Syllabus

1. Introduction
   - Survival data
   - Censoring mechanism
   - Application in medical field
2. Concepts and definitions
   - Survival function
   - Hazard function
3. Non-parametric approach
   - Life table
   - Kaplan-Meier survival estimate
   - Hazard function
   - Median and percentile survival time
4. **Hypothesis testing**
   - **Overview – hypothesis, test statistics, p-values**
   - **Log-rank**
   - **Wilcoxon**
   - **Gehan test**
5. Study design and sample size estimation
   - Overview
   - Survival sample size estimation
   - Accrual time and Study duration
6. Semiparametric model – proportional hazard model
   - Partial likelihood
   - Inference
   - Time varying covariates
   - Stratification
7. Model checking in the PH model
   - Model checking
   - Residuals
8. Parametric model
   - Parametric proportional hazard model
   - Accelerate failure model
9. Other topics
   - Competing risk
   - Recurrent events
   - Non-proportional hazard ratio
   - Interval censoring

Recap

- Life-table
  - Discrepancy between lecture note and SAS manual
- Confidence interval for median survival

Life-table Estimate

- Also know as the actuarial estimate
  - Used in continuous survival data
  - Grouped data – similar to discrete survival time
- Divide survival data *T* into intervals, for the i^{th} interval
  - $t_{i-1} \leq t < t_i \;\; or \;\; [t_{i-1}, t_i) \;\; i = 1, \ldots, s$
  - The intervals may or may not be of equal length

Life-table Estimate

- Within the i^{th} interval
  - d_i, number of events
  - c_i, number of censors
  - n_i, number of subjects at risk at t_i
  - $n_i' = n_i - c_i/2$, average number of subjects at the interval

- Why $n_i' = n_i - c_i/2$,
- $i = 1, \ldots, s$

Life-table Estimate – Conditional Probability

- For the i^{th} interval
  - Conditional probability of surviving through the i^{th} interval
    $$\hat{p}_i = \frac{n_i' - d_i}{n_i'}$$
  - Conditional probability of experiencing an event in the i^{th} interval
    $$\hat{q}_i = 1 - \hat{p}_i = \frac{d_i}{n_i'}$$
- Why $n_i' = n_i - c_i/2$,
  - Not $n_i' = n_i$, underestimate the risk $\hat{q}_i$
  - Not $n_i' = n_i - c_i$, overestimate the risk $\hat{q}_i$
  - $n_i' = n_i - c_i/2$, assuming constant censoring rate

Life-table Estimate – Survival Function

- In the i^{th} interval
  - Survival function at the end of the i^{th} interval

$$\hat{S}_L(t_0) = 1$$

$$\hat{S}_L(t_i) = \hat{S}_L(t_{i-1})\left(1 - \frac{d_i}{n_i'}\right)$$

$$\text{var}\{\hat{S}_L(t_{i-1})\} = \hat{S}_L{}^2(t_{i-1}) \sum_{j=1}^{i-1} \frac{d_j}{n_j'(n_j' - d_j)}$$

$$i = 1, \ldots, s$$

Life-Table – SAS Manual Notation

SAS/STAT User's Guide

Life-Table Method

The life-table estimates are computed by counting the numbers of censored and uncensored observations that fall into each of the time intervals $[t_{i-1}, t_i)$, $i = 1, 2, \ldots, k+1$, where $t_0 = 0$ and $t_{k+1} = \infty$. Let n_i be the number of units that enter the interval $[t_{i-1}, t_i)$, and let d_i be the number of events that occur in the interval. Let $b_i = t_i - t_{i-1}$, and let $n'_i = n_i - w_i/2$, where w_i is the number of units censored in the interval. The effective sample size of the interval $[t_{i-1}, t_i)$ is denoted by n'_i. Let t_{mi} denote the midpoint of $[t_{i-1}, t_i)$.

The conditional probability of an event in $[t_{i-1}, t_i)$ is estimated by

$$\hat{q}_i = \frac{d_i}{n'_i}$$

and its estimated standard error is

$$\hat{\sigma}(\hat{q}_i) = \sqrt{\frac{\hat{q}_i \hat{p}_i}{n'_i}}$$

where $\hat{p}_i = 1 - \hat{q}_i$.

