FITTING SEMI PARAMETRIC AFT MODEL IN SURVIVAL DATA

Dr. Rinku Saikia[1]

[1] Assistant Professor, Golaghat Commerce College, Golaghat

Email: saikiarinku25@gmail.com

Abstract

Semi parametric Accelerated failure time (AFT) model is a log-linear model. This model is a combination of parametric part of regression coefficient and a nonparametric part for unknown error distribution. The main objectives of this paper is to fit semi parametric AFT model by rank estimation method and to study the effect of different factors on the survival of esophagus cancer patients by using semi parametric AFT model.

Key words: Semi-parametric, AFT, rank estimation, factor

1. INTRODUCTION

Semi parametric Accelerated failure time (AFT) model is a log-linear model. This model is a combination of parametric part of regression coefficient and a nonparametric part for unknown error distribution. The probability distribution with censored observations in which the error term is specified is called accelerated failure time model. The AFT model with unspecified error distribution considered as a semi parametric model can be the alternative to the Cox model because of its simple interpretation. The AFT model gives a direct linear relationship between the failure time and covariates i.e., the logarithmic function of the failure time model has linear effect on the covariates. However, estimation and inference process of this model is difficult in the presence of censoring. The AFT models are introduced by Cox (1972) and considered by Prentice (1978) discussed by Kalbfleisch & Prentice (2002) and Lawless (2002).

Jin and Ying (2004) considered asymptotic theory of rank estimation for AFT model under fixed censorship. Zhou (1992) and Jin (2007) considered M- estimation for the AFT models. For

the heteroscedastic errors, Stute (1993, 1996) measured convergence properties of weighted estimators and Zhou et al (2012) started an empirical likelihood method under hypothesis testing. Jin (2016) discussed the semi parametric AFT model in case of right censored data.

Semi Parametric AFT Model

The linear model with the log-transformation is called the accelerated failure time model (AFT). Let T_i be the failure time for the i[th] patient, $i = 1,2, \dots n$. Due to censoring $C_i, T_i = \min(T_i, C_i)$ and $\delta_i = I\{T_i, C_i\}$ which takes value 1 if $T_i \leq C_i$ and 0 otherwise.

The semi-parametric AFT model is given by

$$\log(T_i) = X'_i\beta + \varepsilon_i$$

Where $X_i \; i = 1,2, \dots \dots, n$ are observed covariates, β is $px1$ unknown parameter vector and $\varepsilon_i \; i = 1,2, \dots, n$ are independent and unobserved errors and has a distribution form that does not involve X. It is also assumed that the mean of the ε_i, is not actually 0 i.e., $E(\varepsilon_i) = 0$. The semi-parametric AFT model is an efficient semi parametric linear regression model and its regression coefficient β directly estimates the impact of covariates X on the survival time rather than hazard rates.

It is presumed that the conditional vector X_i has the p $x1$ covariate for the i[th] subject, C_i is the censoring variable and $T_i \; i = 1,2, \dots \dots, n$ is the failure times which are independent, and $\varepsilon_i \; i = 1,2, \dots, n$ are independent and identically distributed random variables whose distribution function is totally unidentified. (Cox and Oakes, 1984; Kalbfleisch and Prentice, 2002; Lawless, 2003).

Suppose $X_i' = Z, \varepsilon_i$ follows exponential function with hazard function $h_0(t)$ then the hazard function for semi parametric AFT model is

$$h(t;x) = \exp(-Z'\beta)h_0(te^{-Z'\beta}) \qquad 0$$

And the survival function is

$$S(t;x) = \exp\{-\int_0^t \exp(-Z'\beta)\; h_0\left(ute^{-Z'\beta}\right)du\}$$

The density function is

$$f(t;x) = h(t;x)S(t;x)$$

The interpretation of the semi parametric AFT model is straightforward and the model specifies that the effect of the covariate is multiplicative on survival time t rather than the hazard function. There is a baseline hazard function $h_0(t)$ which applies when $Z = 0$. It is assumed that the role of Z is to accelerate (or decelerate) the time to failure. If the probability distribution of the error terms in the model is one of the well- known statistical distributions, then the AFT model is called parametric, otherwise, the AFT model is semi-parametric. The frequently used distributions for parametric AFT model are log-logistic, exponential, and Weibull etc. To estimate the regression parameters in a parametric AFT model, maximum likelihood estimation can be used, while the parameter estimates in a semi parametric AFT model can be obtained by using rank-based estimators.

2. OBJECTIVE OF THE STUDY

The main objectives of this paper are:

(i) to fit semi parametric AFT model by rank estimation method and

(ii) to study the effect of different factors on the survival of oesophagus cancer patients by using semi parametric AFT model.

3. RESULTS OF FITTING SEMI PARAMETRIC AFT MODEL

In this study, the semi parametric AFT model is estimated with the explanatory variables age, sex, cancer directed treatment, socio economic status, location and stage of the patients in case of esophagus cancer patients. The following table shows the results of this study.

