Lecture Five: Comparing Multiple Samples: Non-Parametric tests (Cont.)

1. Weighted log rank Tests

- For each time interval $(t_{(j-1)}, t_{(j)}]$, in which there is only one distinct failure time (allow ties), we have a 2 by 2 table

Group	# of deaths at $t_{(j)}$	# of surviving beyond $t_{(j)}$	# at risk just before $t_{(j)}$
I	d_{1j}	$n_{1j}- d_{1j}$	n_{1j}
II	d_{2j}	$n_{2j}- d_{2j}$	n_{2j}
Total	d_j	n_j-d_j	n_j

The expected events:

$$e_{1j} = d_j*n_{1j}/n_j$$
$$e_{2j} = d_j*n_{2j}/n_j$$

$d_{1j}|d_j$ has hypergeometric distribution with

$$E(d_{1j}|d_j) = e_{1j}$$
$$Var(d_{1j}|d_j)= v_{1j} = \frac{n_{1j}n_{2j}d_j(n_j-d_j)}{n_j^2(n_j-1)}$$

- **A family of weighted log rank statistics**

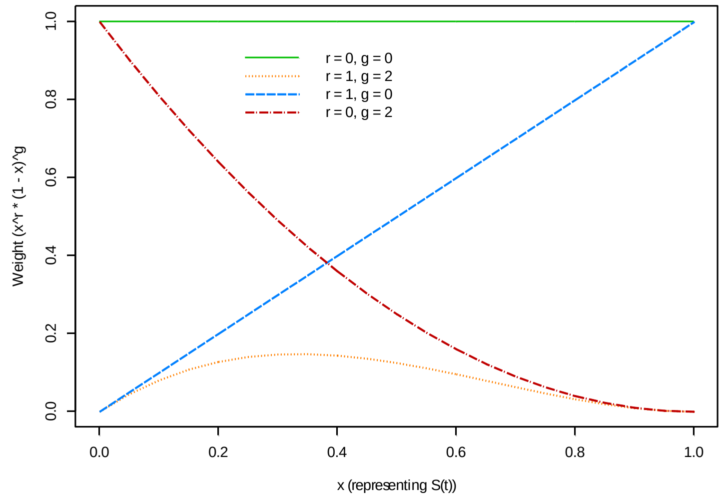
$$U_{WT} = \sum_{j=1}^{r} w_j(d_{1j}-e_{1j})$$

  - A general weighting scheme

$$w_j = S(t_j)^\rho(1-S(t_j))^\gamma$$

  Here, $\rho >= 0$, $\gamma >= 0$, $S(t_j)$ is the KM estimate pooled from both groups.

  - Effects of weights
    - $\rho = 0$ and $\gamma = 0$: equal weight
    - $\rho > 0$ and $\gamma > 0$: more weight on difference in the middle
    - $\rho > 0$ and $\gamma = 0$: more weight on earlier difference
    - $\rho = 0$ and $\gamma > 0$: more weight on later difference



- Splus implementation : $\gamma = 0$ and $w(t) = S(t)^{\rho}$ (Ref.: *Biometrika* vol. 69, pp. 553-566 (1982) by Harrington and Fleming)
  - Splus function: survdiff()
  - $\rho = 0$: w(t) = 1, log-rank/Mental-Haenszel
  - $\rho = 1$: w(t) = S(t), Peto Peto/Prentice (generalized Wilcoxon)
  - $\rho > 0$: more weight on earlier difference (S(t) is non-deceasing function)
  - $\rho < 0$: more weight on later difference (interpretation less natural)
- SAS implementation: strata statement (test option)

- **Other weighted rank based Tests**

$w_j = n_j$ (The Gehan (1965) statistic)

$w_j = n_j^{1/2}$ (One of Tarone and Ware (1977) test statistics)

- **The Wilcoxon Test**

$$U_W = \sum_{j=1}^{r} n_j (d_{1j} - e_{1j})$$

The variance of the Wilcoxon statistic above is

$$V_w = \sum_{j=1}^{r} n_j^2 v_{1j}$$

and the Wilcoxon test statistic is

$$W_W = U_W^2 / V_W \sim \chi^2(1),$$

when the null hypothesis is true (why?).

- **SAS implementation:** see options of *strata statement* of PROC LIFETEST.

- **Example 2.13: Wilcoxon test (see output for example 2.12)**

- **Comparison of the log rank and Wilcoxon tests**

  - Equal weight (detect difference that is consistent over time) for log rank test, more weight on the earlier difference for Wilcoxon test.
  - Log rank: more suitable when assumption of proportional hazards is satisfied ($h_1(t) = \varphi h_2(t)$)
  - Necessary (not a sufficient) condition for proportional hazards: The true survivor functions do not cross ($S_1(t) = [S_2(t)]^{\varphi}$)
  - Example 2.14: KM plot