The estimate of the survival function at t_i is

$$\hat{S}(t_i) = \begin{cases} 1 & i = 0 \\ \hat{S}(t_{i-1}) p_{i-1} & i > 0 \end{cases}$$

and its estimated standard error is

$$\hat{\sigma}\left(\hat{S}(t_i)\right) = \hat{S}(t_i) \sqrt{\sum_{j=1}^{i-1} \frac{\hat{q}_j}{n'_j \hat{p}_j}}$$

SAS Life Table

Life Table Survival Estimates

Interval												Evaluated at the Midpoint of the Interval			
[Lower,	Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error	PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	1	456	0	2418.0	0.1886	0.00796	1.0000	0	0	5.3313	0.1749	0.1886	0.00796	0.208219	0.009698
1	2	226	39	1942.5	0.1163	0.00728	0.8114	0.1886	0.00796	6.2499	0.2001	0.0944	0.00598	0.123531	0.008201
2	3	152	22	1686.0	0.0902	0.00698	0.7170	0.2830	0.00918	6.3432	0.2361	0.0646	0.00507	0.09441	0.007649
3	4	171	23	1511.5	0.1131	0.00815	0.6524	0.3476	0.00973	6.2262	0.2361	0.0738	0.00543	0.119916	0.009154
4	5	135	24	1317.0	0.1025	0.00836	0.5786	0.4214	0.0101	6.2185	0.1853	0.0593	0.00495	0.108043	0.009285
5	6	125	107	1116.5	0.1120	0.00944	0.5193	0.4807	0.0103	5.9077	0.1806	0.0581	0.00503	0.118596	0.010589
6	7	83	133	871.5	0.0952	0.00994	0.4611	0.5389	0.0104	5.5962	0.1855	0.0439	0.00469	0.1	0.010963
7	8	74	102	671.0	0.1103	0.0121	0.4172	0.5828	0.0105	5.1671	0.2713	0.0460	0.00518	0.116719	0.013545

Confidence Interval for Median and Quantiles Survival Time

K-M Estimator Median and Quantiles

- Recall, the p^{th} quantile of the survival function is

$$S(t_p) = P(T > t_p) = p$$

$$t_p = \inf\{\text{t}: S(t_p) \leq p\}$$

- So, for the K-M estimator, $\hat{S}_K(t_p) = p$

$$t_p = \inf\{\text{t}: \hat{S}_K(t_p) \leq p\}$$

Recall the example of the death events from the colon cancer

Confidence Interval for Median and Quantiles Survival Time

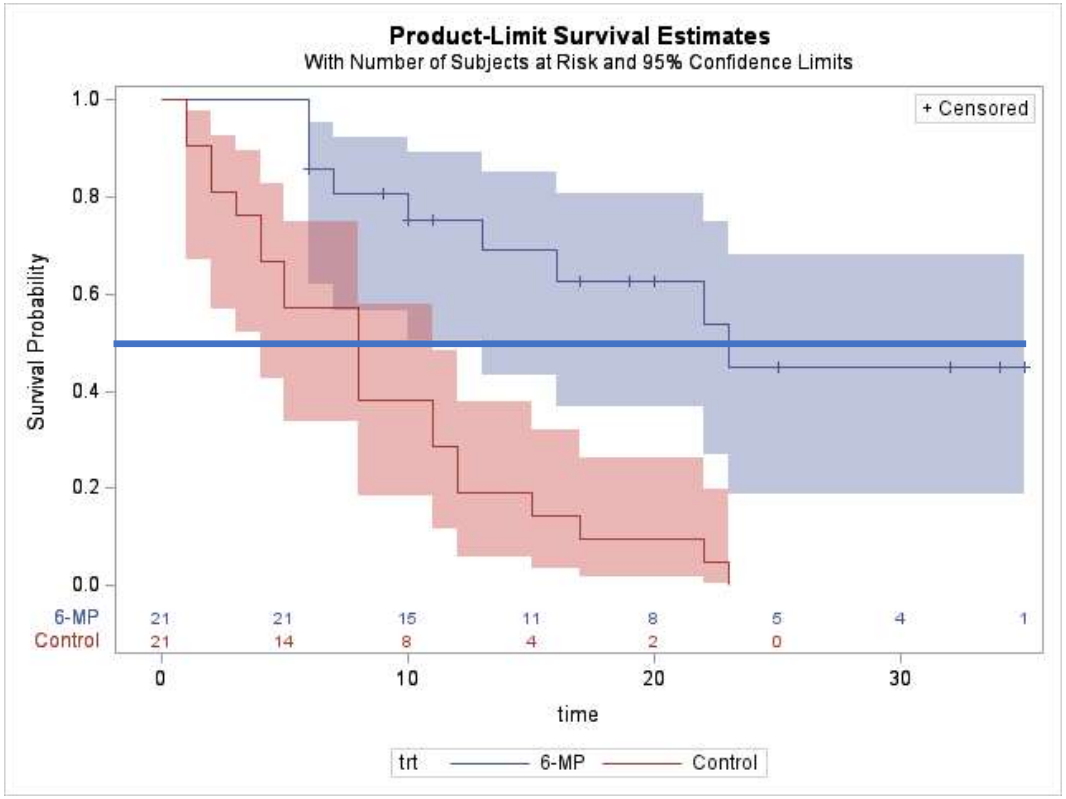
- For the K-M estimator, $\hat{S}_K(t_p) = p$

$$Var\left(\hat{S}_K(t_p)\right) = \left(\frac{d\hat{S}_K(t_p)}{dt_p}\right)^2 \text{var}(t_p)$$

$$= \left(-\hat{f}_K(t_p)\right)^2 \text{var}(t_p)$$

$$\text{var}(t_p) = \frac{1}{\left(\hat{f}_K(t_p)\right)^2} Var\left(\hat{S}_K(t_p)\right)$$

