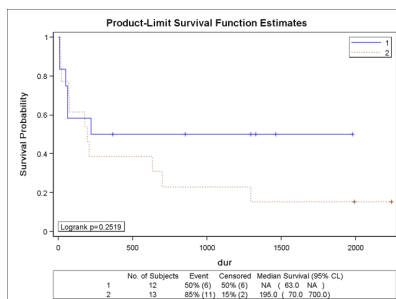


Logrank test for two samples

Introduction

- 25 patients with myelomatosis.
- Patients were randomly assigned to two drug treatments.
- Time in days from the point of randomization to either death or censoring is observed.
- Is there any difference in patient survival for two treatment groups?

Introduction



Introduction

- We have discussed both point and interval estimation of the survival function.
- By plotting the survival curves and confidence intervals, we can visually compare if there is difference in the survival of different groups.
- However, we need a formal statistical procedure to test for a difference between groups.

Hypothesis

- Suppose we wish to compare survival functions of two different populations.
- Consider the following two-sided hypothesis.
 $H_0: S_1(t) = S_2(t)$ for all $t \in \tau$;
 $H_1: S_1(t) \neq S_2(t)$ for some $t \in \tau$.

Data

- The data available to test the above hypothesis consists of independent right-censored samples for each population.
- Let $t_1 < t_2 < \dots < t_D$ be the distinct event times in the pooled sample.
- At time t_i define
 - Y_i : total # of individuals at risk across all samples just prior to t_i
 - d_i : total # of deaths at t_i
 - Y_{ij} : total # of individuals at risk from the j th sample just prior to t_i
 - d_{ij} : total # of deaths at t_i from the j th sample

Data at time t_i

At each event time t_i , we have the following contingency table for all subjects in the risk set:

	# of Death	# of Alive	Total
Group 1	d_{i1}	$Y_{i1} - d_{i1}$	Y_{i1}
Group 2	d_{i2}	$Y_{i2} - d_{i2}$	Y_{i2}
Total	d_i	$Y_i - d_i$	Y_i

Logrank test statistic

- Given the margins of the contingency table at time t_i , if the null hypothesis is true ($S_1 = S_2$), then

$$d_{i1} \sim \text{Hypergeometric}(Y_i, d_i, Y_{i1})$$

- Under null hypothesis, the expected value and variance of d_{i1} are

$$e_{i1} = d_i \frac{Y_{i1}}{Y_i}$$

$$v_{i1} = d_i \frac{Y_{i1}}{Y_i} \frac{Y_i - Y_{i1}}{Y_i - 1} \frac{Y_i - d_i}{Y_i - 1}$$

Logrank test statistics

- Then at time t_i ,

$$u_i = \underset{\text{Observed}}{d_{i1}} - \underset{\text{Expected under } H_0}{e_{i1}}$$

- Then u_i measures how group 1 is different from the pooled sample at time t_i .
- Under H_0 ,

$$E(u_i) = 0, \quad \text{var}(u_i) = v_{i1}.$$

- The **logrank statistic** is given by

$$\text{Logrank} = \frac{[\sum_{i=1}^D u_i]^2}{\sum_{i=1}^D v_{i1}} = \frac{[\sum_{i=1}^D (d_{i1} - e_{i1})]^2}{\sum_{i=1}^D v_{i1}} \sim \chi_1^2.$$

Logrank test statistics

- The logrank test is the most popular method for comparing two survival time distributions.
- The logrank test is essentially the same as the Cochran-Mantel-Haenszel (CMH) test in categorical data analysis, with time as the stratification variable.
- The logrank test is non-parametric test. The test statistics only depends on the **rank**s of observed times; **not the exact values** of event or censoring times.
- The logrank test is the most powerful when the ratio between two hazard functions is constant across time (proportional hazard).

Proportional Hazard

