UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

PhD School in Statistics – XXV cycle

Short Course

ANALYSIS OF SURVIVAL DATA

by

Prof. Donald Alan Pierce

Oregon Health & Science University, USA

Calendar

Thursday	May 20/2010	10.00 - 13.00	
Thursday	May 27/2010	10.30 - 12.30 14.30 - 16.30	Aula Uggè
Thursday	June 03/2010	10.00 - 13.00	

Prof. Alessandra Salvan
School Director

http://www.stat.unipd.it/phd/courses_2010

via C. Battisti, 241 - 35121 Padova, Italia - tel 0498274168 – fax 0498274170

Analysis of Survival Data

Short course (3 2/4-hour sessions)

by Donald Pierce

This course will presume some prior basic knowledge of survival analysis, along lines of the regular Padua course whose outline was provided to me. An emphasis will be that analysis of *rates* is not quite the same as analysis of *lifetimes*. In particular, this distinction becomes important when the time variable is age, and subjects are followed up for only a small part of their life. This is common in medical studies. Attention will be given to issues that have arisen in the analysis of excess cancer among the atomic bomb survivors, these being of general interest for any dose response study. In less depth, practical issues arising from other real datasets will be emphasized, including real-time analyses during lectures using the software package STATA. Students will be required to make considerable use of this software. Although I aim further to indicate briefly why martingale theory has been so useful, introducing the basic notation for that, the course is intended to be quite practical and to complement what the students have already learned.

1. Basic notions

a. Survival times and rates

b. Censoring and delayed entry
   - i. Choice of time scale, connections with delayed entry
   - ii. Survival data specifications in STATA

c. Nonparametric (distribution) estimation
   - i. Counting process notation
   - ii. Kaplan-Meier and Nelson-Aalen: *exactly* the same estimator
   - iii. Martingale approach to variance of these
   - iv. Difficulties arising from delayed entry
   - v. Advantages of inference about rates rather than survival times

2. Relative risk (rate) regression

a. Combining nonparametric estimation and regression modeling: Cox (semiparametric) regression, partial likelihood

b. Using Poisson regression for time-grouped data (in STATA)

c. Some advantages of grouping on time: choice of time scale, and flexible modeling of baseline rates

d. Likelihood-based, score-based, and MLE-based inference: logrank test for comparing distributions

3. Examples pertaining to above issues

a. Study on diabetics: choice of time scale, dealing with delayed entry, use of grouped data

b. Atomic bomb survivor data: attempt to use STATA for this, additional needs, software Epicure developed for such needs

4. Further issues in semiparametric relative risk regression

a. Time dependent covariables: dealing with in STATA

b. Model checking: time dependent covariables, graphical methods, definitions of residuals

c. A-Bomb survivor data, Stanford Heart Transplant Data

d. Role of martingale theory: distinctions between full and partial likelihood

e. Connections of partial likelihood and rank methods

5. Fully parametric models

a. Likelihood function for censored and truncated data

b. Accelerated failure times and proportional hazards

c. Fitting parametric models in STATA