Leukemia Data



Leukemia Data

```
ods graphics on;
proc lifetest data=example method=KM
plots=survival (cl atrisk=0 to 35 by 5)
conftype=loglog outsurv=survival;
  time time*event(0);
  strata trt;
  run;
ods graphics off;
```

Obs	trt	time	_CENSOR_	SURVIVAL	SDF_LCL	SDF_UCL	STRATUM
1	6-MP	0	.	1.00000	1.00000	1.00000	1
2	6-MP	6	0	0.85714	0.61972	0.95155	1
3	6-MP	6	1	0.85714	.	.	1
4	6-MP	7	0	0.80672	0.56315	0.92281	1
5	6-MP	9	1	0.80672	.	.	1
6	6-MP	10	0	0.75294	0.50320	0.88936	1
7	6-MP	10	1	0.75294	.	.	1
8	6-MP	11	1	0.75294	.	.	1
9	6-MP	13	0	0.69020	0.43161	0.84907	1
10	6-MP	16	0	0.62745	0.36751	0.80491	1
11	6-MP	17	1	0.62745	.	.	1
12	6-MP	19	1	0.62745	.	.	1
13	6-MP	20	1	0.62745	.	.	1
14	6-MP	22	0	0.53782	0.26778	0.74679	1
15	6-MP	23	0	0.44818	0.18805	0.68014	1
16	6-MP	25	1	.	.	.	1
17	6-MP	32	1	.	.	.	1
18	6-MP	32	1	.	.	.	1
19	6-MP	34	1	.	.	.	1
20	6-MP	35	1	.	.	.	1
21	Control	0	.	1.00000	1.00000	1.00000	2
22	Control	1	0	0.90476	0.67005	0.97529	2
23	Control	2	0	0.80952	0.56891	0.92389	2
24	Control	3	0	0.76190	0.51939	0.89326	2
25	Control	4	0	0.66667	0.42535	0.82504	2
26	Control	5	0	0.57143	0.33798	0.74924	2
27	Control	8	0	0.38095	0.18307	0.57779	2
28	Control	11	0	0.28571	0.11656	0.48182	2
29	Control	12	0	0.19048	0.05948	0.37743	2
30	Control	15	0	0.14286	0.03566	0.32116	2
31	Control	17	0	0.09524	0.01626	0.26125	2
32	Control	22	0	0.04762	0.00332	0.19704	2
33	Control	23	0	0.00000	.	.	2

Leukemia Data

Quartile Estimates 6-MP				
Percent	Point Estimate	95% Confidence Interval		
		Transform	[Lower	Upper)
75	.	LOGLOG	23.0000	.
50	23.0000	LOGLOG	13.0000	.
25	13.0000	LOGLOG	6.0000	22.0000

Quartile Estimates				
Percent	Point Estimate	95% Confidence Interval		
		Transform	[Lower	Upper)
75	.	LINEAR	23.0000	.
50	23.0000	LINEAR	13.0000	.
25	13.0000	LINEAR	6.0000	23.0000

Quartile Estimates Control				
Percent	Point Estimate	95% Confidence Interval		
		Transform	[Lower	Upper)
75	12.0000	LOGLOG	8.0000	22.0000
50	8.0000	LOGLOG	4.0000	11.0000
25	4.0000	LOGLOG	1.0000	5.0000

Quartile Estimates				
Percent	Point Estimate	95% Confidence Interval		
		Transform	[Lower	Upper)
75	12.0000	LINEAR	8.0000	17.0000
50	8.0000	LINEAR	4.0000	11.0000
25	4.0000	LINEAR	2.0000	8.0000

Optional Readings

- Brookmeyer R, Crowley J. A confidence interval for the median survival time. Biometrics 1982;38:29-41.

Topics – 2 Independent Samples

- Overview – hypothesis testing
- Overview – nonparametric tests

- The log-rank test
- The Wilcoxon test
- Gehan's tests

    - Gehan (1965, Biometrika): modifying rank tests to allow censoring
    - Mantel (1966, Cancer Chem): adapting data to use methods for several 2x2 tables

Compare What?

