International Journal of Scientific Research in Mathematical and Statistical Sciences **Research Paper**
Vol.6, Issue.1, pp.187-192, February (2019) **E-ISSN:** 2348-4519
DOI: https://doi.org/10.26438/ijsrmss/v6i1.187192

Comparison of Different Parametric Modeling for Time-to-Event Data among Cancer Patients

K. Srividhya[1*] and A. Radhika[2]

[1]Department of Statistics, Periyar University, Salem-11
[2]Department of Statistics, Periyar University, Salem-11

Available online at: www.isroset.org

Received: 06/Feb/2019, Accepted: 14/Feb/2019, Online: 28/Feb/2019

Abstract **-**Parametric models are widely used in the modelling of survival data under various diseases. These parametric models were applied to the data of 350 patients of uterus cancer. The main objective of this paper is to compare the results of survival analysis of uterus cancer patients by using different parametric models like Exponential distribution, Weibull distribution, Gompertz distribution, Log - Normal distribution, Log - Logistic distribution and Generalized Gamma distribution by two approaches, the one by Deviance method and the other by Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values. As a result except Exponential and Gompertz distribution all the other distributions gives approximately relative results by these two methods. The model selection of the data is carried out by using Statistical Software STATA 12.

Keywords **-***Survival analysis, Deviance, AIC and BIC values, Uterus Cancer, Parametric Models.*

I. INTRODUCTION

Survival analysis is generally defined as a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest. The event can be death, occurrence of a disease, marriage, divorce, etc. The time to event or survival time can be measured in days, weeks, years, etc. For example, if the event of interest is heart attack, then the survival time can be the time in years until a person develops a heart attack.

Survival analysis techniques used for dealing with censored data can be broadly classified into three techniques.

- Parametric (Exponential, Weibull, Gompertz, Log- Normal, Log –Logistic and Generalized Gamma etc),
- Semi-parametric (Cox Proportional Hazard Method) and
- Nonparametric (Kaplan - Meier method, Log-Rank test).

Two survival regression methods, Cox regression and parametric models were compared and concluded that in univariate analysis the data strongly supported the Log - Normal regression among parametric models and it can be lead to more precise results as an alternative to Cox [1]. The parametric regression models were discussed and showed that Log - Normal model is better than other models [2]. The best parametric model was determined as Log - Logistic model when compared to exponential and Weibull model [3]. The AIC values are compared the survival analysis by using Weibull, Gamma, Gompertz, Log - Logistic and Log - Normal models and concluded that the Gompertz distribution model was more suitable for these survival data [4]. Four parametric survival models are discussed using AIC values and finally concluded that the Log - Normal survival model is the best model [5].

II. DATA FOR STUDY AND DISTRIBUTION

This computed data were taken from NCBI of the uterus cancer patients. This Uterus Cancer data consists of 350 patients with 9 covariates, i.e., 18 variables. The event of interest is survival time. The covariates are given below, 1. Age (Years), 2. Menopausal status (1 = Yes and 2 = No), 3. Hormone Therapy (1 = Yes and 2 = No), 4. Tumor Size (mm), 5.Number of Nodes involved (1 - 51), 6. Tumor Grade(1 - 3), 7. Number of Progesterone Receptors (1 – 2380), 8. Number of Estrogen Receptors (1 - 1144) and9. CA125 (38 -50 units / ml).Event is coded as 1 and censoring is coded as 0.

A. Exponential Distribution

The following are the probability density, survivorship and hazard functions of Exponential distribution with a random variable T is given by

$$f(t,\lambda)=\lambda e^{-\lambda t}$$

$$S(t,\lambda)=\exp\left\{-\int_0^t \lambda du\right\}=e^{-\lambda t}$$

$$h(t,\lambda)=\lambda,\ where\ \lambda>0,\ for\ 0\le t<\infty.$$

The exponential distribution is widely used because of its simplicity and of theoretical (Bain 1964). A special case of Exponentiated Exponential Model was focused for spinal Tuberculosis data while compared to Exponential, Weibull and Log - Logistic models [6]. For modeling real lifetime data the exponential and Lindley distributions were tried to yield better fit for the data [7]. Three-parameter extension of Generalized Exponential Distribution were discussed and moment generating functions (MGF's) are computed [8].

