1/20/2023

V. The Log-Rank Test for Several Groups

- Ho: All survival curves are the same
- Log-rank statistics for >2 groups involve variances and covariances of $O_i - E_i$.
- There is a convenient matrix formula for the test statistic.

Use a computer

If G is the number of groups, then the log-rank statistic is $\sim \chi^2$ with (G-1) degrees of freedom.

VI. Alternatives to the Log Rank Test

- Wilcoxon, Tarone-Ware, Peto, Flemington-Harrington
- These all apply different weights to the f-th failure time.

This could be of use to my application! ~~Since I'm~~ The initial condition of the simulation may be very unlikely under the stationary distribution, but as the simulation goes on, the initial condition for each message will be more likely under the stationary distribution; therefore, it could be useful to assign higher weights to later messages.

Wilcoxon Test:
- Weight at t_f is n_f, so: earlier failures have higher weight.

Matrix formula

$$(O-E)^T [Cov]^{-1} (O-E)$$

Weights Used for Various Test Statistics

Test statistic	$w(t_{(f)})$
log rank	1
Wilcoxon	n_f
Tarone-Ware	$\sqrt{n_f}$
Peto	$\hat{S}(t_{(f)})$
Flemington-Harrington	$\hat{S}(t_{(f-1)})^p \times [1 - \hat{S}(t_{(f-1)})]^q$

"The survival estimate."

$p \& q$ are parameters the user sets

- The best choice is the test with the most power.
- There may be clinical reasons to choose a given weighting (i.e., use Wilcoxon or Tarone-Ware if the treatment is known to wear off over time.
- Choose weights a priori: No p hacking!

Stratified log rank test
- O-E scores calculated within strata
- O-E scores then summed across strata
- Allows control over the stratified variable.
- Reduces the sample size in each stratum.

Ch 3 shows how modeling can be used to test an association of a predictor variable while simultaneously controlling for other covariates.

VII. Confidence intervals for KM curves

$$\hat{S}_{KM}(t) \pm 1.96\sqrt{\hat{Var}[\hat{S}_{KM}(t)]},$$

where Greenwood's formula is given by

# of failures at $t_{(f)}$

$$\hat{Var}[\hat{S}_{KM}(t)] = (\hat{S}_{KM}(t))^2 \sum_{f: t_f \leq t} \left(\frac{m_f}{n_f(n_f - m_f)} \right)$$

VIII. Confidence intervals for the median survival time.
Don't need this right now.

Ch. 3: The Cox Proportional Hazards Model and Its Characteristics

Introduction

- Show similarity between Cox proportional hazards model and linear regression
- Cox model, properties, assumptions, etc.
- Describe how and why to use "age as the time scale" rather than "time in follow-up" as the outcome variable.