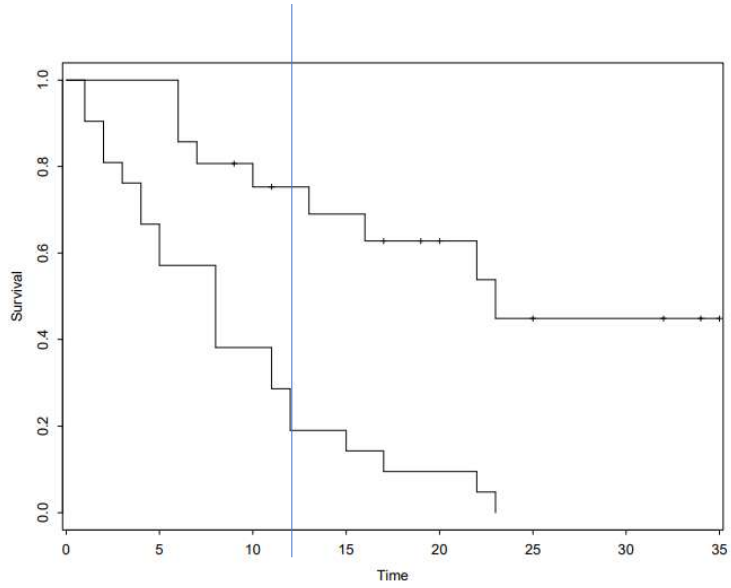


Figure 1: Time to remission of leukemia patients

- Survival probability at certain time
  - Test difference at 1 year survival
  - Can we do this with what we have learned so far?
- Will 6-MP treatment prolong survival?
  - Comparing two survival curve

Leukemia Data
Compare Survival Rates at 1 Year

Stratum 1: trt = 6-MP

Product-Limit Survival Estimates						
time		Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000		1.0000	0	0	0	21
6.0000		.	.	.	1	20
6.0000		.	.	.	2	19
6.0000		0.8571	0.1429	0.0764	3	18
6.0000	*	.	.	.	3	17
7.0000		0.8067	0.1933	0.0869	4	16
9.0000	*	.	.	.	4	15
10.0000		0.7529	0.2471	0.0963	5	14
10.0000	*	.	.	.	5	13
11.0000	*	.	.	.	5	12
13.0000		0.6902	0.3098	0.1068	6	11
16.0000		0.6275	0.3725	0.1141	7	10
17.0000	*	.	.	.	7	9
19.0000	*	.	.	.	7	8
20.0000	*	.	.	.	7	7
22.0000		0.5378	0.4622	0.1282	8	6
23.0000		0.4482	0.5518	0.1346	9	5
25.0000	*	.	.	.	9	4
32.0000	*	.	.	.	9	3
32.0000	*	.	.	.	9	2
34.0000	*	.	.	.	9	1
35.0000	*	.	.	.	9	0

Stratum 2: trt = Control

Product-Limit Survival Estimates						
time		Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000		1.0000	0	0	0	21
1.0000		.	.	.	1	20
1.0000		0.9048	0.0952	0.0641	2	19
2.0000		.	.	.	3	18
2.0000		0.8095	0.1905	0.0857	4	17
3.0000		0.7619	0.2381	0.0929	5	16
4.0000		.	.	.	6	15
4.0000		0.6667	0.3333	0.1029	7	14
5.0000		.	.	.	8	13
5.0000		0.5714	0.4286	0.1080	9	12
8.0000		.	.	.	10	11
8.0000		.	.	.	11	10
8.0000		.	.	.	12	9
8.0000		0.3810	0.6190	0.1060	13	8
11.0000		.	.	.	14	7
11.0000		0.2857	0.7143	0.0986	15	6
12.0000		.	.	.	16	5
12.0000		0.1905	0.8095	0.0857	17	4
15.0000		0.1429	0.8571	0.0764	18	3
17.0000		0.0952	0.9048	0.0641	19	2
22.0000		0.0476	0.9524	0.0465	20	1
23.0000		0	1.0000	.	21	0

Compare Survival Rate at 1 Year

- $S_{6-MP}(12) = 0.753 \quad se\left(S_{6-M}\ (12)\right) = 0.0963$
- $S_{Control}(12) = 0.191 \quad se\left(S_{Control}(12)\right) = 0.0857$
- Test the survival rate difference

$$\frac{S_{6-MP}(12) - S_{Control}(12)}{\sqrt{se\left(S_{6-MP}(12)\right)^2 + se\left(S_{S_{Control}}(12)\right)^2}} = \frac{0.753 - 0.191}{\sqrt{0.0963^2 + 0.0857^2}} = 4.36$$

$$\text{p} - \text{value} \leq 0.001$$

Compare Two Survival Curves

- Rank based tests

Review

- Hypothesis Testing
- General parametric vs Nonparametric tests

- Background knowledge for nonparametric tests of survival data
  - 2x2 tables
    - Hypergeometric test
    - Cochran-Mantel-Haenszel (CMH) test
  - Wilcoxon rank sum test

Hypothesis Testing

- Hypothesis testing answer questions such as
  - Consider one group receive a new treatment
    - Have subjects improved after the new treatment ?
  - Consider two groups
    - Group 1 – new treatment
    - Group 0 – placebo or control
    - Have the subjects in Group 1 became better in general than those in Group 0?
- Hypothesis testing could be
  - Parametric vs. nonparametric
  - One-sided vs. two-sided

