Dhaka Univ. J. Sci. 58(2): 215-217, 2010 (July)

Application of Non-Parametric, Semi-Parametric and Parametric Survival Methods on Infant Mortality Data of Bangladesh

Muhammad Iftakhar Alam and M H M Imrul Kabir

Institute of Statistical Research and Training (ISRT), Dhaka University, Dhaka-1000, Bangladesh

Received on 12. 07. 2009. Accepted for Publication on 16. 06. 2010.

Abstract

The statistical analysis of lifetime or survival data has become a topic of considerable interest of statisticians. The field has expanded rapidly in recent years. In this study, we consider three different approaches to survival analysis. The Kaplan-Meier method as a non-parametric method, Cox's proportional hazards model as a semi-parametric method and accelerated failure time model as a parametric method have been used in this study to analyze the infant mortality data of Bangladesh Demographic Health Survey (BDHS) 2004 to assess the influence of several factors like sex of child, mother's age, mother's education, birth order and type of place of residence on infant survival. It has been found that all these factors have significant influence on infant mortality.

I. Introduction

Survival analysis deals with time until an event occurs. In the current case, the event is death. Survival data can take so many different forms such as censored, uncensored repeated events, multiple states, clustered etc. Censoring is a mechanism for which survival data are different from usual data. It occurs when one has some information about individual survival time, but one doesn't know the survival time exactly. One way to examine the relationship of explanatory variables to life time is through a regression model in which life time depends upon the explanatory variables. This involves specifying a model for the contribution of survival time given covariates. There are many approaches to regression analysis for survival time data. One employs parametric families of lifetime distributions and extends model such as the exponential, Weibull, and lognormal model to include explanatory variable. Another approach is distribution free and assumes less about the underlying distribution than do the parametric methods. One such model is proportional hazards model. A proportional hazard family is a class of model with the property that different individuals have hazard functions, which are proportional to one another. The Cox's proportional hazards model is usually in terms of the hazards model formula. This model gives an expression for the hazard at a specific time for an individual with a given specification of a set of explanatory variables. Cox and Oakes (1984) and Kalbfleisch and Prentice (1980) provided the details introduction to standard hazards model developed initially by Cox (1972). One can use a parametric model if the failure time follows a known parametric model. Thus if the distribution is unknown as is typically the case, the Cox's model will give reliable enough result. So that it is a safe choice of model and the user does not need to worry about whether the wrong parametric model is chosen.

II. Objectives of the Study

Determination of relative contribution of different risk factors on infant mortality and the comparison of different survival methods are the two main goals of this study.

III. Data and Variables

The data from Bangladesh Demographic and Health Survey (BDHS), 2004 have been used here. This study focused on children under one year of age called infant. Five explanatory variables for infant mortality have been selected for evaluation. These are:

- Mother's age at birth of the index child.
- Birth order of the index.
- Mother's education.
- Sex of the index child.
- Type of place of residence.

IV. Methodology

In this study, mortality is defined as the length of the interval between the birth of the child and the time of interview if the child is alive until the survey is conducted, otherwise the age at death is considered. Total subsequent mortality in this one-year period amounted to 1,412 (about 42 per cent of the total). The estimate of proportion of infant died subsequent to alive was obtained by product limit method. The estimate for median birth intervals were then obtain on the basis of P-L estimate for the survival function to take account of censoring (the case is alive). The advantage of using the P-L method is that censoring is taking into account in estimating the survivor functions. The P-L method is utilized in this study to take account of the issue of censoring in infant mortality.

The P-L method of the survival function may be defined as follows:

where, $$\hat{S}(t_i) = \prod_{j=1}^{i} \left(1 - \frac{d_j}{n_j}\right)$$

d_j is the number of infant died at time t_j.

n_j is the number of infant at risk of being died and t_j, denotes the time since the failure of a child.

Cox's Proportional Hazards Model

The Cox's proportional hazards model is the most popular model for describing the relationship between risk factors and survival time. This model has been employed here to explore the effects of risk factors on survival. In 1972 Cox proposed the most popular and flexible proportional hazards model where the hazard function at time t for an individual with covariate vector **x** is given as

$$h(t; \mathbf{x}) = h_0(t)\, e^{\boldsymbol{\beta}\mathbf{x}} \qquad (2.1)$$

where $h_0(t)$ is an arbitrary unspecified baseline hazard function for continuous T, i.e., $h_0(t)$ be the value of hazard

function with $\mathbf{x} = \mathbf{0}$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)$ be the regression co-efficient corresponding to the covariate vector $\mathbf{x}=(x_1, x_2, \ldots, x_p)$. This proportional hazards model is non-parametric in the sense that it involves an unspecified function in the form of an arbitrary baseline hazard function.

