CODEN [USA]: IAJPBB ISSN: 2349-7750

INDO AMERICAN JOURNAL OF
PHARMACEUTICAL SCIENCES

http://doi.org/10.5281/zenodo.845389

Available online at: http://www.iajps.com *Case Study*

A COMPARATIVE ANALYSIS OF TWO-PARAMETER WEIBULL AND GENERALIZED WEIBULL MODELS WITH BAYESIAN ESTIMATIONS FOR MODELLING SURVIVAL DATA WITH INTERVAL CENSORING TO DETERMINATION OF AGE AT DEATH. CASE STUDY: DISCOVERED HUMAN REMNANTS FROM CEMETERY OF THE HISTORICAL SITE OF SHAHR-I-SOKHTA, ZABOL-IRAN

Abdolkarim Shadmehr[1], Ebrahim Hajizadeh[2*], Anoshirvan Kazemnejad[2], Ahmadreza Baqestani[3], Moein Yoosefi[4]

[1]Ph.D. Student in Biostatistics, Faculty of Medicine Sciences, Tarbiat Modares University, Tehran, Iran
[2] Professors of Biostatistics, Faculty of Medicine Sciences, Tarbiat Modares University, Tehran, Iran
[3]Associate Professor, Faculty of Paramedicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[4] MSc students, Faculty of Paramedicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Abstract:
Human skeletons are among the most important data in archaeological records; it is important to study these data in order to extract biological and cultural information. One of the most widely used statistical methods for analyzing such data is the survival analysis. In this paper, we try to compare the two models of survival analysis for estimating age at death in presence of covariates in an ancient site. This is a retrospective cohort study based on the findings from twelve seasons of excavation at the site of Shahr-i-Sokhta in southeastern Iran, from 1997 to 2008. Out of the 758 discovered skeletons, 349 were complete and measurable, which were selected based on an estimated age of over 12 years at the time of their death. For the analytical purposes we used a Two-parameter Weibull regression model as well as a generalized Weibull regression model, both with Bayesian approaches for estimating their relevant parameters. Out of 349 study subjects 206 (59%) were female. The fitting results of both models showed a significant relationship that could be between age at death and the number of buried objects along with skeletons. In other words, for individuals who were associated with more than five objects, the risk of death might have been lower. The other exploratory variables including sex, the presence of sacrifices, and the existence of prestigious goods did not show any significant relationship. The fitting results of these two models on the basis of the data showed that the estimates of the generalized Bayesian model had smaller standard deviations than the corresponding ones from two-parameter Weibul model. The model's adequacy index "Deviance Information Criterion" (DIC) in the generalized Bayesian Weibull model was also smaller than that of the other model. As a result, in this case study, the generalized Bayesian Weibull model was found as more efficient.
Key-Words: *Excavated Skeleton, Shahr-i-Sokhta, Weibull Model, Bayesian Approach.*

Corresponding author:
Ebrahim Hajizadeh,
Professors of Biostatistics,
Faculty of Medicine Sciences,
Tarbiat Modares University, Tehran, Iran.
Email: hajizadeh@modares.ir

QR code

Please cite this article in press as Ebrahim Hajizadeh et al, ***A Comparative Analysis of Two-Parameter Weibull and Generalized Weibull Models with Bayesian Estimations for Modelling Survival Data with Interval Censoring to Determination of Age at Death. Case study: Discovered Human Remnants from Cemetery of the Historical Site of Shahr-I-Sokhta, Zabol-Iran****, Indo Am. J. P. Sci, 2017; 4(08).*

INTRODUCTION:

