ORIGINAL RESEARCH PAPER

Statistics

Volume : 6 | Issue : 12 | December : 2016 | ISSN - 2249-555X | IF : 3.919 | IC Value : 79.96

A STUDY ON COX REGRESSION PROCEDURES IN SURVIVAL ANALYSIS

KEYWORDS

cox regression, time-dependent covariates, time-varying coefficients, failure time, survival time, or event time.

Mr.M.Palanivel

Assistant Professor in statistics, Mahindra Arts and Science College, Kalippatti.

ABSTRACT *The term "survival analysis" pertains to a statistical approach designed to take into account the amount of time an experimental unit contributes to a study. That is, it is the study of time between entry into observation and a subsequent event. In logistic regression, we were interested in studying how risk factors were associated with presence or absence of disease. Sometimes, though, we are interested in how a risk factor or treatment affects time to disease or some other event. Or we may have study dropout, and therefore subjects who we are not sure if they had disease or not. In these cases, logistic regression is not appropriate. This effectively creates a time-varying coefficient that is easily estimated in software such as SAS and R. However, the usual programming statements for survival estimation are not directly applicable. Unique data manipulation and syntax is required, but is not well documented for either software. This paper covers a tutorial in survival estimation for the time-varying coefficient model, implemented in SAS and R. We provide a macro coxtvc to facilitate estimation in SAS where the current functionality is more limited. The macro is validated in simulated data and illustrated in an application.*

INTRODUCTION

The analysis consisted of following the subject until death. The uses in the survival analysis of today vary quite a bit. Applications now include time until onset of disease, time until stock market crash, time until equipment failure, time until earthquake, and so on. The best way to define such events is simply to realize that these events are a transition from one discrete state to another at an instantaneous moment in time. Of course, the term "instantaneous", which may be years, months, days, minutes, or seconds, is relative and has only the boundaries set by the researcher.

COX MODEL

A Cox model is a well-recognized statistical technique for exploring the relationship between the survival of a patient and several explanatory variables. Survival analysis is concerned with studying the time between entry to a study and a subsequent event (such as death). Censored survival times occur if the event of interest does not occur for a patient during the study period.

A Cox model provides an estimate of the treatment effect on survival after adjustment for other explanatory variables. It allows us to estimate the hazard (or risk) of death, or other event of interest, for individuals, given their prognostic variables. Even if the treatment groups are similar with respect to the variables known to effect survival, using the Cox model with these prognostic variables may produce a more precise estimate of the treatment effect (for example, by narrowing the confidence interval).

Interpreting a Cox model involves examining the coefficients for each explanatory variable. A positive regression coefficient for an explanatory variable means that the hazard is higher, and thus the prognosis worse, for higher values. Conversely, a negative regression coefficient implies a better prognosis for patients with higher values of that variable

Survival analysis is a branch of statistics which analysis of time to events, such as death in biological organisms and failure in mechanical systems. This topic called reliability theory or reliability analysis in engineering, and duration analysis or duration modeling in event history analysis in sociology. Survival analysis attempts to analysis the proportion of a population which will survive past a certain time. The Cox regression model (Cox, 1972) is the most popular method in regression analysis for censored survival data. However, due to the very high dimensional space of the predictors, the standard maximum Cox partial likelihood method cannot be applied directly to obtain the parameter estimates. To deal with the problem of co linearity, the most popular approach is to use the penalized partial likelihood which was proposed by Tibshirani (1995) and is called the least absolute shrinkage and selection operator (Lasso) estimation. In the case of biological survival, unambiguous, but for mechanical reliability, failure well-defined, for there may well be mechanical systems in which failure is partial, a matter of degree, or not otherwise localized in time. Even in biological problems, some events (for example, heart attack or other organ failure). More generally, survival analysis involves the modeling of time to event data; in this context, death or failure is considered an "event" in the survival analysis literature traditionally only a single event occurs for each subject, after which the organism or mechanism is dead or broken. The study of recurring events is relevant in systems reliability, and in many areas of social sciences and medical research. The survival function, also known as a survivor function or reliability function, is a property of any random variable that maps a set of events, usually associated with mortality or failure of some system, the term survival function is used in a broader range of applications, including human mortality.

