International Journal of Multidisciplinary Research and Development
Online ISSN: 2349-4182, Print ISSN: 2349-5979; Impact Factor: RJIF 5.72
Received: 12-10-2019; Accepted: 14-11-2019
www.allsubjectjournal.com
Volume 6; Issue 12; December 2019; Page No. 57-59

Statistical comparison of survival models between cox model and exponential model for analysis of breast cancer patients

Elsamoual Mohammed Kurtokeila[1*], Dr. Hamza Ibrahim Hamza[2], Dr. Amal Alsir Alkhidir[3]
[1, 3] Department of Statistics, Sudan University of Science and Technology, Sudan
[2] Alzaiem Alazhari University, Sudan

Abstract
This paper discuss the comparison between semi parametric model (Cox regression) and parametric model (Exponential model) according to survival time distribution. To evaluate the factors that effect on the survival of the breast cancer patients, to determine the best model Akaike Information Criterion (AIC) and Bayesian information criterion (BIC) was used. A sample of 246 was taken from Khartoum Oncology Hospital, 150 (61%) died. The results of data analysis showed that the Hazard Ratios for the factors studied in Exponential Model is lower than the Hazard Ratios for the factors studied in Cox Model. AIC and BIC criterion in Exponential model is less than Cox Model, this indicate that the Exponential model is a favourable for the data of this study.

Keywords: Brest cancer, cox regression, exponential model – Aic - Bic

1. Introduction

Prognosis, the prediction of the future of an individual patient with respect to duration, course, and outcome of a disease plays an important role in medical practice. Before a physician can make a prognosis and decide on the treatment, a medical history as well as pathologic, clinical, and laboratory data are often needed. Therefore, many medical charts contain a large number of patient prognostic factors, and it is often difficult to sort out which ones are most closely related to prognosis. The physician can usually decide which characteristics are irrelevant, but a statistical analysis is usually needed to prepare a compact summary of the data that can reveal their relationship. One way to achieve this purpose is to search for a theoretical model (or distribution), that fits the observed data and identify the most important factors. These models, usually regression models, in this paper we focus on semi-parametric and parametric regression models [7].

Cox's proportional hazard model once of survival models, it's a semi-parametric model used to study explanatory variables affecting time until event occurs. It is flexible model, is no need to assume a particular form of probability distribution for the survival time [3, 9].

If the assumption of a particular probability distribution for the data is valid, inferences based on such an assumption will be more precise. Models in which a specific probability distribution is assumed for the survival times are known as parametric model. A probability distribution that plays a central role in the analysis of survival data is the Exponential.

The Exponential model is once of parametric model. If the parametric models better fit the data, a more precise estimation of coefficients would be achieved.

Maximum likelihood is used for estimation of coefficients in survival parametric model, while partial likelihood used for Cox's proportional hazard model [8].

2. Materials and Methods

Data were taken from the Khartoum Hospital for Oncology in the period from June 2013 to April 2017 with various factors information. Multivariate analysis of prognostic factors was carried out by two methods: Cox proportional hazard (PH) model (as semi-parametric method) and Exponential model (as parametric methods) [3, 4]. Cox's model has become the most used procedure for modelling the relationship of covariates to a survival or other censored outcome [10]. However, it has some restrictions. One of the restrictions to using the Cox model with time-fixed covariates is its proportional hazards assumption, it means the hazard ratio between two sets of covariates is constant over time. This is due to the common baseline hazard function cancelling out in the ratio of the two hazards.

The general form of the Cox PH model is

$$h(t/x) = h_0(t).e^{\sum_{i=1}^{p}\beta_i x_i}$$

Where $h_0(t) \equiv$ Baseline hazard function, exponential expression (e) to the linear sum of $\beta_i x_i$ where the sum is over the P explanatory X variables [6]. For Exponential model Baseline hazard function is the parameter of Exponential distribution, Hazards are constant but not necessarily constant in the Cox PH model [6]

$$h_0(t) = \lambda$$

$$h(t) = \exp\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right)$$

3. Evaluation Criteria

To set a comparison among models we used Akaike Information Criterion (AIC) and Bayesian information criterion (BIC).

AIC is an estimator of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models, the selection of the model is important, as under-fitting a model may not capture the true nature of the variability in the outcome variable, while an over-fitted model loses generality [1].

BIC is an estimate of a function of the posterior probability of a model being true, under a certain Bayesian setup, so that a lower BIC means that a model is considered to be more likely to be the true model [6].

AIC and BIC showed that this selection of the "best" model is determined by:

An AIC score

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

Where p is the number of estimable parameters (degrees of freedom) and $\log(likelihood)$ is the log-likelihood at its maximum point of the model estimated. The constant"2"remains"forhistorical reasons, $(k=1)$ for exponential model. Lower AIC indicates better likelihood.

A BIC score

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

Where n is the sample size.

an estimate of a function of the posterior probability of a model being true, under a certain Bayesian setup, so that a lower BIC means that a model is considered to be more likely to be the true model [2]. All calculations were carried out by STATA statistical software (version 10).

