UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

UNIVERSITÀ DEGLI STUDI DI PADOVA

PhD SCHOOL IN STATISTICS
Department of Statistical Sciences

Specialist Course

ANALYSIS OF SURVIVAL DATA

Donald Alan Pierce

Oregon Health & Science University, USA

Aula Cucconi

June 17th – June 24th, 2013

Schedule

Monday	June 17	14.00 – 17.00
Tuesday	June 18	10.00 – 13.00
Friday	June 21	10.00 – 12.00
Monday	June 24	10.00 – 12.00 14.00 – 17.00

http://www.stat.unipd.it/phd/courses_2013

Topics in Survival Analysis

Short course (10 hours lecture plus student homework presentations)

by Donald Pierce

This course will presume some prior basic knowledge of survival analysis, along lines of the regular Padua "reliability" course whose outline was provided to me. An emphasis will be that analysis of *rates* can be quite different from analysis of *lifetimes*. For example, 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 and actuarial work. Practical issues arising with real datasets for medical studies will be emphasized, including real-time analyses during lectures using the software package STATA. Students will be asked to make considerable use of STATA for homework, but anyone objecting to this can probably do the same things in R: Survival. Issues arising in quantitative dose-response studies will be addressed in terms of analysis of radiation-related cancer among the A-bomb survivors. 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. Analysis of prognostic indicators for malignant melanoma, and of screening effects for cervical cancer.
    c. Dealing with non-proportional hazards: gastric cancer data.
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