Skeletal and dental remnants, and the information they provide, including the average age of the society, the type of relations between buried people, the society's health etc., are at the center of attention for archaeological studies [1]. The site of Shar-i-Sokhta is one of the earliest urban centers in the eastern part of the Iranian plateau. The site is located about 60 km to the east of Zabol, in Sistan va Baluchestan province. The human settlements at Shahr-i-Sokhta go back to the fifth millennium B.C. The site extends over an area of 152 hectares, and was a focal center for commerce and arts on the eastern fringe of the Iranian plateau [2, 3]. The knowledge derived from the archaeological evidence overrides old theories regarding the absolute central role of Mesopotamia as the Cradle of civilization [4]. The discovered remains include "residential areas", "industrial zones", "monuments," and "cemeteries" and other structures, encompassing 1,200 years of life in this ancient city from the end of the fourth millennium to the end of the third millennium BC. The cemetery is one of the richest ancient cemeteries, providing valuable information as well as having diverse objects [2]. The cemetery preserves the remnants of generally similar bodies, but with differences in time and also in the cultural traditions that belonged to various communities that lived here together. It is estimated that the Shahr-i-Sokhta cemetery contains 20,000 graves [5, 6]. Nowadays, statistical analysis has found a special place in archaeological researches. Due to advances in laboratory technologies, many historical cultural materials discovered in ancient sites can be quantified, and the creation of a database as a reservoir of the findings of these cultural materials leads to massive quantitative collections that require proper analyses. One important piece of information that can be obtained and analyzed from the human bone remnants of the artifacts from ancient sites is an estimate of the age of individuals at death, or the estimation of life spans. Analysis and estimation of life spans is one of the most important issues in statistical methods which can be considered as a widely used branch of statistical methods. In this branch, the statistical methods called survival analysis can fit different models into lifetime data models and evaluate the relationship between different variables. One of the factors that make survival analysis more relevant is the ability to incorporate censored data into related models, data that relates to things that their real-time information is not available. Because the time of death occurrence event for the bone remnants is determined at an approximate time interval, the desired survival data can be considered as a kind of censorship [7]. In survival analysis, in general, two categories of regression models are used to explore the statistically significant relationships between time, as a dependent variable, and a set of covariates called as independent variables or predictors as well. The first kind of these models for the above-mentioned purpose is the Cox proportional hazards model as a semi-parametric model and the second kind is accelerated failure time as parametric models, although Cox regression is the most practical model for survival analysis, but parametric models for analysis of survival data in some situations might be more appropriate [8, 9]. Obtaining parametric distribution of survival time is of special importance, because if the parametric model has a proper fitting on the data, it will result in a more precise estimation for the parameters compare to semi-parameter method. The reason for this is that, the estimates are based on the parameters, in other words, the model could be obtained in simpler way [7]. Also, from the distribution probability in subsequent studies could be used for Bayesian methods. Despite the advantages of Bayesian approach over the classical approach [10], however, in most studies, tendency to use the classical approach is more than the Bayesian approach, because Bayesian models require the use of sophisticated methods such as Markov Chain Monte Carlo (MCMC) method and the Gibbs Sampling Algorithm. Statistical distributions such as Weibull, exponential, Gompertz, Gamma, and Log-normal, logistic as parametric models are frequently used for survival analysis, so that the Weibull distribution is one of the most distributions for survival analysis. According to surveys conducted by Shott at the New Guinea Site on the set of stone tools to determine their use life Weibull doubled parameters [11, 12, 13] was used. Recently, Godde [14] benefited from Gompertz model (a special case of the Weibull distribution) to estimate age at the time of death without considering covariates. Accordingly, we can say that the use of generalized Weibull on the survival of discovered human skeletons in an archaeological site in Iran can be considered as an original research, especially considering interval censoring and incorporating covariates into the model. In this paper, statistical models for estimating age at the time of death were obtained by using the generalized Weibull and Weibull distribution with the Bayesian approach to estimate their relevant parameters and coefficients for the factors affecting the survival of the remains of the discovered human skeletons at Shahr-i-Sokhta were studied and finally, these two models were compared, based on the adequacy criterion of fitting.