4. Results and Discussion

Table 1

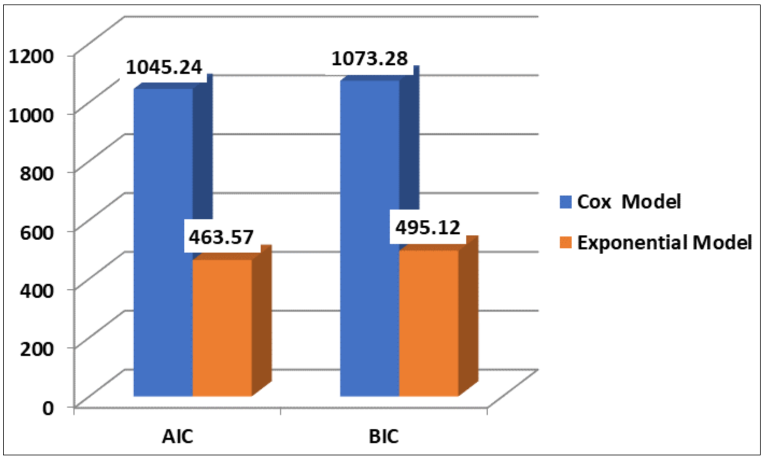
Factors	Cox proportional hazard Model			Exponential Model		
	Hazard Ratio	[95% Conf. Interval]	p-value	Hazard Ratio	[95% Conf. Interval]	p-value
Stage (I, II,III,IV, V)	0.812	(0.693 - 0.950)	0.009	0.797	(0.692 - 0.918)	0.002
Radiotherapy (given, not given)	2.784	(1.878b - 4.128)	0.000	2.062	(1.449 - 2.934)	0.000
Hormonal (given, not given)	2.442	(1.607 - 3.711)	0.000	1.596	(1.098 - 2.320)	0.014
Surgical (Yes, No)	2.901	(1.481 - 5.682)	0.002	2.256	(1.179 - 4.315)	0.014
Chi- square (P-value)	80.02 (0.000)			52.10 (0.000)		

Source: prepared by the researcher by using STATA (v10), 2019.

Table 2

Criterion	Cox proportional hazard Model	Exponential Model
AIC	1045.24	463.57
BIC	1073.28	495.12

Source: prepared by the researcher by using STATA (v10), 2019.



Source: prepared by the researcher by using Excel.

Fig 1

There were 246 patients 150 (61%) died. We compared semi parametric (Cox model) and parametric (Exponential model) by using AIC and BIC.

Table 1: shows the results of multivariate analysis using Cox model and alternative Exponential model. According to Cox regression analysis, among variables that entered to

model (age, education level, marital status, work, stages, radiotherapy, chemotherapy, hormonal, surgical and time from date of diagnosis until death or last follow-up or end date of study per week) the factors influencing on survival of patients were: stages, radiotherapy, hormonal, surgical) ($p<0.05$). In Cox Model at any time during the study, rate of death per-week among stage is 0.812, among radiotherapy is 2.784, among Hormonal is 2.442, among Surgical is 2.901 and the value of Chi-square for Cox Model (80.02) with P-value (0.000). In Exponential Model at any time during the study, rate of death per-week among stage is 0.797, among radiotherapy is 2.062, among Hormonal is 1.596, among Surgical is 2.256, and the value of Chi-square for Exponential Model (52.10) with P-value (0.000).

Table 2: shows the results AIC (1045.24) and BIC (1073.28) for Cox Model. And AIC (463.57) and BIC (495.12) for Exponential Model.

The results of data analysis showed that the Hazard Ratios for the factors studied in Exponential Model is lower than the Hazard Ratios for the factors studied in Cox Model. AIC and BIC criterion in Exponential model is less than Cox Model (figure 1), this indicate that the Exponential model are the best model in multivariate analysis.

5. Conclusion

The Cox model is the most commonly used model in survival analysis studies because it does not depend on knowing the distribution of survival time when estimating model parameters.

This paper aimed to investigation the comparative performance of Cox Model and Exponential model in a survival analysis of patients with Breast cancer. We used Akaike Information Criterion (AIC) and Bayesian information criterion (BIC) to evaluate among models.

The results of data analysis showed that the Hazard Ratios for the factors studied in Exponential Model is lower than the Hazard Ratios for the factors studied in Cox Model. AIC and BIC criterion in Exponential model is less than Cox Model, this indicate that the Exponential model are the best model in multivariate analysis. When survival time distributed according to known probability distribution the parametric model (Exponential Model) is best than the semi parametric (Cox Regression Model).

6. References

1. Petrov BN, Csaki F. (Eds.), Second International Symposium on Information Theory, Akademiai Kiado, Budapest, 1973, 267-281.
2. Chris T Volinsky, Adrian E Raftery. Bayesian Information Criterion for Censored Survival Models. 2000; 56(1):256-262.
3. Collett D. Modelling survival data in medical research. Boca Raton, Fla: Chapman & Hall/CRC, 2nd edition, 2003.
4. Collett D. Modelling survival data in medical research. Chapman & Hall/CRC, 3rd edition, 2014.
5. Cox DR. Regression models and life tables (with discussion), Journal of the Royal Statistical Society, 1972, 34.
6. David, Mitchel, Survival analysis 3rd, 2012.
7. Elisa T Lee, John Wang. Statistical Methods for Survival Data Analysis Statistical Methods for Survival Data Analysis, Third Edition, Hoboken, New Jersey, John Wiley & Sons, 2003.
8. Faiz AM Elfaki. Exponential Model for Survival Analysis, International Journal of Computer Science and Network Security, 2017, Vol.17.
9. Kleinbaum DG, Klein M. The Cox Proportional Hazards Model and Its Characteristics. In: Survival Analysis. Statistics for Biology and Health. Springer, New York, NY, 2012.
10. Therneau T, Grambsch P. Modeling Survival Data: Extending the Cox Model, New York: Springer-Verlag, 2000.