Time to Event Analysis (Part 2)

Christof Seiler

Stanford University, Spring 2016, STATS 205

Overview

- Last time:
  - Survival data
  - Survival function
  - Hazard function
  - Kaplan-Meier estimator
- Today:
  - Confidence bands with the Boostrap
  - Comparison of survival curves

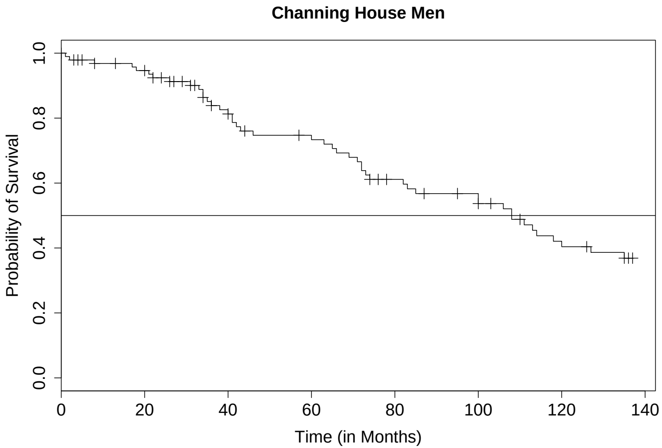
Confidence bands with the Boostrap

- Efron 1981 on Channing house data

```
##    sex entry exit time cens
## 1 Male   782  909  127    1
## 2 Male  1020 1128  108    1
## 3 Male   856  969  113    1
## 4 Male   915  957   42    1
## 5 Male   863  983  120    1
## 6 Male   906 1012  106    1
```

- 97 men in retirement house in Palo Alto
- From opening 1964 until data collection day in July 1975
- 46 were observed exactly, the men died while in the Channing house
- The remaining 51 were censored, five moved elsewhere, and 46 were still alive at data collection day

Confidence bands with the Boostrap



Confidence bands with the Boostrap

- The quantify of interest was the median survival time
- Randomly censored data consist of iid pairs of observations $(X_i, \delta_i), i = 1, \delta, n$,
  - if $\delta_i = 0$ then X_i denotes a censored observation, and
  - if $\delta_i = 1$ then X_i denotes an exact "survival" time
- Efron takes a random sample with replacement from $(X_1, \delta_1), \ldots, (X_n, \delta_n)$
- Then recomputes the survival function on the bootstrap sample
- There are theoretical developements by Akritas in 1986 showing that the boostrap approach works

Log Rank Test

- We could use use our asymptotic results on the KM estimator $\widehat{S}(t)$ from last time or the bootstrap esimator to devise a test equality of survivar function at some time t
- But taking advantage of the entire function will make more efficient use of the data
- The most commonly used statistics are based on nonparametric rank tests

Log Rank Test

- Let failure time $t_1 < \cdots < t_k$ over both samples
- We construct contingency tables at every event time t_i
- d_{0i} and d_{1i} are the number of deaths in group 0 and 1
- n_{0i} and n_{1i} are the number at risk in groups 0 and 1

Group	Failures	Survivors	Total
0	d_{0i}	$n_{0i} - d_{0i}$	n_{0i}
1	d_{1i}	$n_{1i} - d_{1i}$	n_{1i}
Total	d_i	$n_i - d_i$	n_i

- Under the null hypothesis $S_1(t) = S_0(t), 0 < t < \infty$, d_{1i} follows a hypergeometric distribution conditional on the margins

Log Rank Test

- With the hypergeometric distribution we can get

$$E_i = \mathrm{E}(d_{1i}) = n_{0i}\frac{d_i}{n_i}$$

$$V_i = \mathrm{Var}(d_{1i}) = \frac{\sum_{i=1}^{k} n_{1i} n_{0i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)}$$

- And observed is $O_i = d_{0i}$
- Using this expectation and variances and summing over all timepoints t_k, we get the log rank statistics

$$\chi^2 = \frac{\left(\sum_{i=1}^{k} (O_i - E_i)\right)^2}{\sum_{i=1}^{k} V_i}$$

Log Rank Test

- Symmetric in two groups
- The log rank statistic depends on ranks of event times only
- $O - E$ from the k two by two tables are treated as independent
- The $O - E$ in the numberator can be written as

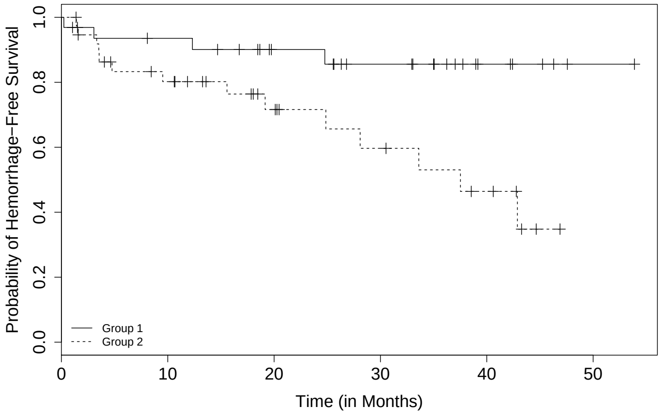
$$\frac{d_{0i} d_{1i}}{d_i} (\widehat{\lambda}_{1i} - \widehat{\lambda}_{0i})$$

- So it quantifies the sums of differences in the hazard function over all k time points scaled by by the fraction of failures

Log Rank Test (Example)

- **Study**: Patients who had survived a lobar intracerebral hemorrhage and whose genotype was known
- **Data**: Survival times (in months) for 71 subjects
- **Event**: Time until recurrence of lobar intracerebral hemorrhage
- **Question**: Genetic effect on recurrence in two gropus with different genotype
- One subject's genotype information is missing and is excluded from analysis
- Of the remaining 70 subjects, 32 are in Group 1 and 38 are in Group 2
- A$+$ sign indicates a censored observation; meaning that at that point in time the subject had yet to report recurrence

Log Rank Test (Example)



Log Rank Test (Example)

```
## Call:
## survdiff(formula = Surv(time, recur) ~ genotype)
##
## n=70, 1 observation deleted due to missingness.
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## genotype=0 32        4     9.28      3.00      6.28
## genotype=1 38       14     8.72      3.19      6.28
##
##  Chisq= 6.3  on 1 degrees of freedom, p= 0.0122
```

- Note that the log-rank test statistic is 6.3 with p-value 0.0122 based on a null χ^2-distribution with 1 degree of freedom
- Thus the log-rank test confirms the difference in survival time of the two groups

References

- The Statistical Analysis of Failure Time Data (2002). Kalbfleisch and Prentice
- Lecture Notes (2005). Ibrahim
- Efron (1981). Censored Data and the Bootstrap
- Akritas (1986). Bootstrapping the Kaplan-Meier Estimator