Logrank test for k samples

Logrank test

- We have introduced the logrank test and weighted logrank test for comparing two survival time distributions.
- We can extend the tests to compare survival distributions of more than two groups.

An example

- BMT data of 137 bone marrow transplant patients extracted. (SAS manual)
- At the time of transplant, each patient is classified into one of three risk categories: ALL (acute lymphoblastic leukemia), AML (acute myeloctic leukemia)-Low Risk, and AML-High Risk.
- Of interest to compare the disease-free survival time for patients in three different risk categories.

Hypothesis

- Suppose we wish to simultaneously compare survival functions of k $(k \geq 2)$ populations.
- Consider the following two-sided hypothesis.

H_0: $S_1(t) = S_2(t) = \cdots = S_k(t)$ $for\ all\ t \in \tau$
H_1: $at\ least\ one\ of\ S_i(t)'s\ is\ different\ for\ some\ t \in \tau.$

Data

- The data available to test the above hypothesis consists of independent right-censored samples for each of the k populations.
- Let $t_1 < t_2 < \cdots < t_D$ be the distinct event times in the pooled sample.
- At time t_i define
  - Y_i :total # of subjects at risk just prior to t_i among all k samples
  - d_i : total # of deaths at t_i among all k samples
  - Y_{ij}: total # of subjects at risk from the jth sample just prior to t_i
  - d_{ij}: total # of deaths at t_i from the jth sample
- Note that $Y_i = \sum_{j=1}^{k} Y_{ij}$, and $d_i = \sum_{j=1}^{k} d_{ij}$.

Data at time t_i

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

	Group 1	⋯	**Group k**	**Total**
Death	d_{i1}	⋯	d_{ik}	d_i
Alive	$Y_{i1} - d_{i1}$	⋯	$Y_{ik} - d_{ik}$	$Y_i - d_i$
Total	Y_{i1}	⋯	Y_{ik}	Y_i

Logrank test statistics

- Given the margins of the contingency table at time t_i, if the null hypothesis is true, then
  $(d_{i1}, \cdots, d_{ik}) \sim$ Multivariate Hypergeometric
- Under H_0,

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

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

Test Statistic

- For each group j = $1, \ldots, k$, define
  $$Z_j = \sum_{i=1}^{D} W(t_i) \left\{ d_{ij} - \frac{Y_{ij}}{Y_i} d_i \right\}$$
  - Z_j measures the difference between observed and expected (under null hypothesis) for group j.
- Note that $Z_1, Z_2, \ldots, Z_k$ are linearly dependent that
  $$\sum_{j=1}^{k} Z_j = \sum_{j=1}^{k} \sum_{i=1}^{D} W(t_i) \left\{ d_{ij} - \frac{Y_{ij}}{Y_i} d_i \right\} = 0.$$
- The test statistic is constructed by selecting $k-1$ of the Z_j's.
- The test statistics is the same for any $k-1$ of the Z_j's.
- Without loss of generality, we take the first $k-1$ elements with
  $$\boldsymbol{Z} = (Z_1, \ldots, Z_{k-1})^T.$$

Test Statistic

- Let $\boldsymbol{Z} = (Z_1, \ldots, Z_{k-1})^T$.
- Let the variance-covariance matrix of Z be $\boldsymbol{\Sigma} = \left(\sigma_{jj'}\right)_{j,j'=1}^{k-1}$.
- Then Log-rank test statistics for comparing k survival curves is defined as
  $$\text{Logrank} = \boldsymbol{Z}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{Z}.$$
- Under null hypothesis,
  $$\text{Logrank} = \boldsymbol{Z}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{Z} \to \chi^2_{k-1}.$$

Summary

- Log-rank test for comparing two or more groups
- Weighted Log-rank test by including weights
- Powerful nonparametric test for survival data