COX'S REGRESSION MODEL:

This model one of important models published by (D.R. Cox in 1972) and is one of most frequently articles in statistics and medicine, which usually associated with mortality or failure of some system, he suggested that model depend on (hazard rate) in time (t), as the follows:

$$\lambda(t;z) = \lambda_0(t)\exp(\underline{\beta}'\underline{Z}) = \lambda_0(t)\exp\left(\sum_{i=1}^{p}\beta_i Z_i\right)$$

$\lambda_0(t)$: Initial hazard function when all values of $(\underline{Z} = 0)$.
$\underline{\beta}$: are unknown's regression coefficients.
$(\underline{Z})$: is the *p*-dimensional vector of covariates.

We can write survival function of (10) as follows:

$$S(t;\underline{Z}) = \{S_0(t)\}\exp\left(\sum_{i=1}^{p}\beta_i Z_i\right)$$

Where exp $(\sum_{i=1}^{p}\beta_i Z_i)$ is the proportional hazard function. But, Cox model is a semi- parametric model with free distribution. So, the estimation problem for $(\beta)$is the same under any transform. Only the rank statistic $r(.)$ can carry information about (β) when λ_0 is completely unknown. It follows that the rank statistic is marginally sufficient to estimate (β).To apply the rank statistic to get inferences about (β),one would use the marginal distribution of the ranks and the marginal likelihood.

Volume : 6 | Issue : 12 | December : 2016 | ISSN - 2249-555X | IF : 3.919 | IC Value : 79.96

EXPERIMENTAL PART

Description data:

In this research we use a real data of (Apollo - hospital) in Chennai city- India, we have (72) patients with Leukemia disease from (1-8-2015) to (31-11-2016). We study the following variables:

T: Survival Time
Z_1: The age of patient in first visit hospital
Z_2: The gender of patient takes (1 for male and 2 for female)
Z_3: Disease type takes follow:

1- Acute myeloid leukemia /AML
2- Acute Lymphocytic leukemia / ALL
3- Chronic mylelogenous leukemia/ CML
4- Chronic Lymphocytic leukemia/CLL

Z_4: Treatment type takes follow:
1- Biological Treatment
2- Chemical Treatment

Z_5: Address type takes follow:
1- Towner patient
2- Out of the town

Z_6: State of patient Anemia takes follow:
1- Patient with Anemia
2- Patient without Anemia

Z_7: State of Censored takes follow:
0- Patient with Censored
1- Patient exit

After analysis the data of patients in each four month of period study, we get the following results:

ESTIMATION COX REGRESSION PARAMETERS

After gets proportional Hazard assumption, now we can find the estimation parameters (Z1, Z2, Z3, Z4 Z5 Z6 with survival time (T) as in equation as follows:

The results of Cox- Regression model estimation

Variables	$\hat{\beta}$	S.E.	Wald	d.f.	Sig.
Age	0.000	0.007	0.001	1	0.981
Gender	0.118	0.311	0.145	1	0.703
Disease	0.123	0.183	0.452	1	0.501
Treatment	-.937	0.314	8.880	1	0.003
Address	0.331	0.273	1.469	1	0.226
Anemia	0.612	0.272	5.045	1	0.025

The results of (Backward method) for select sig.variables

Step No.	Variables in model	$\hat{\beta}$	S.E.	Wald	d.f.	Sig.
Step 1	Age	0.000	0.007	0.001	1	0.981
	Gender	0.118	0.311	0.145	1	0.703
	Disease	0.123	0.183	0.452	1	0.501
	Treatment	-.937	0.314	8.880	1	0.003
	Address	0.331	0.273	1.469	1	0.226
	Anemia	0.612	0.272	5.045	1	0.025
Step 2						
	Gender	0.116	0.288	0.161	1	0.688
	Disease	0.122	0.176	0.477	1	0.490
	Treatment	-.939	0.308	9.263	1	0.002
	Address	0.330	0.272	1.475	1	0.225
	Anemia	0.611	0.272	5.064	1	0.024
Step 3						
	Disease	0.109	0.175	0.386	1	0.534
	Treatment	-.938	0.309	9.189	1	0.002
	Address	0.329	0.272	1.465	1	0.226
	Anemia	0.584	0.263	4.930	1	0.026
Step 4						
	Treatment	-.957	0.310	9.508	1	0.002
	Address	0.292	0.265	1.215	1	0.270
	Anemia	0.606	0.260	5.423	1	0.020
Step 5						
	Treatment	-.967	0.310	9.747	1	0.002
	Anemia	0.622	0.261	5.703	1	0.017

Therefore we get only two significant variables with Cox model as follows:

$$Y = \lambda_0(t) \exp(-0.967Z_4 + 0.622Z_6) \text{ and}$$

$$Y = \text{Ln}\frac{\lambda(t/Z_1,Z_2,Z_3,Z_4,Z_5,Z_6)}{\lambda_0(t)} = -0.967Z_4 + 0.622Z_6$$