B. Weibull Distribution

The following are the probability density, survivorship and hazard functions of Exponential distribution with arandom variable T is given by

$$f(t,\lambda,\gamma)=\lambda\gamma\, t^{\gamma-1}\exp\left(-\lambda t^{\gamma}\right)$$

$$S(t,\lambda,\gamma)=\exp\left\{-\int_0^t \lambda\gamma\, u^{\gamma-1}du\right\}=e\left(-\lambda t^{\gamma}\right)$$

$$h(t,\lambda,\gamma)=\lambda\gamma\, t^{\gamma-1},\ where\ \lambda>0,\gamma>0.$$

The Weibull distribution is widely used in modeling weather forecasts in meteorology, and defining the distribution of wind speed in radar modeling. Weibull distribution is favored for performing survival data analysis in industrial engineering[9].In a study was conducted on the nationwide estimators were made for defining the parameters of the Weibull distributions [10].To evaluate possible prognostic factors and to assess the relationship between survival times were focused by Weibull parametric model [11].

C. Gompertz Distribution

The following are the probability density, survivorship and hazard functions of Exponential distribution with a random variable T is given by

$$f(t,\lambda,\gamma)=\lambda\exp(\gamma t)\exp\left(\frac{\lambda}{\gamma}(1-\exp(\gamma t))\right)$$

$$S(t,\lambda,\gamma)=\exp\left[\frac{\lambda}{\gamma}\left(1-e^{\gamma t}\right)\right]$$

$$h(t,\lambda,\gamma)=\lambda\exp(\gamma t),\ where\ \lambda>0,\gamma\in(-\infty.\infty)$$

Gompertz model repeatedly used by medical researchers and biologists in modeling the mortality ratio data was formulated by Benjamin Gompertz in 1825. The Gompertz distribution is applied to describe the distribution of adult lifespans through demographers [12]. Estimate the life expectancy of childhood acute lymphoblastic leukemia by using Gompertz model and life expectancy tables [13].Generalized Gompertz distribution for modelling lifetime data [14].

D. Log - Normal Distribution

The following are the probability density, survivorship and hazard functions of Exponential distribution with a random variable T is given by

$$f\left(t,\mu,\sigma^2\right)=\frac{1}{t\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}\left(\log t-\mu\right)^2\right]t>0$$

$$S\left(t,\mu,\sigma^2\right)=\frac{1}{\sigma\sqrt{2\pi}}\int_t^\infty\frac{1}{x}\exp\left[-\frac{1}{2\sigma^2}\left(\log x-\mu\right)^2\right]dx$$

$$h\left(t,\mu,\sigma^2\right)=\frac{f(t)}{S(t)},\ where\ \sigma>0.$$

The Log- Normal distribution theory was described by McAlister in 1897. In medicine field there are many examples for Log – Normal distribution.The studies on resolving the survival time in cancer, and to resolve the beginning age of Alzheimer's disease be the examples for the medical field studies [15]. To analyze the survival times by using the survival analysis methods like Kaplan – Meier curves, Cox proportional hazard model, Boag Log – normal and Log – Normal [16].The analysis of survival times,distinctivelyin modelling the effects of prognostic factors by Log – Normal distribution [17].

E. Log - Logistic Distribution

The following are the probability density, survivorship and hazard functions of Exponential distribution with a random variable T is given by

$$f\left(t,\lambda,\gamma\right)=\frac{\lambda\gamma\,t^{\gamma-1}}{\left(1+\lambda\,t^{\gamma}\right)^2}$$

$$S\left(t,\lambda,\gamma\right)=\frac{1}{\left(1+\lambda t^{\gamma}\right)}$$

$$h\left(t,\lambda,\gamma\right)=\frac{\lambda\gamma\,t^{\gamma-1}}{1+\lambda t^{\gamma}},\ where\ \lambda>0,\gamma>0.$$

The Log- Logistic distribution is best suitable in analyzing survival data conducted by Cox, Cox and Oakes, Bennet, O'Quinley and Sruthers[18]. A study highlighted that the Maximum Likelihood Estimation was the best suitable method in estimating the parameters using Log – Logistic distribution on grouped failure time data [19].

F. Generalized Gamma Distribution

The generalized gamma distribution is a continuous probability distribution with three parameters. It is a generalization of the two-parameter gamma distribution.

The following are the probability density, survivorship and hazard functions of Exponential distribution with a random variable T is given by

$$f\left(t,\alpha,\beta,\lambda\right)=\frac{\lambda\beta}{\Gamma(\alpha)}\left(\lambda t\right)^{\alpha\beta-1}e^{-(\lambda t)^{\beta}},$$

$$S\left(t,\alpha,\beta,\lambda\right)=P(T>t)=1-F(t)$$

$$h\left(t,\alpha,\beta,\lambda\right)=\frac{f(t)}{S(t)},\ where\ t>0,\alpha,\beta,\lambda>0$$

The proposed taxonomy of Hazard function using Generalized Gamma distributions is to study the survival data [20]. The applications of parametric survival models and scrutinized the parametric and semi parametric models in different proportional hazards (PH) and Accelerated Failure Time (AFT) assumptions [21]. A new Mixture Generalized Gamma (MGG) distribution is attained by mixing Generalized Gamma (GG) distribution and Length Biased Generalized Gamma (LGG) distribution [22].

