Survival Analysis

Introduction to Survival Analysis

Survival analysis models factors that influence the time to an event. Ordinary least squares regression methods fall short because the time to event is typically not normally distributed, and the model cannot handle censoring, very common in survival data, without modification.

Non Parametric Methods for Survival Analysis

Nonparametric methods provide simple and quick looks at the survival experience, and the Cox proportional hazards regression model remains the dominant analysis method.

Note: A number of sub-sections are titled Background. These provide some statistical background for survival analysis for the interested reader (and for the author of the seminar!). Provided the reader has some background in surival analysis, these sections are not necessary to understand how to run survival analysis in SAS. These may be either removed or expanded in the future.

Note: The terms event and failure are used interchangeably in this seminar, as are time to event and failure time.

Censoring

Types of Censoring

Left censoring a data point is below a certain value but it is unknown by how much.

Interval censoring a data point is somewhere on an interval between two values.

Right censoring a data point is above a certain value but it is unknown by how much.

Type I and II Censoring

Type I censoring occurs if an experiment has a set number of subjects or items and stops the experiment at a predetermined time, at which point any subjects remaining are right-censored.

Type II censoring occurs if an experiment has a set number of subjects or items and stops the experiment when a predetermined number are observed to have failed; the remaining subjects are then right-censored.

- Random (or non-informative) censoring is when each subject has a censoring time that is statistically independent of their failure time. The observed value is the minimum of the censoring and failure times; subjects whose failure time is greater than their censoring time are right-censored.
- Interval censoring can occur when observing a value requires follow-ups or inspections. Left and right censoring are special cases of interval censoring, with the beginning of the interval at zero or the end at infinity, respectively.
- Estimation methods for using left-censored data vary, and not all methods of estimation may be applicable to, or the most reliable, for all data sets.

3.2. Nelson-Aalen estimator of the cumulative hazard function

The NelsonAalen estimator is a non-parametric estimator of the **cumulative hazard rate function** in case of censored data or incomplete data.

It is used in survival theory, reliability engineering and life insurance to estimate the cumulative number of expected events.

An "event" can be the failure of a non-repairable component, the death of a human being, or any occurrence for which the experimental unit remains in the "failed" state (e.g., death) from the point at which it changed on.

Formula

The Nelson-Aalen estimator is a non-parametric estimator of the cumulative hazard function and is given by:

$$\tilde{H}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i},$$

with d_i the number of events at t_i and n_i the total individuals at risk at t_i.(where di is the number who failed out of ni at risk in interval ti.)

The curvature of the NelsonAalen estimator gives an idea of the hazard rate shape. A concave shape is an indicator for infant mortality while a convex shape indicates wear out mortality. It can be used for example when testing the homogeneity of Poisson processes. Because of its simple relationship with the survival function, $S(t) = e^{H(t)}$ the cumulative hazard function can be used to estimate the survival function.

The estimator is calculated, then, by summing the proportion of those at risk who failed in each interval up to time t.

Implementation with SAS

The Nelson-Aalen estimator is requested in SAS through the nelson option on the proc lifetest statement. SAS will output both Kaplan Meier estimates of the survival function and Nelson-Aalen estimates of the cumulative hazard function in one table.