Basics of Survival Analysis

Ian Crandell

April 23rd, 2015

Outline

1. What is survival analysis?
2. Censoring
3. Survival Function and Hazard Function
4. Data Description
5. Hypothesis Testing
6. Cox Proportional Hazards Model

What is Survival Analysis?

Survival Analysis (SA) is a branch of statistics which deals with measuring lifetimes, or more generally, the time to some event. Some examples:

- Medicine: time to death, or time to relapse of a disease
- Engineering: time until a component breaks down
- Finance: time until a stock reaches a certain value

Censoring

A common feature of survival data is *censoring*. This is when we don't know the exact lifetime of a subject, but only a lower or upper bound for that lifetime. Examples:

- In a medical study about time until relapse, a patient does not relapse before the conclusion of the study. This is called right censoring.
- In a study about the length of the time until marijuana was first used, one subject said "I've used it but I don't know when." This is left censoring, since we know the event happened before some time.

Survival Function and Hazard Function

- The survival function S gives the proportion of subjects which have survived at a specific time t.
- By comparing the survival functions of two populations, it is possible to make claims about which is longer lived.
- The hazard function h can be thought of as the instantaneous chance of death, or of the event. Higher hazard means the event is more likely to occur in the next interval of time.

$$S(x) = \prod_{x_j \leq x} 1 - h(x_j)$$

Survival Function and Hazard Function

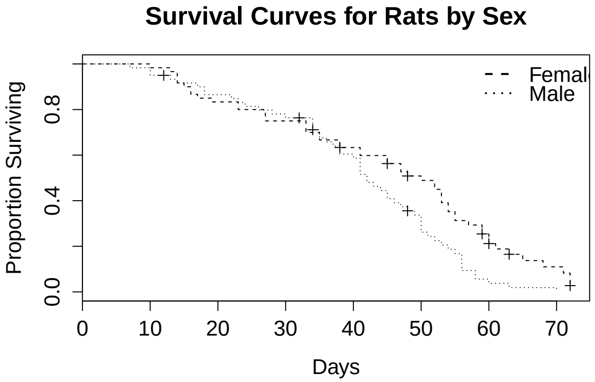

Figure: Survival curves for male and female rats. Pluses denote censored times. Function was estimated using the Kaplan Meier method.

Survival Function and Hazard Function

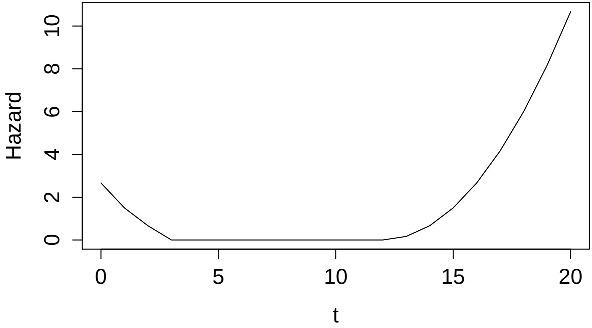

Figure: Hazard curve for an organism. Hazard is high for newborns, constant for adults, and increasing for the elderly.

Data Description

- These data come from an experiment on the efficacy of a rat poison.
- Rats were fed the poison in varying doses in their diets and time until death was recorded.
- There is right censoring. Rats were entered in the study as they entered the lab. Some of them did not die before the study period concluded.

Data Description: Sample Data

lifetimes	sex	poisonconc	died
68	F	0	1
71	M	0	1
33	F	1	1
72	M	1	0
33	F	2	1
52	M	2	1

Table: Lifetime is in days since the poison regimen began. died $= 1$ when the rat died during the study period. died $= 0$ indicates right censoring.

Hypothesis Testing

Hypothesis tests are generally tests on hazard functions. Consider the case of K different hazard functions.

- $H_0 : h_1(t) = \ldots = h_K(t)$
- H_a : At least one hazard function is different.

It's possible to emphasize different parts of the curve to give a weighted hypothesis test. We will focus on the case of an equally weighted curve.

Proportional Hazards Model

The proportional hazards model is used to quantify the effect of a covariate on the hazard function. For a person with covariate vector Z, their hazard function is modeled by

$$h(t) = h_0(t) \times \exp\{\beta^T Z\}.$$

Here, β are regression coefficients and $h_0(t)$ is a baseline hazard function estimated by, say, the Kaplan Meier estimate.

Hazard Ratio Example

Take for example sex as a covariate. We let Z denote sex, where $Z = 0$ for females and $Z = 1$ for males. Here, females are called the *baseline group* since their hazard function is just

$$h(t) = h_0(t),$$

whereas for males it is

$$h(t) = h_0(t) \times \exp\{\beta\}.$$

The *hazard ratio* is the ratio of these hazards, simply equal to $\exp\{\beta\}$. We see $\exp\{\beta\}$ determines the degree to which males have a higher or lower hazard than females.

References

Kaplan, E. L. & Meier, P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53, 1958, 457-48