CONCLUSION

This real data example using Cox modeling will show what increased risk for hospitalization for an event of interest might look like graphically and in a risk of event type ratio. The study of recurring events is relevant in systems reliability, and in many areas of social sciences and medical research. The survival function, also known as a survivor function or reliability function, is a property of any random variable that maps a set of events, usually associated with mortality or failure of some system, the term survival function is used in a broader range of applications, including human mortality. From the results of Cox- reg. model we have only two sig. variables (Treatment and Anemia). 4- Most risk at survival time at (0- 20) days with probability survival (0.0039) and we find that the risk of death is increasing with time, that is mean the disease still continuo.

REFERENCE

1. Aalen, O. (1976) Nonparametric inference in connection with multiple decrement models, Scanadian Journal of Statistics, 3, 15-27.
2. Akritas, M.G. and van Keilegom, LV. (2003) Estimation of bivariate and marginal distributions with censored data, Journal of the Royal Statistical Society B, 65, 457-471.
3. Andersen, P.K., Abildstrom, S.Z. and Rosthoj, S. (2002) Competing risks as a multi-state model, Statistical Methods in Medical Research, 11,203-215.
4. Andersen, P.K., Borgan, 0., Gill and RD., Keiding, N. (1993) Statistical Models Based on Counting Processes, Springer-Verlag, New York.
5. Andersen, P.K. and Gill, RD. (1982) Cox's regression model for counting processes: a large sample study, The Annals of Statistics, 10, 1100-1120.
6. Antony, A.A. and Sankaran, P.G. (2005) Estimation of bivariate survivor function under masked causes of failure, Journal of Statistical Theory and Applications, 4, 401-423.
7. Arnold, B.c. and Strauss, D. (1988) Bivariate distributions with exponential conditionals, Journal of the American Statistical Association, 83, 522-527.
8. Arnold, B.C. and Zahedi, H. (1988) On multivariate mean remaining life functions, Journal of Multivariate Analysis, 25,1-9.
9. Bartholomew, DJ. (1957) A problem in life testing, Journal of the American Statistical Association, 52, 350-355.
10. Basu, A.P. (1971) Bivariate failure rate, Journal of the American Statistical Association, 66, 103-104.
11. Braekers, R and Veraverbeke, N. (2005) Cox's regression model under partially informative censoring, Communications in Statistics-Theory and Methods, 34, 1793-1811.
12. Buchanan, W.B. and Singpurwalla, N.D. (1977) Some stochastic characterization of multivariate survival, Theory and Applications of Reliability, 1, Eds. c.P. Tsokos and I.N.Shimi, Academic Press, New York, 329-348.
13. Burke, M.D. (1988) Estimation of a distribution function under random censorship, Biometrika, 75, 379-382.
14. Cai, J. and Douglas, E.S. (2004) Analysis of recurrent event data, Handbook of Statistics, 23, 603-623.
15. Cai, J. and Prentice, RL. (1995) Estimating equations for hazard ratio parameters based on correlated failure time data, Biometrika, 82, 151-164.
16. Campbell, G. and Foldes, A. (1982) Large sample properties of nonparametric bivariate estimators with censored data, Colloquia Mathemetica -Societatis, Janos Bolyai, 32, 103-121.
17. Chang. S.H. (2000) A two-sample comparison for multiple ordered event data, Biometrics, 56, 183-189.
18. Chang, S. H. and Wang, M. C. (1999) Conditional regression analysis for recurrence time data, Journal of the American Statistical Association, 94, 1221-1230.
19. Chen, Y.Q. and Cheng, S. (2005) Semiparametric regression analysis of mean residual life with censored survival data, Biometrika, 92, 19-29.
20. Chen, Y.Q. and Cheng, S. (2006) Linear life expectancy regression with censored data, Biometrika, 93, 303-313.
21. Chen, Y. Q., Wang, M.C. and Huang, Y. (2004) Semi parametric regression analysis on longitudinal pattern of recurrent gap times, Biostatistics, 5, 277-290.
22. Cheng, S.C., Fine, J.P. and Wei, L.1. (1998) Prediction of cumulative incidence function under the proportional hazards model, Biometrics, 54, 219-228.
23. Cox, D.R (1972) Regression models and life tables (with discussion), Journal of the Royal Statistical Society B, 34,187-220.