MATERIALS AND METHODS

This is a retrospective cohort study based on the findings from twelve seasons of excavation at site of Shahr-i Sokhta in southeastern Iran, from 1997 to 2008 [2]. Out of the 758 discovered skeletons, 349 were complete and measurable, which were selected based on an estimated age of over 12 years at the time of their death for entering the study. Calculation of age at death was associated with an estimated age variation in each case, with regard to the growth of organs, bones, joints, skull steator, opiate long bones, teeth growths and their corrosion, therefore a range of age at death was determined. For this reason, the age variable at death was considered as the response variable with an interval censoring. Other information that included covariates such as gender, the existence of sacrifices beside the bodies, the number of objects in the grave and the existence of prestigious goods in grave were measured and used in this study. Parametric models, because of their flexibility and a closed form for survival function and risk function, have a lot of use in analyzing survival data and lifetime. The Weibull model is one of the most popular parametric models and two forms of it are among the most common ones, i.e. models with two parameters and three parameters. In this study, a three-parameter model, generalized Weibull model, and a two-parameter model, the ordinary versions have been used to determine the age at death in the presence of covariates. A Bayesian approach has been used to estimate the model's parameters as well. Finally, the performance of the models was also compared. The survival functions for the two-parameter and generalized Weibull models are as follows [15, 16]:

Weibull model:

$$S(t,\alpha,\beta) = (e^{-\alpha t})^{\beta} \quad t,\alpha,\beta > 0$$

Generalized Weibull model:

$$S(t,\alpha,\beta,\gamma) = 1 - F(t,\alpha,\beta,\gamma) = 1 - (1 - (e^{-\alpha t})^{\beta})^{\gamma} \quad t,\alpha,\beta,\gamma > 0$$

Since there were no information on the prior distribution of the parameters, a non-informative uniform distribution with a range of (-100 , 100) was considered as the prior distribution of parameters, and finally, based on one of the criteria used in the Bayesian benchmarking section, so called "Deviations Information Criterion" (DIC), presented by SPIEGEL HATTER et al [17]. was selected as the best-fit model for choosing a model with the lowest DIC. In this study, SAS software (Version 9.4) has been used to fit statistical models of the research and compare them with each other.

FINDINGS:

A total of 349 of adult researchable skeletons were detected from the Shahr-i-Sokhta as subjects of the study. Out of these 349 triable subjects, 206 (59%) were female. In addition to these discovered skeletons, 26 (7%) scarifies and 36 (10%) prestigious goods were respectively studied. The mean and standard deviation of the number of buried objects with the discovered skeletons were 4.8 with a standard deviation of 9.21. Mean (IQR) of estimated age of these skeletons at the time of death was 32.5 (20) years (Table 1). The findings of the fitting of the Two-Parameter Weibull model with Bayesian method showed that the Bayesian confidence interval for the variable coefficient "number of buried objects alongside skeletons" is not including zero; therefore, the age at death with the variable of the corresponding regimes was statistically significant: In such a way that for skeletons with more than 5 objects the risk of death was observed at lower age. However, the effect of sex, the existence of sacrifices and the existence of prestigious objects were not statistically significant (Table 2). Based on table 3, the findings related to the fitting of the generalized Weibull Bayesian model with interval censoring to the data, likewise the two-Parameter Weibull model, showed a statistically significant relationship between the age at death and the number of buried objects (The obtained Bayesian confidence interval did not include zero). But the three other variables, i.e. gender, sacrifices and existence of prestigious objects did not show any significant relationship with age at the time of death. The observed difference between the fittings of these two models was that the Bayesian confidence interval obtained in the generalized Weibull was shorter than that obtained in the two-parameter Weibull, which indicates a higher accuracy of the generalized Weibull model than the other model. The Deviation Information Criterion index (DIC)[17], was used in order to compare the fitting of the two models. The values of the index for the two-parameter model and generalized model were DIC = 1647 and DIC = 1624, respectively. Therefore, it can be concluded that the generalized Bayesian Weibull model had a better fit on the data of this research than the Bayesian Weibull model.

Table 1: Descriptive Statistics for General Characteristics of Study Subjects.

Characteristics	Categories	Descriptive statistics
Survival Time (Year)	-	33.5 ± 12.01 (Range: 12-85)
Total Number of Objects	-	4.8 ± 9.21 (Range: 0-75)
Gender	**Male**	143 (41%)
	Female	206 (59%)
Sacrifice	**Presence**	26 (7%)
	Absence	323 (93%)
Prestigious Goods	**Presence**	36 (10%)
	Absence	313 (90%)

*Mean± SD for continues variables and Number (%) for the categorical variables

Table 2: Results of the Bayesian Weibull Model for Assessing the Effect of Different Factors on the Age at Death Time.

