Pop 509: Survival Analysis

http://data.princeton.edu/pop509

Course Description

This half-course offered in the first half of the spring term focuses on the statistical analysis of time-to-event or survival data. We introduce the hazard and survival functions; censoring mechanisms, parametric and non-parametric estimation, and comparison of survival curves. We cover continuous and discrete-time regression models with emphasis on Cox's proportional hazards model and partial likelihood estimation. We discuss competing risk models, unobserved heterogeneity, and multivariate survival models including event history analysis. The course emphasizes basic concepts and techniques as well as applications in social science research using the statistical package Stata.
Prerequisite:WWS509 or equivalent.

Contents

The topics to be covered in each of the six weeks are as follows:

1. Parametric Survival Models

The hazard and survival functions in continuous time. Parametric forms and the distribution of log time. The exponential, Weibull, Gompertz, Gamma, Generalized Gamma, Coale-McNeil, and generalized F distributions. The U.S. life table.

Approaches to modelling the effects of covariates. Parametric families. Proportional hazards models (PH). Accelerated failure time models (AFT). The intersection of PH and AFT. Proportional odds models (PO). The intersection of PO and AFT. Recidivism in the U.S.

2. Non-Parametric Survival Models

One-sample estimation with censored data. The Kaplan-Meier estimator. Greenwood's formula. The Nelson-Aalen estimator. Expectation of life. Comparison of several groups: Mantel-Haenszel and the log-rank test.

Regression: Cox's model and partial likelihood. The score and information. The problem of ties. Tests of hypotheses. Time-varying covariates. Estimating the baseline survival. Martingale residuals.

3. Models for Discrete Data and Extensions

Cox's discrete logistic model and logistic regression. Modeling grouped continuous data and the complementary log-log transformation. Piece-wise constant hazards and Poisson regression.

Current status data versus retrospective data. Open intervals and time since last event. Backward recurrence times. Interval censoring.

4. Competing Risks

Modeling multiple causes of failure. Research questions of interest. Cause-specific hazards. Overall survival. Cause-specific densities. Estimation: one-sample and the generalized Kaplan-Meier and Nelson-Aalen estimators. The Incidence function.

Regression models. Weibull regression. Cox regression and the partial likelihood. Piece-wise exponential survival and multinomial logits. The identification problem. Multivariate and marginal survival. The Fine-Gray model.

5. Unobserved Heterogenity

Heterogeneity of frailty. Frailty distributions: gamma and inverse Gaussian. Subject-specific and population-average models. Heterogeneity's ruses.

Non-parametric estimation of the mixing distribution. Relaxing assumptions about the hazard. The identification problem. Heterogeneity and time-dependence.

6. Multivariate Survival

Kindred lifetimes and shared frailty. Models for husbands and wives. Sibling survival. Event-history models. The choice of time scale. Testing goodness of fit.

Bibliography

These are a few useful texts on survival analysis. The stars indicate the two texts that come closest to our coverage.

Andersen, P.K; O. Borgan, R. Gill, N. Keiding (1991). Statistical Models Based on Counting Processes. New York: Springer-Verlag.

Blossfeld, H-P; K. Golsch, G. Rohwer (2007). Event History Analysis with Stata. New Jersey: Lawrence Erlbaum Associates.

Box-Steffensmeier, J. and B. S. Jones (2004). Event History Modeling: A Guide for Social Scientists. Cambridge, England: Cambridge University Press.

*Cleves, M.; W. G. Gould, and J. Marchenko (2016). An Introduction to Survival Analysis Using Stata. Revised Third Edition. College Station, Texas: Stata Press. (Earlier editions in 2002, 2004, 2008 and 2010).

Cox, D. and D. Oakes (1984). Analysis of Survival Data. London: Chapman-Hall.

*Kalbfleisch J. D and R. L. Prentice (2002). The Statistical Analysis of Failure Time Data. Second Edition. New York: John Wiley. (First Edition 1980).

Royston, P. and P. C. Lambert (2011). Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model. College Station, Texas: Stata Press.

Singer, J.D and J. B. Willett (2003) Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford, England: Oxford University Press.

Therneau, T. M. and P. M. Grambsch (2000). Modeling Survival Data: Extending the Cox Model. New York:Springer.