CS&SS 544 Event History Analysis –Spring 2009

Scope: Event history analysis has become an important analytical tool in many fields of the social sciences. This course covers applied event history analysis. We will examine the standard tools used in the field—things like life tables, Kaplan Meier estimates, Cox proportional hazards model, and parametric survival models. Additionally, we will build a tool kit for developing custom models that involve "non-standard" methods like subgroup heterogeneity, incorporation of "immune" individuals, mixture models, models for clustered observations, multi-state models and social diffusion models. This course is not specific to any field within the social sciences, although many of the examples in this course are taken from demography.

Objective: (1) provide you with tools and concepts for solving quantitative problems involving the statistical analysis of time to events; (2) provide a took kit for developing custom event history models, (3) provide sufficient historical, intellectual, and mathematical background so that you can evaluate contemporary research using event history methods.

Instructor: Darryl Holman, 418 Denny Hall, 206-543-7586, djholman@u.washington.edu

Course web page: http://faculty.washington.edu/~djholman/csss544

Classes: Tuesday and Thursdays at 1:30-3:20 in 217 Denny Hall.

Office hours: I will usually be available after class for office hours. Other times can be arranged. Feel free to use email (djholman@u.washington.edu) to contact me with questions or to set up an appointment.

Textbook and readings: The textbooks are

- Allison PD (1995) *Survival Analysis Using the SAS System: A Practical Guide*. Cary, NC: SAS Institute Inc.
- Box-Steffensmeier JM, Jones BS (2004) *Event History Modeling: A Guide for Social Scientists.* Cambridge: Cambridge University Press.

Additional readings and handouts will supplement the text. These readings will illustrate principles discussed in lecture and the text, and will also be used as the basis for some class discussions. Readings will be available at https://djholman.csde.washington.edu/csss544/readings.

Grades: There will be 5 problem sets (12% each) that will make up 60% of your final grade, and a final project (40%). There are no exams for this course.

Problem sets: The five problem sets will consist of analytical exercises and other short problems. Frequently, the problems will require the use of computer software. I recommend that you get an account on the CSDE Windows network. The CSDE systems have many useful programs for doing event history analysis (request a Windows account here: http://www.csde.washington.edu/services/computing/access.shtml). Data sets for this course will be available on both the course web site and the CSDE server. You can use books, readings, notes, and web pages to help you work on the problems. In fact, you can work in groups on most exercises. **Grades for late problem sets will depreciate by 10% per day**, including any fraction of a day late.

Software: You can use any software that works for you and gets the job done. For example, when we work with the Cox proportional hazards regression model, almost any standard statistical software will work. For other assignments only a few "packages" will be able to easily perform the analysis. One option, and one I encourage, is that you begin working with a statistical programming language. Perhaps the best overall statistical programming language is *R*. However, the language *mle* written by your instructor is a good choice as well for advanced modeling. If there is sufficient interest, I will offer optional weekly sessions in a computer lab that introduces *mle* programming. There are a number of short courses and online tutorials that introduce *R* (or *S-plus*).

The *mle* package is freely available from http://mlelabs.com for use on your Windows or Linux computer. Extensive documentation is available online. You can download a pdf version of the documents for browsing or printing. The *mle* program is also installed on the CSDE terminal servers. Most of the exercises that you can do in *mle* can also be done in other statistical programming languages (*S-plus, R, Matlab, Gauss, Octave*). You are free to use any of these for your work under the idea that learning one such language will help you understand any other.

Projects: 40% of your course grade will be based on a project. This project can take one of several forms: (1) a new research proposal in preliminary form incorporating one or more of the methods covered in the course; (2) a completed existing research or funding proposal (a proposal started in another course, for example) that is revised to include one or more of the methods covered in this course; (3) a new manuscript in which you have applied methods covered in this course; (4) a term paper in which a dataset has been analyzed using methods from this course, (5) a completed poster presentation with original research, analysis, and presentation of an event history project. For options 3, 4, and 5, you can do a team project with another person in the course.

Schedule

Week 1: Introduction to Event History Analysis (Mar 31, Apr 2)

Topics: Course stuff, introduction, event data, probability, distributions.

Readings: Box-Steffensmeier and Jones (BSJ) Ch 1, 2; Allison Ch 1; Lecture 1 Notes; Notes on Writing an Event History Analysis Paper (Tuma).

Week 2: Parametric Survival Models (Apr 7, Apr 9)

Topics: Likelihood estimation, complete data, right censoring.

Readings: BSJ Ch 3; Allison Ch 2; Distributions Handout; Likelihood Handout.

Problem set 1 assigned (*Tuesday*)

Week 3: More Parametric Survival Models (Apr 14, 16)

Topics: Interval censoring, truncation, accelerated failure time models, hazards models.

Readings: BSJ Ch 3; Allison Ch 4. Messy Data Handout; Covariates Handout.

Problem set 1 due and problem set 2 distributed (*Tuesday*).

Week 4: Empirical and seemingly empirical models (Apr 21, 23)

Topics: Life tables, Kaplin-Meier estimation, piecewise models.

Readings: Gehan (1969); Blossfeld and Rohwer Ch 3; Allsion Ch 3.

Problem set 2 due, and problem set 3 distrubuted (*Tuesday*).