Table 1: Fitting of Semi parametric AFT model

Variable	Coefficient	Standard error	Z value	P value	95% Confidence Interval	
(Intercept)	6.397	0.322	19.838	<2e-16	5.766	7.028
Age						
Less than 50	Reference					
50 to 70	0.031	0.179	0.172	0.863	-0.319	0.382
	-0.452	0.281	-1.607	0.108	-1.003	0.099

70 and above						
Sex Male						
Female	Reference					
	-0.087	0.182	-0.477	0.633	-0.444	0.269
Cancer treatment						
Surgery & others	Reference					
Other than Surgery	-0.270	0.200	-1.347	0.178	-0.662	0.122
	-0.582	0.282	-2.062	0.039	-1.135	-0.029
No treatment						
Location						
Rural Urban	Reference					
	-0.158	0.148	-1.071	0.284	-0.448	0.132
Socio Economic						
Lower	Reference					
Middle Higher	0.494	0.209	2.362	0.018	0.084	0.904
	0.892	0.324	2.754	0.006	0.257	1.527
Stage						
Localized	Reference					
Regional	-0.660	0.182	-3.620	<2e-16	-1.017	-0.303
Distant	-1.758	0.238	-7.393	<2e-16	-2.224	-1.292
Unknown	-1.187	0.237	-5.015	<2e-16	-1.652	-0.722

The table 1 shows the impact of various factors such as age, sex, location, cancer directed treatment, socio economic status and stage of the diseases of esophagus cancer patients by fitting semi parametric AFT model. In this study, it is seen that, in case of age of the patients, the coefficients of age groups 50 to 70 and 70 and above age group are 0.031 and -0.452 respectively with confidence interval (0.319 0.382) and (-1.003 0.099). In this study, there is no significant difference in survival among patients belonging to different age groups since the p values of these age groups are greater than 05. Again, in case of the factor sex of the patients, considering male patients as reference category and the coefficient of the female patients is -0.087 with 95%

confidence interval (-0.444 0.269) which is also not significant. Considering rural as reference category it is also found that urban patients having coefficient -0.158 with 95% confidence interval (0.448 0.132) is not also a significant factor in case esophagus cancer data.

But in case of the patients from socio economic status, it is seen that middle and higher socio-economic group of people have lower risk of dying than lower socio-economic people. The coefficients of middle and higher socio-economic group are 0.494 and 0.892 respectively with 95% confidence interval (0.084 0.904) and (0.257 1.527) respectively and both these categories are significant since their p values are0.018 and 0.006. Again, stage of the patients is also a responsible factor in case of this esophagus cancer data. From this study, it is seen that the coefficients of Regional, Distant, and unknown stage are -0.660, -1.758 and -1.187 respectively with 95% confidence interval (-1.017 -0.303) and (-2.224 -1.292) and (-1.652 -0.722) and they are significant since their p values are less than .05. It means stage of the patients is also a prominent factor for this study. Cancer directed treatment plays an important role in case of survival data. Considering surgery as a reference category it is found that the coefficients of other than surgery and the patients who do not take any treatment are -0.270 and -0.582 with 95% confidence interval (-0.662 0.122) and (-1.135 -0.029) respectively and it is also a significant factor in case of this survival data.

4. SUMMARY

The survival data of esophagus cancer patients of North East India have been taken into account for analysis in this chapter. To analyze this data semi-parametric AFT model is used and from this study it is observed that stage of the patients, socio economic status and cancer directed treatment have a significant role. With reference to lower socio-economic status patients, the middle and higher socio-economic patients have lower risk of dying. The stage of diagnosis of patients is also a responsible factor in case of survival of esophagus cancer patients. The probability of survival of a patient diagnosed in early stage is significantly higher than patients diagnosed in advance stages. The patients who undergo cancer directed treatment other than surgery has higher risk of dying than the patients who undergo the treatment of surgery and its combinations. From the observations it is found that, the age of the patients at the time of diagnosis has no significant impact on the esophagus cancer patients. Also, the patients belonging to both rural and urban areas are experiencing more or less similar risk of time. The sex of the patients is also not found to be a significant factor which can influence the survival.

References

[1] Cox, D.R. (1972). Regression Models and Life Tables, *Journal of the Royal Statistical Society*,Series B, (Methodological), 34 (2), 187-220.

[2] Cox, D.R. (1975). Partial Likelihood, *Biometrika*, 62(2), 269-276.

[3] Cox, D.R. & Oakes, D.(1984). Analysis of Survival Data, 1st edition, *Chapman and Hall*, London -NewYork.

[4] Jin, Z. (2007). M-estimation in Regression Models for Censored Data, *Journal of Statistical Planning and Inference*, 137, 3894-3903.

[5] Jin, Z. (2016). Semiparametric Accelerated Failure Time Model for the Analysis of Right Censored Data, *Communications for Statistical Applications and Methods*, 23 (6), 467–478.

[6] Jin, Z., & Ying, Z. (2004). Asymptotic Theory in Rank Estimation for AFT Model Under Fixed Censorship, Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis, and Quality of Life, *Springer*, 107-120.

[7] Kalbfleisch, J. D., & Prentice, R. L. (2002). The Statistical Analysis of Failure Time Data, 2nd edition, *John Wiley & Sons*, New-York.

[8] Lawless, J. F. (2003). Statistical Models And Methods For Lifetime Data Analysis,2nd edition, *Wiley*, New York.

[9] Stute, W. (1993). Consistent Estimation under Random Censorship when Covariables are Present, *Journal of Multivariate Analysis*, 45, 89-103.

[10] Stute, W. (1996). Distributional Convergence Under Random Censorship when Covariables are Present, *Journal of Statistics*, 23, 461- 471.

[11] Zhou, M (1992). M-estimation in Censored Linear Models, *Biometrika*, 79, 837-841.

[12] Zhou, M., Kim, M., & Bathke, A. (2012). Empirical Likelihood Analysis for the Heteroscedastic Accelerated Failure Time Model, *Statistica Sinica*, 22, 295-316.