G. Parametric error measurements

To determine the best parametric model in Uterus cancer data, Akaike Information Criterion (AIC) and Bayesian information criterion (BIC) will be calculated.

- **Akaike Information Criterion**

Akaike Information Criterion (AIC) is a measure of selecting a model from a set of models. AIC estimates the quality of each model, relative to each of the other models. The AIC is given by

$$AIC = -2 * \log(likelihood) + 2(k)$$

Where k is the number of parameters in model. Thus, k = 1 for the exponential model, k= 2 for the Weibull, Gompertz, Log-Normal and log-Logistic models and k = 3 for the Generalized Gamma model. Smaller AIC indicate a better model fit.

- **Bayesian Information Criterion**

The Bayesian information criterion (BIC) or also known as Schwarz criterion (SBC, SBIC) is a criterion for model selection among a finite set of models. It is based on the likelihood function and hence, it is closely related to the Akaike Information Criterion (AIC). The BIC is given by:

$$BIC = -2 * \log(likelihood) + k \log(n)$$

Where n is the sample size and k is the number of covariates including an intercept. Same as AIC, BIC also has the value of k =1 is for the exponential model, k = 2 for the Weibull, Gompertz, Log-Normal and log-Logistic models and k = 3 for the Generalized Gamma model. Smaller BIC indicate a better model fit. The BIC generally penalizes free parameters more strongly than does the Akaike Information Criterion, though it depends on the size of n and relative magnitude of n and k.

III. ANALYTICAL METHOD

A. Likelihood Ratio Test

A likelihood ratio test is a statistical test which is used to compare the goodness of fit of two statistical models. The test is based on the likelihood ratio, which expresses how many times more likely the data are under one model than the other. This likelihood ratio can be used to calculate a p-value to decide whether to reject the null model in favor of the alternative model. Both models are fitted to the data and their log-likelihoods are recorded.

IV. MODEL RESULTS

Table 4.1. Parametric Regression model Fitted to Uterus Cancer Data.

Covariates	Exponential		Weibull		Gompertz	
	Haz. Ratio	**S. E**	**Haz. Ratio**	**S. E**	**Haz. Ratio**	**S. E**
Age	1.00150	0.02306	1.00593	0.02329	1.00519	0.02330
Size	1.01536*	0.00661	1.01727*	0.00680	1.01696*	0.00675
Grade	0.90708	0.41707	1.02646	0.48229	0.99926	0.46796
Nodes	1.05759*	0.01964	1.06656*	0.02013	1.06468*	0.02014
Prog_recp	0.99144*	0.00250	0.99063*	0.00257	0.99063*	0.00258
Estrg_recp	1.00026	0.00119	1.00031	0.00120	1.00032	0.00120
CA125	1.07676	0.09554	1.0606	0.09460	1.06537	0.09491
Menopause	1.04464	0.39051	1.03182	0.38292	1.03382	0.38245
Hormone	0.87841	0.24388	0.77022	0.21415	0.76825	0.21453
Deviance	**361.2679**		**338.94936**		**347.56082**	

Covariates	Log - Normal		Log - Logistic		Generalized Gamma	
	Coef.	**S.E.**	**Coef.**	**S.E.**	**Coef.**	**S.E.**
Age	-0.00142	0.01479	-0.00118	0.01372	-0.00139	0.01430
Size	-0.01085*	0.00478	-0.00984*	0.00421	-0.01081*	0.00450
Grade	0.07763	0.27895	-0.01530	0.26882	0.03226	0.28143
Nodes	-0.04287*	0.01390	-0.04236*	0.01234	0.04173*	0.01308
Prog_recp	0.00474*	0.00121	0.00504*	0.00140	0.00485*	0.00129
Estrg_recp	0.00018	0.00066	-0.00003	0.00062	0.00009	0.00068
CA125	-0.07238	0.05450	-0.04314	0.05303	-0.05841	0.05700
Menopause	-0.10153	0.23591	-0.08361	0.22371	-0.08314	0.23112
Hormone	0.16533	0.17718	0.17300	0.16700	0.16377	0.17211
Deviance	**335.89582**		**335.81358**		**335.41242**	

Table 4.2.Parametric Model Selection in Uterus Cancer Data.

