MDSC 643.01, Biostatistics I
Brant, Fall 2002

Time to Event Data

Survival Analysis

Methods for examining time to occurrence of event, often death or morbidity (e.g. cancer recurrence, MI)

Times measured from defined time origin (usually subject specific)
- definition depends on purpose

Censored observations: subjects where follow-up ends before occurrence of event due to:
- end of pre-determined follow-up
- unintended loss to follow-up

In absence of censoring, can easily plot survival curve

y = (decreasing) proportion of subjects still "event-free"

x = follow-up time

Allowing for censoring?

large samples - life table approach

smaller samples - Kaplan Meier curve.

Hazard rate at time t = probability that an event occurs (given subject hasn't already had the event.

Estimated as $\dfrac{\text{\# experiencing event at time t}}{\text{\# still event free up to time t}} = \dfrac{\text{\# events}}{\text{\# at risk of event}}$

- can be calculated even in presence of censoring

Constructing the survival curve:

S(t) (the survival function) = probability of survival (proportion surviving) up to time t

If events can occur at $t_1, t_2, t_3, \ldots t_i$, S(t) can be expressed as

$$S(t_i) = [1 - h(t_1)] \times [1 - h(t_2)] \times [1 - h(t_3)] \times \ldots [1 - h(t_i)]$$

$\hat{S}(t_i)$ derived by estimate h 's as proportion in relative "risk set"

Plotting a Survival Curve, $\hat{S}(t)$

- steps occur at "events"
- censoring times often indicated with "|"'s
- provides a valid estimate of true "uncensored" survival curve if censoring is independent of event (non-informative censoring)

Comparing Survival Curves

Graphical: superimposed Kaplan-Meier curves for two groups

Formal test of hypothesis: Log-rank test (variant of Mantel-Haenszel, time as stratifying variable)

Assumptions of the Log-Rank test

Independent samples of survival data

Non-informative censoring in both samples (but may not be "uniform")

Proportional hazards assumption

Consider TRUE hazard function $h_i(t)$ for group i (i = 1,2)

$\frac{h_1(t)}{h_2(t)} = \lambda$, same value for all values of t.

λ is "relative hazard" - similar to relative risk

Log-rank test

Based on above assumptions, test $H_0 : \lambda = 1$

O = observed number deaths in one group (e.g. Group 1)

E = expected number, based on follow-up pattern

V = variance of O (messy)

$X^2 = \frac{(|O-E|-.5)^2}{V}$ refer to χ^2 on 1 degree of freedom

Checking assumptions

Informal: Check for "parallel" pattern in survival curves

Log-log plot: Plot of $-\log_e[\log_e(\hat{S}(t))]$ versus $\log_e(t)$

- improves visual acumen