2. **Comparison of more than two samples**

- Same idea as in two group case: measuring discrepancy
- Kruskal-Wallis tests (more general than Wilcoxon tests)
- log-rank tests based on sequence of 2 by g tables (g > 2)

$$U_{Lk} = \sum_{j=1}^{r}(d_{kj} - \frac{n_{kj} d_j}{n_j}), \text{ [Wilcoxon test:} U_{Wk} = \sum_{j=1}^{r} n_j (d_{kj} - \frac{n_{kj} d_j}{n_j})\text{]}$$

for k = 1, 2, …, g-1.The variance matrix for log-rank test is

$$V_L = (V_{Lkk'}),$$

where

$$V_{Lkk'} = \sum_{j=1}^{r} \frac{n_{kj} d_j (n_j - d_j)}{n_j (n_j - 1)} (\delta_{kk'} - \frac{n_{kj}}{n_j}).$$

- The test statistic: $U_L' V_L^{-1} U_L \sim \chi^2(g-1)$ (why?)

3. **Further Generalizations**

- **Stratification within a treatment group is necessary when subjects are not homogenous: Section 2.8**

- Handle additional covariates (confounding variables).
- Example: Multi-center clinical trial (stratified by center); stratified by sex or other potential risk factors.
- Stratified log-rank/Wilcoxon test: Basically, Calculating the values of U- and V-statistics for each stratum, then combine them (see following test statistic).
- Test statistic

$$W_S = \frac{(\sum_{k=1}^{s} U_{Lk})^2}{\sum_{k=1}^{s} V_{Lk}} \sim \chi^2(1)$$

- Example 2.15: Two vaccines after surgery for melanoma patients

Summarized output from following SAS program:

Age group	U_L	V_L	W_L (U_L^2 / V_L)
21-40	-0.2571	1.1921	0.055
41-60	0.4778	0.3828	0.596
61-	1.0167	0.6497	1.591
Total	1.2374	2.2246	

$W_S = 1.2374^2/2.2246=0.688$. Test statistic $W_S \sim \chi^2(1)$. P-value = 0.41.

/* SAS program: melanoma.sas (SAS Version 8) */

```
options pagesize=60 linesize=79 nodate nonumber;
libname fu '../../sdata';
data fu.melanoma;
infile '../../data/melanoma.dat';
input age tx survt censor;
data w1;
    set fu.melanoma;
if age = 1;
proc lifetest notable;
    time survt*censor(0);
    strata  tx;
data w2;
    set fu.melanoma;
if age = 2;
proc lifetest notable;
    time survt*censor(0);
    strata  tx;
data w3;
    set fu.melanoma;
if age = 3;
proc lifetest notable;
```

```
        time survt*censor(0);
        strata  tx;
run;
```
/* SAS program: melanoma.sas (SAS Version 9) */
```
options pagesize=60 linesize=79 nodate nonumber;
libname fu '../../sdata';
data w;
       set fu.melanoma;
proc lifetest notable;
      time survt*censor(0);
strata age / group = tx;
run;
```

- It's not flexible as Cox model (proportional hazards model).

- **When treatment groups are ordered in some way: Log-rank test for trend**

  - Examples: groups correspond to increasing doses of a treatment; the stage of a disease, or age group.
  - Log-rank test may not lead to a significant difference among groups even though the hazard of death increase or decrease across the groups

  - Mathematically, the alternative hypothesis is

$$H_A : S_1(t) < S_2(t) < \ldots < S_g(t)$$

  - Log-rank test for trend statistic:

$$U_T = \sum_{k=1}^{g} w_k (d_{k.} - e_{k.}),$$

where w_k is a code assigned to the k'th group, k = 1, 2, …, g and

$$d_{k.} = \sum_{j=1}^{r_k} d_{kj}, \quad e_{k.} = \sum_{j=1}^{r_k} e_{kj}$$

are the observed and expected numbers of deaths in the k'th group.
The variance of U_T is given by

$$V_T = \sum_{k=1}^{g} (w_k - \bar{w})^2 e_{k.},$$

where

$$\bar{w} = \frac{\sum_{k=1}^{g} w_k e_{k.}}{\sum_{k=1}^{g} e_{k.}},$$

Then, the statistic $W_T = U_T^2 / V_T \sim \chi^2(1)$ under $H_0 : S_1(t) = S_2(t) = \ldots = S_g(t)$

- Example 2.16: Melanoma patients (BCG arm only: trend over age?) (page 55)

SAS output:

Trend Tests

Test	Test Statistic	Standard Error	z-Score	Pr > |z|
Log-Rank	2.5692	1.5465	1.6613	0.0967
Wilcoxon	25.0000	1 4.4568	1.7293	0.0838

SAS program:

```
options pagesize=60 linesize=79 nodate nonumber;
libname fu '../../sdata';
data w;
        set fu.melanoma;
        if tx = 1;
proc lifetest notable;
        time survt*censor(0);
        strata age / trend;
```

- More flexible approach: Cox model (next chapter)

- **Renyi type of test (Similar to Kolmogorov-Smirnov test, but with censored data)**

See pages 223-224 of Klein & Moeschberger's book (reference #1 in the syllabus).

Reading assignment: Read section 2.10