Distributions	**Measurements**	
	AIC	**BIC**
Exponential	363.26790	363.81197
Weibull	342.94936	344.03750
Gompertz	351.56082	352.64896
Log-Normal	339.89582	340.98396
Log-Logistic	339.81358	340.90172
Generalized gamma	341.40522	343.04462

V. RESULTS AND DISCUSSION

Two methods were discussed for finding the best parametric model for uterus cancer patients data, First method is Deviance method and the other method is Akaike Information Criterion and Bayesian Information Criterion. The parametric models were fitted using Statistical software STATA 12 and the results are presented in Table 4.1. From the table we conclude that the three covariates namely Size, Nodes and Progesterone receptors are significantly associated with the survival time under all the model assumptions. The deviance of parametric models like Weibull, Log –Normal, Log - Logistic and Generalized gamma distribution are very close to each other when compared to Exponential and Gompertz distribution. Finally we conclude that the above mentioned four parametric models were gives the approximate results.

Besides, the measurements of model selection are calculated by using Akaike Information Criterion and Bayesian Information Criterion. Smaller AIC and BIC indicates a better model fit. Among all parametric distribution given, the distributions like Weibull, Log - Normal, Log - Logistic and Generalized gamma distribution were given the smallest value in AIC and BIC that is very close to each other compared to Exponential and Gompertz distribution are presented in Table 4.2.From the table, we conclude that the above mentioned four parametric models gives the approximate relative results.

VI. CONCLUSION

The aim of this study is to determine the best parametric model for Uterus cancer patient's data by two different approaches. i.e., by deviance and by AIC and BIC values. We conclude that among all distributions given, Weibull distribution, Log - Normal distribution, Log - Logistic distribution and Generalized Gamma distribution were given the approximate results when compared to Exponential and Gompertz distribution. The model selection of Uterus cancer patient's data were carried out by using Statistical Software STATA12.

REFERENCES

[1]. Mohamad Amin Pourseingholi, Ebrahim Hajizadeh, Bijan Moghimi Dehkordi, Azadeh Safaee, Alireza Abadi, Mohammad Reza Zali,"*Comparing Cox Regression and Parametric Models for Survival of Patients with Gastric Carcinoma*", Asian Pacific Journal of Cancer Prevention,Vol. **8,** pp.**412 – 416, 2007**.

[2]. V. Vallinayagam, S. Prathap, P. Venkatesan,"*Parametric Regression Models in the Analysis of Breast Cancer Survival Data"*.International Journal of Science and Technology, Vol.**3,** Issue. **3,** pp.**163 – 167, 2014**.

[3]. Syahila Enera Amran, M. Asrul Afendi Abdullah, Kek Sie Long, Siti Afiqah Muhamed Jamil,"*Analysis of Survival in Breast Cancer Patients by Using Different Parametric Models*". IOP Conf. Series: Journal of Physics, Conf. Series **890, 2017**.

[4]. Elvan Akturk Hayat, Asli Suner, Burak Uyar, Omer Dursun, Mehmet N.Orman,Gul Kitapcioglu MD,"*Comparison of Five Survival Models: Breast Cancer Registry Data from Ege University Cancer Research Center*", TurkiyeKlinikleri Journal of Medical Science, Vol. **30,** Issue.**5,** pp.**1665-1674, 2010**.

[5]. Phillip Oluwatobi Awodutire, Oladapo Adedayo Kolawole, Oluwatosin Ruth Ilori,"*Parametric Modeling of Survival Times Among Breast Cancer Patients in a Teaching Hospital, Osogbo*",Journal of Cancer Treatment and Research,Vol.**5,**Issue. **5,**pp. **81 – 85, 2017**.

[6]. P. Venkatesan and N. Sundaram,"*Exponentiated Exponential Models for Survival data*", Indian Journal of Science and Technology,Vol. **4,** Issue. **8,** pp. **923-930, 2011**.

[7]. Rama Shanker, Hagos Fesshaye and Sujatha Selvaraj,"*On Modeling of Lifetimes Data Using Exponential and Lindley Distributions*".Biometrics & Biostatistics International Journal, Vol.**2,** Issue. **5, 2015**.

[8]. N.A. Rather and T.A.Rather, "*New Generalizations of Exponential Distribution with Applications*",Journal of Probability and Statistics,**Vol. 2017,2017.**

[9]. W. Weibull,"*A Statistical Distribution functions of Wide Applicability*",J ApplMech, Vol,**18,** Issue.**2,**pp. **293 – 297,1951**.

