prediction, International Research Journal of Finance and Economics, Vol. 16, Issue 16.

[19] Tim Drye, Graham Wetherill, and Alison Pinnock (2001). Journal of Targeting, Measurement, and Analysis for Marketing, Vol. 10, 2, 179-188.

[20] Michael J. A. Berry (2009). The Application of Survival Analysis to Customer-Centric Forecasting, NESUG Proceeding, Data Miners, Inc., Cambridge, MA.

[21] Rashid K. (2019). Ph. D. Thesis, Survival Analysis To Understand Customer Retention

[22] Goetz, S. J. (1992). A Selectivity Model of Household Food Marketing Behaviour in Sub-Saharan Africa.American Journal of Agricultural Economics,74 (2), 444–452.

[23] Scherm, H. andOjiambo, P. S. (2004). Applications of survival analysis in botanical epidemiology. Phytopathology, 94, 1022– 1026.

[24] Hay, F. R., Mead, A., and Bloomberg, M. (2014). Modeling seed germination in response to continuous variables: Use and limitations of probit analysis and alternative approaches. Seed Science Research, 24, 165– 186.

[25] Isabel Lambrecht Catherine Ragasa (2016). A Survival Analysis of Contract Farming Participation in Northern Ghana, IFPRI Discussion Paper 01575.

[26] Załuski, D., Mielniczuk, J., Bronowicka-Mielniczuk, U., Stolarski, M. J., Krzyżaniak, M., Szczukowski, S., & Tworkowski, J. (2018). Survival analysis of plants grown in long-term field experiments. Agronomy Journal, 110, 1791– 1798.
[27] Onofri A., Piepho H. P. and Kozak M. (2019). Analyzing censored data in agricultural research: a review with examples and software tips. Ann. Appl. Biol. 2019; 174: 3–13.
[28] Santosha R., Amit Saha, and Kanchan Sinha (2020). Introduction to Survival Analysis and Application in Agricultural Research, Food, and Scientific Reports, 1(4), 28-30.
[29] Benedikt Kramer, Anke Schorr, Reiner Doluschitzand Markus Lips (2019). Survival Analysis for the Adjustment Phase Following Investment in Swiss Dairy Sheds, *Agriculture*, *9*(11), 238.
[30] Cox, D. R. and Oakes D. (2001). Analysis of survival data. London, England: Chapman and Hall.
[31] Lee, H. P. (1982). On clinical trials and survival analysis, Singapore Med J., 23:164–7.
[32] Booth, J. G. and Hirschl, T. A. (2005). Life Table analysis using weighted survey data.
[33] Namboodiri, N. K. and Suchindran, C. M. (1987).*Life table technique and their applications*, Orlando, Florida, Academic Press.
[34] Fergany (1971). On the Human Survivorship Function and Life Table Construction, *Demography*,8(3), 331-334.
[35] SADC Report (2020). SADC selected economic and social indicators, SADC Secretariat, Gaborone, Botswana.
[36] Abbott, R. D. (1985). Logistic regression in survival analysis, American Journal of Epid., 121, 465–471.
[37] Fabsic, P., Evgeny, V. and Zemmer, K. (2011).Seminar in Statistics: Survival Analysis Presentation 3: The Cox proportional hazard model and its characteristics. Zurich.
[38] Mark, S. (2009).An introduction to survival analysis, EpiCentre, IVABS, Massey University.
[39] Zwiener, I., Blettner, M. and Hommel, G. (2011). Survival analysis,Dtsch Arztebl Int., 108,163–169.