Time-to-event Data Analysis Project Guidelines

General Information

- The objective of this project is to analyze a time-to-event dataset, write a report, and give a short (15 minutes to 20 minutes) oral presentation of the result.
- Students may work alone or in groups of size 2.
- The project will be done in stages, so that I can provide feedback along the way. The due dates are given below.

Stage	Due date
Milestone 1	Wednesday, October 28, 2009
Milestone 2	Wednesday, November 11, 2009
Milestone 3	Friday, December 04, 2009

- Milestones
  - Milestone 1: Description of the dataset, including copies of the pages from the reference source, if it comes from something previously published (including on the web).

    The description does not need to be very long, half page is enough. Only one copy of the description is needed for each team, with both team members' names on it, if you decide to work with someone else.
  - Milestone 2: Preliminary analysis of the data.

    You only need to submit some initial graphical analysis results. Submit one or two pages of graphical results is enough. For example, the plots of the Kaplan-Meier curves.
  - Milestone 3: Final project report and presentation.

Data Analysis

This is somewhat open-ended. At a minimum, you should have the estimates for the survival curves, do some hypothesis testing, fit a Cox proportional hazard model, and do some model diagnostics.

I can give you some suggestions once I have looked at your dataset.

Written Report

The written description should be from 5 pages to 10 pages (this is not a constraint, but an expectation). Include all interesting and relevant graphics and tables. Describe the purpose of the study, how the data were collected, which model and methods were used, how the analysis proceeded, the results of the analysis, and conclusions drawn. State necessary assumptions needed for conclusions and indicate which can and cannot be checked by looking at the data.

The report should be done in LaTeX or MS Word and should be supplemented with graphs or tables (extracted from computer output or done by hand). These should be integrated neatly into the report. Do *NOT* hand in raw computer output. You may attach your computing codes in the appendix.

The report is due before the presentation on December 04, 2009. The assessment of the project is mainly based on the report.

Oral Presentation

The presentation will be held on Friday, December 04, 2009 during the class. Each team will have about 15-20 mins, depending on how many teams we will have. If there are two members in a team, each member is expected to present part of the results.

How to Find a Time-to-event Dataset

- Find a real time-to-event dataset. Dataset from any area is fine. Your dataset should have the following characteristics:
  - Time to event should be the (possibly censored) response. It would be nice to have some censoring, but this is not a requirement.

  - There should be ***at least one*** explanatory variable. It is better to have two or more explanatory variables.
  - The dataset does not have to be large, but try to find one with at least 10 events.

- Resources:
  - If you have access to real data from your consulting projects, they may be used for the project.
  - Journal articles provide the some interesting datasets.
  - Some web-sites archives datasets.
  - Google!