Parametric

- Distributions of the data are known

$$f(t|\theta)\,, \theta \in \Theta, \text{ where}$$

  θ is a vector parameters describe the distribution of random variable T
  Θ is the parameter space
- The questions of interest can be described by distribution parameters
  - Such as mean or proportions

- The Hypotheses are set up with the parameters
  - One sample:
    - 2-sided: H_0: $\mu = 0\ vs\ \ H_A$: $\mu \neq 0$
    - 1-sided: H_0: $\mu \leq 0\ vs\ \ H_A$: $\mu > 0$
  - Two samples
    - 2-sided: H_0: $\mu_1 = \mu_0$ vs H_A: $\mu_1 \neq \mu_0$
    - 1-sided: H_0: $\mu_1 \leq \mu_0\ vs\ \ H_A$: $\mu_1 > \mu_0$

Nonparametric

- AKA: distribution-free tests
- When to use
  - Difficult to make distribution assumptions
    - Often occurred when sample size is small
  - Data are ordinal
    - Parametric models are often used when the categories are 5 or more
  - Data are skewed
    - Sometimes skewed data are transformed and then use parametric models
- Hypotheses
  - 1-sample 1-sided
    - H_0 : median $\leq$ reference value vs. H_A: median $>$ reference value
  - 2-sample 1-sided
    - $H_0: F_1 \leq F_0 \; vs \;\; H_A: F_1 > F_0$, where F represents data distribution

Nonparametric vs. Parametric tests

Nonparametric

- 1-sample – test median against a reference or targeted value
  - Sign test
  - Wilcoxon signed test – require symmetric distributions such as Cauchy and uniform distributions
- 2 independent samples
  - Wilcoxon rank sum test (Mann Whitney U test)
- 2 or more independent samples
  - Kruskal-Wallis test
- Spearman correlation

Parametric

- 1-sampe t-test
  - Paired t-test
- 2-samples
  - 2-sample t-tests
  - One-way ANOVA
- 2 or more independent samples
  - One-way ANOVA
- Pearson correlation

Review of Wilcoxon Rank Sum Tests

- Consider a randomized and controlled clinical trial testing two treatments

- Group 0
  - n_0 - number of subjects in Group 0
  - $x_{01}, x_{02}, \ldots, x_{0n_0}$ - observed data

- Group 1:
  - n_1 - number of subjects in Group 1
  - $x_{11}, x_{12}, \ldots, x_{1n_1}$ - observed data

Review of Wilcoxon Rank Sum Tests

- Rank data from the entire two groups
- Group 0
  - n_0 - number of subjects in Group 0
  - $x_{01}, x_{02}, \ldots, x_{0n_0}$ - observed data of Group 0
  - $r_{01}, r_{02}, \ldots, r_{0n_0}$ - ranks of Group 0
- Group 1:
  - n_1 - number of subjects in Group 1
  - $x_{11}, x_{12}, \ldots, x_{1n_1}$ - observed data of Group 1
  - $r_{11}, r_{12}, \ldots, r_{1n_1}$ - ranks of Group 1

Review of Wilcoxon Rank Sum Tests

- Sum the ranks of each group

$$R_z = \sum_{i=1}^{n_z} R_{zi} \qquad z = 0,1 - \text{group indicator}$$

$$E(R_1) = \frac{n_1(n_1 + n_0 + 1)}{2}$$

$$E(R_0) = \frac{n_0(n_1 + n_0 + 1)}{2}$$

$$Var(R_z) = \frac{n_0 n_1(n_1 + n_0 + 1)}{12}$$

- The test statistics can be written as

$$\frac{R_1 - E(R_1)}{\sqrt{Var(R_1)}} \sim N(0,1)$$

Review of Wilcoxon Rank Sum Tests

- Re-write the Wilcoxon test in the Mann-Whitney test statistics

$$U_{ij} = \begin{cases} +1 & \text{if } x_{0i} > x_{1j} \\ 0 & \text{if } x_{0i} = x_{1j} \\ -1 & \text{if } x_{0i} < x_{1j} \end{cases}$$

$$U = \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} U_{ij}$$

$$R_1 = \frac{n_1(n_1 + n_0 + 1)}{2} + \frac{U}{2} \quad or$$

$$U = 2R_1 - n_1(n_1 + n_0 + 1)$$

$$E(U) = 0 \qquad Var(U) = \frac{n_0 n_1 (n_1 + n_0 + 1)}{3}$$

- The test statistics is

$$\frac{U}{\sqrt{Var(U)}} \sim N(0,1)$$

Review of Hypergeometric Distribution

- N fishes in a pond
  - N_0 are white
  - N_1 are golden
- Catch d fishes
  - D_0 are white
  - D_1 are golden
- 2x2 tables – all margins are fixed
- D_0 follows hypergeometric distribution
- $D_0 \sim Hyper(N, N_0, D)$

Class	Case	Non-case	Total
White	D_0	$N_0 - D_0$	N_0
Golden	D_1	$N_1 - D_1$	N_1
Total	D	$N - D$	N