Variable	Category	Estimate(β)	SD	95%CI for HP
Numbers of Objects	**<5**	Reference	-	-
	>5	-0.389	0.13	[-0.635;-0.109]
Gender	**Female**	Reference	-	-
	Male	-0.033	0.11	[-0.234; 0.189]
Sacrifice	**Absence**	Reference	-	-
	Present	-0.031	0.25	[-0.525; 0.448]
Prestigious Goods	**Absence**	Reference	-	-
	Present	0.08	0.22	[-0.355; 0.515]

Table 3: Results of the Bayesian Generalized Weibull Model for Assessing the Effect of Different Factors on the Age at Death Time.

Variable	Category	Estimate(β)	SD	95%CI for HP
Numbers of Objects	**<5**	Reference	-	-
	>5	-0.136	0.06	[-0.252;-0.016]
Gender	**Female**	Reference	-	-
	Male	-0.131	0.07	[-0.244; 0.020]
Sacrifice	**Absence**	Reference	-	-
	Present	-0.052	0.10	[-0.258; 0.130]
Prestigious Goods	**Absence**	Reference	-	-
	Present	-0.012	0.11	[-0.199; 0.207]

DISCUSSION:

Discovered skeletal remains from archaeological sites give us important and valuable information about the human societies of the past. Hence, the distribution of estimated longevity of skeletons and objects can reveal a sign of the quality of life of the communities under study. The present paper focused on the use of a generalized Bayesian Model with a Censorship in the presence of covariates. The fitting of this model were done for the first time on the discovered human remains from the archaeological cemetery of Shahr-i Sokhta, Zabol, Sistan va Baluchestan. In this research, the effect of the covariates gender, existence or absence of prestigious goods, the presence or absence of sacrifices and the number of discovered artifacts in the grave were analyzed on the survival of the discovered skeletons from Shahr-i-Sokhta belonging to the third millennium BC. In a study like this we are confronted not with a live and dynamic society, but with the archaeological remains of an ancient community (here about 5,000 years), certainly more research will be required. The Bayesian approach in statistics has been widely used for model building, but, compared to other statistical methods for estimation; this approach has not been extensively utilized with survival analysis data. Since Survival models rely on actual data, calculation of the posterior distribution is not easily possible. In recent years, with the development of computational algorithms such as the MCMC method and also the promotion of computer processing, the application of the Bayesian methods in survival analysis ~~is~~ has become more feasible. Consequently, when information on covariates is scarce, the sample size is small, or the data is of censored type, then the Bayesian approach is preferred. On the other hand, the non-Bayesian approach is based on the normal approximation, which is true only in large samples, and the obtained confidence intervals are reliable only when the size of sample is not small. In fact, such inferences cannot be used for small or medium samples [17]. The results of the fitting of the Weibull model and the generalized Weibulll model with the Bayesian approach in investigating the factors affecting the survival time of the discovered human skeleton from the Shahr-i-Sokhta showed that the life span of the same age at death is statistically meaningful only when determined with the variable of the number of objects. In the graves where the number of objects along with the discovered skeletons exceeded five, greater life spans were observed; in other words, the number of objects showed a significant relationship with the livelihood of the individual. This can function as a sign of the quality of life over the life span of individuals. Therefore, the number of objects in the graves as one of the life quality indicators showed the most significant relationship with the duration of life of the people of that time. As for the other auxiliary variables, they were found to have a direct relationship with survival time, but these relationships were not meaningful from a statistical point of view. Findings for the fitting of the generalized Weibull and Weibull models in other studies prove the superiority of the generalized Weibull model over the normal Weibull [18].

