

**UK      Fall 2013 Course Description**  
**STA635      Survivability and life testing (3 credits)**

---

Prereq: STA606 or STA525

---

We assume students already had a graduate level statistical inference course, like Sta606/525. In particular, you should know exponential distribution and maximum likelihood method.

We will begin with a review of the exponential distributions, and discuss some new property/usage of exponential distributions. In particular we will introduce the concept of hazard function and study the piecewise exponential distributions.

After that, we introduce the concept of censoring and cover various types of censored/truncated data, and discuss the assumptions in the sampling process, the change in the likelihood function and the difficulty they bring to statistical analysis. We cover the parametric inference procedures along with the SAS `proc lifereg`. Or R function `survreg()`. The main approaches are likelihood based: maximum likelihood estimator, likelihood ratio tests etc. We shall cover Weibull/extreme value regression models in some detail.

We next turn to nonparametric and semiparametric models and the statistical inference procedures for those models. Among the topics we will cover are:

- the Nelson-Aalen estimator, the Kaplan-Meier (product limit) estimator (see Java applets at <http://www.ms.uky.edu/~mai/java/AppletIndex.html>) and other related estimators based on them. SAS `proc lifetest`, R function `survfit()`. Greenwood formula, confidence intervals. R packages `emplik`

- log-rank test, weighted log-rank tests. Wilcoxon test. Variance estimate of the tests. One/two/K sample tests. Proportional hazards (Lehmann) alternative. Power calculations. Video game playing analogy of the tests. R function `survdifff()`.

- Cox models, partial likelihood analysis. Time dependent covariates. SAS `proc phreg`. R function `coxph()`. Crazy clock interpretation of the baseline hazard.

I will try to present a balanced approach which shall be interesting to both application minded audience and those looking for research problems. Some real data sets will be analyzed using the SAS software and/or the free software R. Counting process notation will be used (but no theory). Re-sampling method will be introduced. (but no Bayes).

Most topics cited above can be found in our textbooks *Survival Analysis Using The SAS System: a practical guide* by P. Allison (published by SAS Institute, 1995). I will supplement the text with handouts that supply additional details, computing programs, etc.

There will be homeworks (50%), an in-class midterm (25%) and a take-home final (25%),

**Time and Place:**

**Instructor:** Mai Zhou; Office: 343 MDS; Tel.: 257-6912;  
mai@ms.uky.edu;      <http://www.ms.uky.edu/~mai/courses.html>

**Office hours:**

There are a few other books that you may find interesting:

0) Cox and Oakes; Analysis of Survival Data. 1984 (Chapman & Hall) (very well written, if you like the style...)

1) Miller, Survival Analysis. Wiley. (1981) (these are the lecture notes, old but readable....)

2) Kalbfleisch & Prentice, The Statistical analysis of failure time data. Wiley. (1981) second edition (2002) (very good book, ...)

3) Lawless, Statistical Models & Methods for Lifetime Data. Wiley. (good coverage for parametric method ....)

4) Lee, Statistical methods for survival data analysis. 2nd Ed. (1992)

5) Fleming and Harrington, Counting Processes and Survival ... (if you want to learn the counting process martingale technique....)

6) Andersen et.al. Statistical Models Based on Counting Process ... (similar to 5, with many examples....)

7) Kleinbaum, Survival Analysis: A self-learning text 1996 Springer (at a lower level and with less material compared to 8...)

8) Klein & Moeschberger, Survival Analysis 1997 Springer (with many data sets ....)

9) Collett, Modeling Survival Data in Medical Research. 1994 Chapman & Hall

10) Allison, Survival Analysis Using The SAS System, ....

11) Marubini & Valsecchi (1995). Analyzing Survival Data from Clinical Trials and Observational Studies, Wiley.

12) Cantor, Extending SAS survival analysis techniques for medical research (1998) (has a few sas macros...)

13) Applied Survival Analysis: regression modeling of time to event data (1999) Hosmer and Lemeshow (Wiley)

14) Smith, Analysis of failure and survival data. Chapman & Hall 2002 (a text book. Use Splus and minitab. Cover Buckley-James accelerated regression model in detail....)

15) Therneau and Grambsch, Modeling Survival Data: Extending the Cox model. (more advanced compare to 8)

Earlier Tech Report. It is free and covers many topics same as in the book

<http://mayoresearch.mayo.edu/mayo/research/biostat/upload/58.pdf> (1996)

<http://mayoresearch.mayo.edu/mayo/research/biostat/upload/survival.pdf> (1999)

- Review of the exponential distribution, properties.
- Definition of hazard, cumulative hazard function and its connection with density function, CDF. [continuous, discrete and mixed case]
- The piecewise exponential distribution. [It is a continuous distribution.]
- The concept of right censoring and various types of censored/truncated data.
- The assumption of independent censoring in the sampling process, identifiability.
- The likelihood function for censored data.

The parametric inference procedures along with the SAS `proc lifereg`. R function `survreg()`. The main approaches are likelihood based: maximum likelihood estimator, likelihood ratio tests etc. We shall cover Weibull/extreme value regression models in some detail.

- Example of EM algorithm in computing MLE.

We next turn to nonparametric and semiparametric models and the statistical inference procedures for those models.

- Overview of nonparametric methods.
- the Nelson-Aalen estimator, the Kaplan-Meier (product limit) estimator (see Java applets at <http://www.ms.uky.edu/~mai/java/AppletIndex.html>) and other related estimators based on them. SAS `proc lifetest`, R function `survfit()`. Greenwood formula, confidence intervals. R package `emplik`

- log-rank test, weighted log-rank tests. Wilcoxon test. Variance estimate of the tests. One/two/K sample tests. Proportional hazards (Lehmann) alternative. Power calculations. Video game playing analogy of the tests. R function `survdifff()`.

- Cox models, partial likelihood analysis. Time dependent covariates. SAS `proc phreg`. R function `coxph()`. Crazy clock interpretation of the baseline hazard.

Your Name:

Your Major:

Is there a particular statistical procedure you want to learn most?

Have you used any statistical software before? [what is it?]

**General Methodology:** estimation, testing, modeling. parametric/nonparametric. Likelihood theory, Bayesian/non-Bayesian.

**Statistical Model:** time series modeling. linear/nonlinear regression model, logistic regression, mixed model, GLM.

**Special type of data:** binary/categorical type, censored type, spacial data, genetic/gene data. large data, image data, longitudinal type. time series data.

**Design of experiments:** (from Fisher to sequential design, re-randomize), case-control, case-cohort.

**Application oriented:** telephone survey methodology. financial statistics, etc

Special method: sequential method

Computational statistics:

Decision theory: admissible, etc.

Exact distribution theory. Approximate distribution theory.