Review of Hypergeometric Distribution

- $D_0 \sim Hyper(N, N_0, D)$
- $P(D_0 = d_0) = \frac{\binom{D}{d_0}\binom{N-D}{N_0-d_0}}{\binom{N}{N_0}}$
- $\sum_{0 \le d_0 \le \min(D, N_0)} \frac{\binom{D}{d_0}\binom{N-D}{N_0-d_0}}{\binom{N}{N_0}} = 1$

Class	Case	Non-case	Total
White	D_0	$N_0 - D_0$	N_0
Golden	D_1	$N_1 - D_1$	N_1
Total	D	$N - D$	N

Review of Hypergeometric Distribution

- Test independence of caught (case) vs not caught and fish color

- Let $o = D_0$ be the observed
- Let $e = E(D_0) = \frac{DN_0}{N}$ be the expected
- H_0: $o = e$ vs H_0: $o \neq e$

Class	Case	Non-case	Total
White	D_0	$N_0 - D_0$	N_0
Golden	D_1	$N_1 - D_1$	N_1
Total	D	$N - D$	N

- The test statistics

- $\frac{\left(D_0 - \frac{DN_0}{N}\right)^2}{Var(D_0)} \sim \chi_1^2$

- $Var(D_0) = \frac{N_0 N_1 D(N-D)}{N^2(N-1)}$

Review of Cochran-Mantel-Haenszel Test

- Test independence
  - Treatment vs. disease status
- Stratified 2x2 tables, *K* strata
  - Different age groups
  - Investigation centers
- Assume common odds ratio
- $H_0: odds\ ratio = 1$

Class	Case	Non-case	Total
White	D_{0i}	$N_{0i} - D_{0i}$	N_{0i}
Golden	D_{1i}	$N_{1i} - D_{1i}$	N_{1i}
Total	D_i	$N_i - D_i$	N_i

- The test statistics is $\frac{\sum_{i=1}^{k}\left(D_{0i}-\frac{D_{0i}N_{0i}}{N_i}\right)^2}{\sum_{i=1}^{k}\frac{D_i(N_i-D_i)N_{1i}N_{0i}}{{N_i}^2(N_i-1)}} \sim \chi_1^2$

Nonparametric Tests for Survival Data

Nonparametric Tests for Survival Data

- Compare two groups wrt the event time distributions
  - $S_0(t)$ - the control group
  - $S_1(t)$ - the new treatment group

- If the occurrence of the events in the new treatment group delayed in comparison to the control group

$H_0\colon S_1(t) \leq S_0(t)$vs. $H_0\colon S_1(t) > S_0(t)$

Log-rank Test

- AKA
  - Mantel-Haenszel test
  - Mantel-Cox test
  - Peto-Mantel-Haenszel test

- Proposed by
  - Mantel (1966)
  - Cox (1972)

The Log-rank Test

- Keep in mind that survival data (T, Δ)
  - $T = \min(X, C)$
  - $\Delta = I(X < C)$
  - X - event time
  - C – censor time

- Group 0
  - n_0 - number of subjects in Group 0
  - $(T_{01}, \delta_{01}), (T_{02}, \delta_{02}), \dots, (T_{0n_0}, \delta_{0n_0})$,- observed survival data in Group 0

- Group 1:
  - n_1 - number of subjects in Group 1
  - $(T_{11}, \delta_{11}), (T_{12}, \delta_{12}), \dots, (T_{1n_1}, \delta_{1n_1})$ - observed survival data in Group 1

- Number of events in both groups: r
  - Ordered distinct event time: $t_{(1)} < t_{(2)} < \cdots < t_{(k)}$
  - $k \leq r$

The Log-rank Test

- Construct k 2x2 tables at each distinct event time

- At i^{th} event time,
  - n_{0i} number of subjects at risk at $t_{(i)}^{-}$ in Group 0
  - n_{1i} number of subjects at risk at $t_{(i)}^{-}$in Group 1
  - $n_i = n_{0i} + n_{1i}$ total number of subjects at risk at $t_{(i)}^{-}$

  - d_{0i} number of subjects at risk at $t_{(i)}$ in Group 0
  - d_{1i} number of subjects at risk at $t_{(i)}$ in Group 1
  - $d_i = d_{0i} + d_{1i}$ total number of events at $t_{(i)}$

2x2 Table at the i^{th} Event Time

Group	Events occurred at $t_{(i)}$	Number of subjects Survival at $t_{(i)}^{+}$	Number of subject at risk at $t_{(i)}^{-}$
0	d_{0i}	$n_{0i} - d_{0i}$	n_{0i}
1	d_{1i}	$n_{1i} - d_{1i}$	n_{1i}
Total	d_i	$n_i - d_i$	n_i

- n_{0i}, n_{1i}, n_i are fixed at $t_{(i)}^{-}$
- d_i is fixed at $t_{(i)}$
- $d_{0i} \sim$ hypergeometric distribution