CONCLUSION:

Our findings, also consistent with these studies, show a more suitable fit for the generalized Weibull model. The fitting of these two models on the data presented in this study suggests that the estimation of the generalized Weibull model compared to the Bayesian Weibull model presents a smaller standard deviation. Also, the model's adequacy index (DIC) in this generalized Bayesian is smaller than the other model, and therefore, the generalized Bayesian Weibull model proves itself as more adequate. As a proposal for further research, similar investigations on the age at the time of the death in archaeological sites using advanced statistical models such as the Weibull models family are recommended. This is hoped to result in comparisons of the findings of different archaeological sites in appropriate ways and provide a better picture of the life of ancient societies.

ACKNOWLEDGMENTS

This research is some part of a PhD thesis which financially supported by Tarbiat Modares University. The authors would like to sincerely thank Dr. Seyyed Mansour Seyyed Sajjadi, the director of the archeological excavations at Shahr-i-Sokhta for his magnificent support of this research. Our heartfelt gratitude also goes with Mr. Seyyed Mohammad Beheshti, Head to Research Institute of Cultural Heritage and Tourism (RICHT), Dr. Hamideh Chubak, Dean of Iranian Center of Anthropological Research (ICAR), and Mr. Farzad Forouzanfar for providing access to the resources of the present research.

REFERENCES:

1.Roberts CA. Human remains in archaeology: a handbook: Council for British Archaeology, 2009.
2.Forouzanfar F. Reports of Shahr-i Sokhta 3 (Excavations at Graveyards: Anthropological Studies 1997-2008), 2010.
3.Tosi M. Excavations at Shahr-i Sokhta, a Chalcolithic Settlement in the Iranian Sīstān. Preliminary Report on the First Campaign, October-December 1967. East and West. 1968; 18(1): 9-66.
4. Sajjadi S, Mansour S. What do I know about Iran? Shahr-i Sokhta. Tehran: Bureau for Cultural Research, 2011.
5.Piperno M, Salvatori S. The Shahr-I Sokhtya Graveyard (Sistan, Iran): Excavation Campaigns 1972-1978. IsIAO; 2007.
6.Salvatori S, Vidale M, Istituto italiano per l'Africa e l'Oriente. Shahr-i Sokhta 1975-1978: Central Quarters Excavations: Preliminary Report. IsIAO; 1997.
7. Kleinbaum DG, Klein M. Survival analysis: a self-learning text: Springer Science & Business Media, 2013.
8.Lawless JF. Parametric models in survival analysis. Encyclopedia of Biostatistics, 1998.
9.Lindsey JK. A study of interval censoring in parametric regression models. Lifetime data analysis. 1998; 4(4): 329-54..
10.Congdon P. Applied bayesian modelling. John Wiley & Sons, 2014.
11.Shott MJ. Weibull estimation on use life distribution in experimental spear-point data. Lithic Technology. 2002; 27(2): 93-109.
12.Shott MJ, Seeman MF. Curation and recycling: estimating Paleoindian endscraper curation rates at

Nobles Pond, Ohio, USA. Quaternary International. 2015; 361(9): 319-31.
13.Shott MJ, Sillitoe P. Modeling use-life distributions in archaeology using New Guinea Wola ethnographic data. American Antiquity. 2004; 7(3): 339-55.
14.Godde K, Hens SM. Age-at-death estimation in an Italian historical sample: A test of the Suchey-Brooks and transition analysis methods. American journal of physical anthropology. 2012; 149(2): 259-65.
15.Mudholkar GS, Srivastava DK. Exponentiated Weibull family for analyzing bathtub failure-rate data. IEEE Transactions on Reliability. 1993; 42(2): 299-302.
16.Mudholkar GS, Srivastava DK, Kollia GD. A generalization of the Weibull distribution with application to the analysis of survival data. Journal of the American Statistical Association. 1996; 91(436): 1575-83.
17.Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002; 64(4): 583-639.
18.Yoosefi, M., Baghestani, A. R., Khadembashi, N., Pourhoseingholi, M. A., Baghban, A. A., & Khosrovirad, A. Survival analysis of colorectal cancer patients using exponentiated Weibulll model. International Journal of Cancer Management. 2017; 5(2): 61-8.