Week 5: Cox Proportional Hazards Models (Apr 28, 30)

Topics: The Cox model, partial likelihood, interpretation of estimates.

Readings: BSJ Ch 4; Allison Ch 5.

Problem set 3 due (Tuesday) and problem set 4 distributed (*Thursday*)

Week 6: More Cox Model and Piecewise Models (May 5, 7)

Topics: Time varying covariates, the proportionality assumption, more piecewise models.

Readings: BSJ Ch 4; Allison Ch 7.

Problem set 4 due (*Thursday*).

Week 7: Model Selection and Diagnostics (May 12, 14)

Topics: Uncertainty in estimates, computational approaches, information criteria, residuals.

Readings: BSJ Ch 6, 8; Wood et al. (1994).

Week 8: Models of Mixed Populations (Unobserved Heterogeneity I) (May 19, 21)

Topics: Clusters, immune fraction, mixtures of subpopulations.

Readings: Holman (2004).

Problem set 5 distributed (*Thursday*).

Week 9: Continuous Unobserved Heterogeneity (May 26, 28)

Topics: Heterogeneity, frailty models.

Readings: BSJ Ch 9; Vaupel and Yashin (1985).

Week 10 Some Advanced Models (Jun 2, Jun 4)

Topics: Repeated events, competing risks, latent events, social diffusion models.

Readings: Wood et al. (1994); BSJ Ch 10; Allison Ch 6, 8; Strang and Tuma (1993)

Problem set 5 due (*Tuesday*).

Final Projects Due: Friday, Jun 12 by 2:30 pm.

References

[**]Allison PD (1984) *Event history analysis: Regression for longitudinal event data.* Newbury Park, CA: Sage Publications.

[***]Allison PD (1995) *Survival Analysis Using the SAS System: A Practical Guide*. Cary, NC: SAS Institute Inc.

Blossfeld H-P, Hamerle A, Mayer KU (1989). *Event history analysis.* Hillsdale, New Jersey: Lawrence Erlbaum.

[*]Blossfeld H-P, Rohwer G(1995). *Techniques of Event History Modeling.* Mahwah, NJ: Lawrence Erlbaum.

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

[§]Edwards AWF (1972) *Likelihood*. Cambridge: Cambridge University Press.

[**]Elandt-Johnson RC, Johnson NL (1980) *Survival Models and Data Analysis*. New York: John Wiley and Sons.

[§]Evans M, Hastings N, Peacock B (2000) *Statistical Distributions*. Third edition. New York: John Wiley and Sons.

Gehan EA (1969) Estimating survival functions from the life table. *Journal of Chronic Diseases* **13**:629-644.

Holman DJ (2003) *mle: A programming language for building likelihood models.* Version 2.1. Volume 1. User's Manual. and Volume 2, Reference Manual. http://faculty.washington.edu/~djholman/mle.

Holman DJ (2003) Unobserved heterogeneity. In: Lewis-Beck MS, Bryman A, Liao TF, (eds.) *Encyclopedia of Social Science Research Methods*. Thousand Oaks, CA: Sage Publications.

Hosmer DW, Lemeshow S (1999) *Applied Survival Analysis: Regression Modeling of Time to Event Data*. New York: John Wiley and Sons.

Kalbfleisch JD, Prentice RL (1980) *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons.

[*]Klein JP, Moeschberger ML (1997) *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer-Verlag.

London D (1997) *Survival Models*. Winsted, CT:ACTEX Publicatons.

Mayer KU, Tuma NB (eds.) (1990) *Event History Analysis in Life Course Research*. Madison:Univ. Wisconsin Press.

[§]Namboodiri K, Suchindran CM (1987) *Life Table Techniques and their Applications*. Orlando: Academic Press.

Nelson W (1982) *Applied Life Data Analysis*. New York: John Wiley and Sons.

[***]Box-Steffensmeier JM, Jones BS (2004) *Event History Modeling: A Guide for Social Scientists.* Cambridge: Cambridge University Press.

Schoen R (1988). *Modeling Multigroup Populations.* New York: Plenum Press.

Strang D, Tuma NB (1993) Spatial and Temporal Heterogeneity in Diffusion, *American Journal of Sociology* **99**(3):614-639.

Trussell J, Hankinson R, Tilton J (eds.) (1992) *Demographic Applications of Event Histroy Analysis*. Oxford:Clarendon Press

Tuma NB, Hannan MT (1984) *Social Dynamics: Models and Methods.* New York: Academic Press.

Vaupel JW, Yashin AI (1985) Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *American Statistician* **39**:176-85.

Vermunt JK (1997) *Log-linear models for event histories*, London: Sage Publications.

*Lee ET (1992) *Statistical Methods for Survival Data Analysis* (2nd edition). New York: John Wiley & Sons.

Wood JW, Holman DJ, Yashin A, Peterson RJ, Weinstein M, Chang M-c (1994) A multistate model of fecundability and sterility. *Demography* **31**(3):403-426.

*Yamaguchi K (1991) *Event History Analysis*. Newbury Park, CA: Sage Publications, Inc.

* Recommended book for you collection
** Recommended book for this course (optional)
*** Required book for this course
§ Good technical reference.