Hypergeometric Distribution

- Let D_{0i} be the number of events at $\mathrm{t_{0i}}$ in Group 0

$$P(D_{0i} = d_{0i}) = \frac{\binom{d_i}{d_{0i}}\binom{n_i - d_i}{n_{0i} - d_{0i}}}{\binom{n_i}{n_{0i}}}$$

- Combination $\binom{d_i}{d_{0i}}$ - probability of selecting d_{0i} out of d_i without ordering

$$\binom{d_i}{d_{0i}} = \frac{d_i!}{d_{0i}!\,(d_i - d_{0i})!}$$

$$d_i! = d_i \cdot d_{i-1} \cdots 2 \cdot 1$$

2x2 Table at the i^{th} Event Time

- The mean of D_{0i}

$$e_{0i} = n_{0i} d_i / n_i$$

The expected number of events if the survival probability of the two groups are the same – the null hypothesis

- The difference of observed d_{0i} and the expected e_{0i}

$$d_{0i} - e_{0i}$$

Represents deviation from the null hypothesis at the i^{th} event time

The Log-rank Test

- The total deviation from null

$$L = \sum_{i=1}^{k} (d_{0i} - e_{0i})$$

- Variance

$$\text{var}(L) = var\left(\sum_{i=1}^{k} (d_{0i} - e_{0i})\right) \approx \sum_{i=1}^{k} var(d_{0i})$$

From hypergeometric distribution

$$var(d_{0i}) = \frac{n_{0i} n_{1i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)}$$

$$\text{var}(L) = \sum_{i=1}^{k} \frac{n_{0i} n_{1i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)}$$

$$\frac{L}{\sqrt{var(L)}} \sim N(0,1)$$

The Log-rank Test

- If there is no tie at each event time, log-rank test statistics reduces to

$$\frac{\sum_{i=1}^{k}(d_{0i}-n_{0i}/n_i)}{\sum_{i=1}^{k} n_{0i}n_{1i}/{n_i}^2}$$

where d_{0i} can be 0 or 1

The Log-rank Test

- The numerators of the log-rank test is the difference of the observed and expected

$$\left(d_{0i} - \frac{n_{0i}d_i}{n_i}\right) = \frac{n_{0i}n_{1i}}{n_i}\left(\frac{d_{0i}}{n_{0i}} - \frac{d_{1i}}{n_{1i}}\right)$$

$$= \frac{n_{0i}n_{1i}}{n_i}(h_{0i} - h_{1i})$$

- Like the CMH test for several 2x2 tables, the log-rank test has the optimal power when the hazard ratio is constant over time, i.e.

$$h_0(t) = \lambda h_1(t) \quad \text{or}$$

$$S_0(t) = S_1(t)^{\lambda}$$

Example: Leukemia Data

- Group 0: 6+,6,6,6,7,9+,10+,10,11+,13,16,17+,19+,20+,22,23,25+,32+,32+,34+,35+
- Group 1: 1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23
- 17 Distinct event time: 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13,15, 16, 17, 22, 23

Group	Events at $t_{(i)}$	Survival at $t_{(i)}^{+}$	At risk at $t_{(i)}^{-}$
0	d_{0i}	$n_{0i} - d_{0i}$	n_{0i}
1	d_{1i}	$n_{1i} - d_{1i}$	n_{1i}
Total	d_i	$n_i - d_i$	n_i

Example: Leukemia Data

- 16 2x2 tables
- The first 7 2x2 tables

- Can be calculated

$$L = \sum_{i=1}^{k} (d_{0i} - e_{0i}) = -10.25$$

- Variance

$$\mathrm{var}(L) = \sum_{i=1}^{k} \frac{n_{0i} n_{1i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)} = 6.26$$

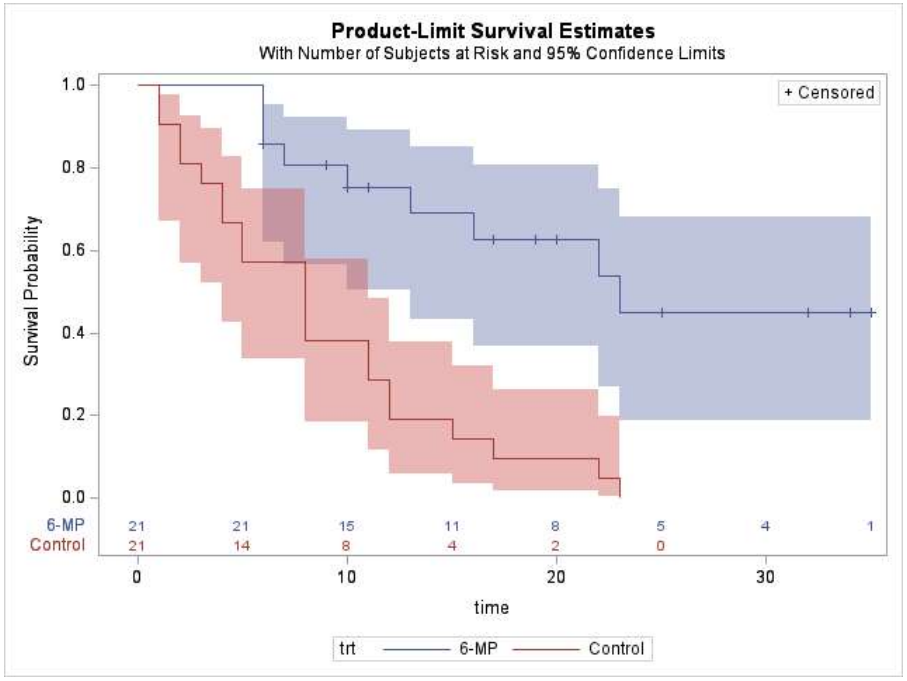
$$\frac{L}{\sqrt{\mathrm{var}(L)}} = 4.098 \sim N(0,1)$$