[10]. R. Inghelmann, E. Grande, S. Francisci et al,"*National Estimators of Cancer Patients Survival in Italy: A Model Based Method*",Tumori,Vol. **91,** Issue.**2,** pp. **109-115, 2005**.

[11]. Ahmad Reza Baghestani, Sahar Saeedi Moghaddam, Hamid Alavi Majd, Mohammad Esmaeil Akbari, Nahid Nafissi, Kimiya Gohari,"*Survival Analysis of Patients with Breast Cancer using Weibull Parametric Model*",Asian Pacific Journal of Cancer Prevention,Vol.**16,**Issue.**18,** pp. **8567-8571, 2015**.

[12]. S. Viscomi, G.Pastore, E. Dama, L. Zuccolo, N. Pearce, F.Merletti& C.Magnani, "*Life Expectancy as an Indicator of Outcome in follow-up of Population- Based Cancer Registries: The Example of Childhood Leukemia*", Annals of Oncology, Vol. **17,**pp. **167 – 171, 2006.**

[13]. Vaupel, James W,"*How Change in Age – Specific Mortality Affects Life Expectancy*", Population Studies, **40 (1),** pp. **147 – 57, 1986**.

[14]. A. El-Gohary, Ahmad Alshamrani, Adel NaifAl-Otaibi, "*The Generalized GompertzDistribution"*,Applied Mathematical Modelling,Vol. **37,**pp.**13-24, 2013**.

[15]. R.D. Horner,"*Age at Onset on Alzheimer's Disease; Clue to the Relative Importance of Etiologic Factors?*", Am J Epidemio,Vol. **126,**Issue.**3,** pp. **409 –414, 1987**.

[16]. P. Tai, J.A. Chapman, E.Yu, D. Jones, C. Yu, F. Yuvan, et al,"*Disease –Specific Survival for Limited – Stage Small – Cell Lung Cancer Affected by Statistical Method of Assessment*", BMC Cancer,**7:31, 2007**.

[17]. P. Royston,"*The Log-Normal Distribution as a Model for Survival Time in Cancer, with an Emphasis on Prognostic Factors*",StatisticaNeerlandica, Vol. **55,** pp. **89-104, 2001**.

[18]. Ramesh C. Gupta, OlcayAkman, Sergey Lvin,"*A Study of Log-Logistic Model in Survival Analysis*",Biometrical Journal,Vol. **41,** pp.**431-443, 1999**.

[19]. Yan YanZhou, JieMi,ShengruGuo,"*Estimation of Parameters in Logistic and Log – Logistic Distribution with Grouped Data*", Lifetime Data Anal,Vol. **13,** pp.**421 – 429, 2007**.

[20]. Christopher Cox, Haitao Chu, Michael F. Schneider, Alvaro Munoz,"*Parametric Survival Analysis and Taxonomy of Hazard Functions for the Generalized Gamma Distribution*", Statistics in Medicine, Vol.**26,** pp.**4352-4374, 2007**.

[21]. AlirezaAbadi, FarzanehAmanpour, Chris Bajdik,ParvinYavari,"*Breast Cancer Survival Analysis: Applying the Generalized Gamma Distribution under Different Conditions of the Proportional Hazards and Accelerated Failure Time Assumptions*", International Journal of Preventive Medicine, Vol.**3**,No**. 9, 2012**.

[22]. SatsayamonSuksaengrakcharoen,WinaiBodhisuwan,"*A New Family of Generalized Gamma Distribution and its Application*", Journal of Mathematics and Statistics, Vol. **10**, Issue.**2,** pp.**211-220, 2014**.

[23]. D. Collett,"*Modelling Survival Data in Medical Research*", Chapman& Hall, London,**1994**.

[24]. David G.Kleinbaum, Mitchel Klein, "*Survival Analysis: A Self – LearningText*", Second Edition, Springer, New York, **2005**.

[25]. Elisa T. Lee, John Wenyu Wang, "*Statistical Methods for Survival Data Analysis*",Third Edition, Wiley, New York,**2003**.

AUTHORS PROFILE

Mrs. K.Srividhya, M.Sc., M.Phil, (Ph.D) is Research Scholar at Department of Statistics, Periyar University, Salem – 636 011. Under the guidance of Dr. A. Radhika.

Dr. A. Radhika, M.Sc., M.Phil., Ph.D. She is currently working as Assistant Professor in Department of Statistics, Periyar University, Salem – 636 011. She is a member of ISPS since 2013. Her main research work focuses on Survival Analysis, Bio Statistics, Clinical Trials, Time Series Analysis and Design of Experiments. She has more than 6 years of research experience.

