台北大學 計系 學士班
存活分析 (Survival Analysis)

林建甫 (Jeff Lin, PhD.)
台北大學統計系助理教授
e-mail: cflin@mail.ntpu.edu.tw
http://web.ntpu.edu.tw/~cflin

Fall, 2006

- Time: Wednesday, 9:10pm – 12:00noon
- Classroom: 簡學大樓 7F26 與 統計系電腦室
- Course Web Page: from http://web.ntpu.edu.tw/~cflin
- 助教: 古惠瑜 gucl@alumni.nccu.edu.tw, 澎湘婷 chirr_peng@yahoo.com.tw
- Office Hours:
  - Tuesday, 三峽, 12:00noon–1:00pm (make appointment by e-mail first)
  - Wednesday, 三峽, 12:00noon–1:00pm (make appointment by e-mail first)
  - Thursday. 台北, 4:00pm–5:00pm
- Textbook:
  - Lecture Notes: 林建甫 (2006), 存活分析, course web page
  - Lecture Slides: course web page
- Software: SAS and R
- Evaluation
  - Assignments
  - Two Exams
1. What Types of Data to Study?

   (a) Births to Deaths (Medicine, Epidemiology, Demography)
   (b) Diagnosis of Disease(s) (Cancer, Stroke) to Death (Medicine, Epidemiology)
   (c) Repair and Replacement in Mechanical System (Engineering, Reliability)
   (d) Marriages to Divorces, Re-marriages to Re-divorces (Sociology, Event History Analysis)
   (e) Immigration and Migration (Politics, Sociology)
   (f) Commit Crimes, Arrests, Probation, and Convictions (Law and Criminal Justice)
   (g) Job Changes, Loss, and Promotions (Business, Sociology)
   (h) Business Closing, Bankruptcies (Business, Tobit Model)
   (i) Customer Loyalty (Business, Data Mining)

2. Survival Analysis

   Survival analysis is the study of the distribution of life times, and is a loosely defined statistical term that encompasses variety of statistical technique for analyzing positive-valued random variables. Typically, the value of the random variable is the times from an initiating event to some terminal event, i.e. from time of birth (start of treatment) to death(relapse). Examples of this time-to-event data arise in diverse field, such as survival rate in medicine, mortality in public health, life table in epidemiology, vital statistics in actuarial science and demography, reliability in engineering, event history analysis in social science, queue process in business, unemployment in economics.

   You probably come up against two intractable difficulties when using ordinary regression or logistic regression:

   (a) Censoring: Nearly every data set contains some cases that do not experience the event. If the dependent variable is the time of the event, what do you do with these cases?
   (b) Time-dependent covariates: Some explanatory variables (like blood pressure value) change over time during the study. How do you handle these variables in a regression analysis?

   Ignoring these questions can lead to severe biases. Survival analysis are designed to deal with censoring and time-dependent covariates with statistically “reasonable” methods. Originally developed by biostatisticians, survival analysis have become popular in sociology, demography, psychology, economics, engineering, political science, and marketing.

3. Course Outlines

   (a) Introduction to Survival Data
   (b) Hazard Function, Survival Function
   (c) Censoring and Truncation
   (d) Nonparametric Methods for One Sample Survival Curve
   (e) Nonparametric Tests for Comparison of Two or More Samples Survival Curves
   (f) Semiparametric Models: Cox Proportional Hazard Models
   (g) Semiparametric Models: Time-Dependent Covariates
   (h) Semiparametric Models: Diagnosis
   (i) Parametric Models: Failure Time Distribution and Accelerated Failure Time Models
References