Accelerated Failure Time Regression Model

Accelerated failure time (AFT) model is a parametric model that provides an alternative to the commonly-used proportional hazards models. Whereas a proportional hazards model assumes that the effect of a covariate is to multiply the hazard by some constant, an AFT model assumes that the effect of a covariate is to multiply the predicted event time by some constant. AFT models can therefore be framed as linear models for the logarithm of the survival time. To be used in an AFT model, a distribution must have a parameterization that includes a scale parameter. The logarithm of the scale parameter is then modeled as a linear function of the covariates. Location-scale distribution has survival function of the form,

$$S(y,u,b) = S_0(\frac{y-u}{b});\ -\infty < y < \infty \qquad (2.2)$$

where $u\,(-\infty < u < \infty)$ is a location parameter, b>0 is a scale parameter, and $S_0()$ is a fully specified survivor function defined on $(-\infty, \infty)$. If T is a life time variable and Y=log T has the distribution (2.2) then it is to be said that T has a log-location distribution. The survivor function T may be written as

$$S^*(t;\alpha,\beta) = S_0(\frac{\log t - u}{b}) = S_0[(\frac{t}{\alpha})^{\beta}] \qquad (2.3)$$

where, $\alpha = \exp(u)$, $\beta = b^{-1}$ and for $0 < w < \infty$, $S_0{}^* = S_0(\log w)$. The weibull, loglogistic and lognormal distribution are all of this form; the corresponding location-scale parameter distributions of Y are the extreme value, logistic and normal respectively.

Log-location-scale distributions are the most widely used parametric lifetime models and regression models in which u (and sometimes b) in (2.2) are functions of covariates are of fundamental importance in both parametric and semi-parametric frameworks. All these methodology were implemented by using statistical software SPSS and R.

V. Analysis

Kaplan-Meier Estimate

In the first stage of the analysis, Kaplan-Meier (K-M) approach is used to estimate the proportion of survived infants.

Table. 1. Kaplan-Meier estimate of the probability of infant death by different demographic and socio-economic characteristics.

Time	Sex of Child		Mother's Age				Birth Order		Type of Place of Residence		Mother's Education			
	Boy	Girl	Under 20	20 to 30	30 to 40	More than 40	First child	Others	Rural	Urban	No edu-cation	Pri-mary	Sec-ondary	Higher
≥1	0.9	.92	.99	.95	.85	.84	.89	.92	.91	.92	.87	.92	.97	.99
≥2	.86	.89	.99	.92	0.8	.75	.86	.88	.87	.88	.81	.89	.96	.97
≥3	0.8	.83	.98	0.9	.71	.63	.79	.83	.81	.82	.74	.83	.94	.96
≥4	.76	.81	.98	.89	.66	.58	.75	0.8	.78	0.8	.69	0.8	.93	.95
≥5	.72	.78	.97	.87	.61	.52	.71	.77	.75	.76	.65	.76	.91	.94
≥6	.67	.73	.97	.84	.55	.42	.67	.71	0.7	.72	.59	.72	.88	.94
≥7	.64	.71	.96	.83	.52	.37	.64	.69	.67	.68	.56	.69	.87	.93
≥8	.61	.68	.96	.81	.48	.33	.61	.66	.64	.66	.52	.67	.85	.93
≥9	.58	.65	.96	0.8	.45	.27	.58	.63	.61	.63	.49	.64	.83	.93
≥10	.56	.62	.96	.79	.43	.23	.56	0.6	.59	.59	.46	.62	.82	.93
≥11	.54	0.6	.95	.77	.41	0.2	.53	.58	.57	.56	.43	.59	0.8	.93
≥12	.49	.53	.95	.75	.36	0.1	.47	.52	.51	.51	.35	.54	.79	.93

It reveals from the table 1 that as age increases probability of infant surviving decreases. The probability of surviving is more among girls than those of boys. Survival probability decreases with the increase in mother's age. Children of young mother exhibit higher survival probabilities. Children whose birth order is second or higher survive longer than children with first birth order. Children of urban area show higher survival probability than children of rural area. It is also seen that survival probabilities increase with the increase of level of mother's education

In the second stage of the analysis, Cox's proportional hazards model is used. The purpose of this model is to show how mortality depends upon different factors. The table 2 gives the coefficients of hazards ratio using Cox's proportional hazards model. The estimated hazard ratio gives how much a level of a covariate is at risk relative to the reference category.

Table. 2. The coefficients and P-values for different levels of covariates using Cox's proportional hazards model

Covariate	Levels	Estimated Hazard Ratio	P-value
Mother's age	20 to 30	7.7	0
	30 to 40	25.59	0
	>40	42.29	0
Sex of Child	Boy	1.14	$1.4e^{-02}$
Birth Order	≥2	0.44	0
Mother's Education	Primary	0.84	$6.2e^{-03}$
	Secondary	0.52	$1.8e^{-10}$
	Higher	0.13	$2.0e^{-08}$

The P-values indicate that all the covariates are highly significant. The hazard ratio 1.44 for boy means that the hazard rate for boy is 1.44 times of the girl. As mother's age increases hazard ratio decreases. It is also seen that advancement in mother's education decreases hazard of children. **Accelerated Failure Time Regression Model**

A technique closely related to P-P and Q-Q plot is used with parametric models for which the survivor function or distribution function can be "linearized". This means that some transform of $S(t;\theta)$ is a linear function of t or of some function of t, that is, $g_1[S(t;\theta)]$ is a linear function of $g_2(t)$ for some function g_1 and g_2.The idea is then plot $g_1(\hat{S}(t))$ versus $g_2(t)$; if the parametric family is appropriate the result should be roughly linear. It has been found that weibull distribution fits the data.

Table. 3. The coefficients and related p-values obtained from accelerated failure time model

Covariate	Levels	Coefficients	P-value
Mother's age	20 to 30	-1.68	$8.75\ e^{-19}$
	30 to 40	-2.66	$2.40\ e^{-43}$
	>40	-3.04	$1.68\ e^{-55}$
Sex of child	Boy	-0.12	$8.56\ e^{-03}$
Birth order	≥2	0.66	$6.42\ e^{-42}$
Type of place of residence	Urban	-0.04	$4.78\ e^{01}$
Mother's education	Primary	0.13	$8.60\ e^{-03}$
	Secondary	0.54	$4.24\ e^{-10}$
	Higher	1.64	$4.39\ e^{08}$

Categories for mother's age are highly significant. The negative values of the coefficients indicate survival probability of children decreases comparing to the children of mother's, aged less than 20 years. The survival probability of a boy is less than a girl. The coefficient value 0.66 means that the survival probability of the infants with birth order 2 or more is more than the infants with birth order 1. It is also seen that with the rise in level of mother's education, survival probability of infants also increases.

VI. Discussion and Conclusion

This study provides some empirical evidence for association between some selected explanatory variables and infant mortality. Among the five explanatory variables that are examined, the mother's education seems to have a very strong effect on survival status of the infants. Other explanatory variables, such as age of the mother, sex of the child, type of place of residence and birth order, also have significant influence on infant mortality.

Among the three methods, the non-parametric Kaplan-Meier method gives the comparison of survival probability within several groups or categories of a covariate. It is very much useful to assess which distribution will fit the data even checking the assumptions of the covariates for Cox's proportional hazards model. Its main problem is that it can't take more than one covariate at a time to analyze. On the other hand, it is easy and simple compared to other survival methods. It is also robust and closely approximate to parametric model. It gives regression coefficients, adjusted survival curves and hazards ratios without specifying the base-line hazard function. It gives hazards ratio which provides information that how much one category is in risk to failure than the other category. The parametric AFT model is the best of all if the data fit any distribution.

Acknowledgements

The authors are thankful to the referee for useful suggestions which helped in improving the paper.

---

1. Ahmed, N.R. (1981). Family size and sex preferences among women in rural Bangladesh, *Studies in Family Planning* **12(3)**, 100-109.
2. Bangladesh Bureau of Statistics (BBS) (1999). *Statistical Year Book* 1999. Bangladesh Bureau of Statistics, Ministry of Finance and Planning.
3. Cox,D.R and Oakes (1994). *Analysis of Survival Data*. Chapman and Hall, New York.
4. Hosmer, D.W. and J. S. Lemeshow, (1999). *Applied Survival Analysis*. John Willey and Sons, New York.
5. Kalbfleisch, J.D. and R. L. Prentice, (1980*).The Statistical Analysis of Failure Time Data*. John Willey and Sons, New York.
6. Lawless, J.F. (2003).*Statistical Methods for Lifetime Data*. John Willey and Sons, New York.
7. Lee, E.T. (1992). *Statistical Methods for Survival Data Analysis.* John Willey and Sons, New York.
8. Chakraborty, N., S. Sharmin and M. A. Islam (1996). Differential Pattern of Birth Intervals in Bangladesh*, Asia-Pacific Population Journal*, **11(4)**, 73-86.