$$p - value = 2.1e^{-5} < 0.0001$$

i	Group	Events at $t_{(i)}$	Survival at $t_{(i)}^{+}$	Expected	At risk at $t_{(i)}^{-}$
1	0	0	21	1	21
	1	2	19	1	21
		2	40		42
2	0	0	21	1.05	21
	1	2	17	0.95	19
		2	38		40
3	0	0	21	0.55	21
	1	1	16	0.45	17
		1	37		38
4	0	0	21	1.14	21
	1	2	14	0.86	16
		2	35		37
5	0	0	21	1.2	21
	1	2	12	0.8	14
		2	33		35
6	0	3	19	1.91	21
	1	0	12	1.09	12
		3	31		33
7	0	1	16	0.6	18
	1	0	12	0.4	12
		1	29		30

Example: Leukemia Data

- SAS outcomes



Rank Statistics		
trt	Log-Rank	Wilcoxon
6-MP	-10.251	-271.00
Control	10.251	271.00

Covariance Matrix for the Log-Rank Statistics		
trt	6-MP	Control
6-MP	6.25696	-6.25696
Control	-6.25696	6.25696

Covariance Matrix for the Wilcoxon Statistics		
trt	6-MP	Control
6-MP	5457.11	-5457.11
Control	-5457.11	5457.11

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	16.7929	1	<.0001
Wilcoxon	13.4579	1	0.0002
-2Log(LR)	16.4852	1	<.0001

The Gehan's Wilcoxon Test

- With right censored data, extending Censoring in the Wilcoxon Test

$$U_{ij} = \begin{cases} +1 & \text{if } t_{0i} > t_{1j}\,;\delta_{1j} = 1 \\ 0 & \text{if otherwise} \\ -1 & \text{if } t_{0i} < t_{1j}\,;\delta_{0i} = 1 \end{cases}$$

$$GW = \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} U_{ij}$$

Under Null

$$E_{H_0}(GW) = 0$$

$$\frac{GW}{\sqrt{Var(GW)}} \sim N(0,1)$$

The Gehan's Wilcoxon Test

- The generalized Wilcoxon procedure proposed by
  - Gehan (1965)
  - Breslow (1970)

- The test statistics can also be formulated as

$$W = \sum_{i=1}^{k} n_i (d_{0i} - e_{0i})$$

- Weighted by n_i, the total number of subjects at risk at $t_{(i)}^{-}$
- The variance

$$\mathrm{var}(W) = var\left(\sum_{i=1}^{k} n_i (d_{0i} - e_{0i}) \right) = \sum_{i=1}^{k} \frac{n_{0i} n_{1i} d_i (n_i - d_i)}{(n_i - 1)}$$

- A weighted log-rank test

General Class of Weighted Log-rank Tests

- Can be written as

$$L_w = \sum_{i=1}^{k} \omega_i (d_{0i} - e_{0i})$$

$$\text{var}(L_w) = \sum_{i=1}^{k} \omega_i^2 \frac{n_{0i} n_{1i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)}$$

Test	Weight ω_i
Log-rank	$\omega_i = 1$
Gehan's Wilcoxon	$\omega_i = n_i$
Peto/Prentice	$\omega_i = S(t_i)$
Fleming-Harrington	$\omega_i = S(t_i)^\rho \quad \rho \geq 0$
Tarone-Ware	$\omega_i = \sqrt{n_i}$

Comparisons

- All tests will control type I error under null
- May obtain optimal power under certain alternatives
- The Log-rank test
  - Tends to be sensitive to distributional differences which are separated late in time
  - Has the optimal power when the hazard ratio is constant

- The Wilcoxon test
  - Tends to be more powerful in detecting early separation between survival functions

Reference: A pretest for choosing between logrank and wilcoxon tests in the two-sample problem. International Journal of Statistics, 2010, vol. LXVIII, n. 2, pp. 111-125

Example – Ovarian Data

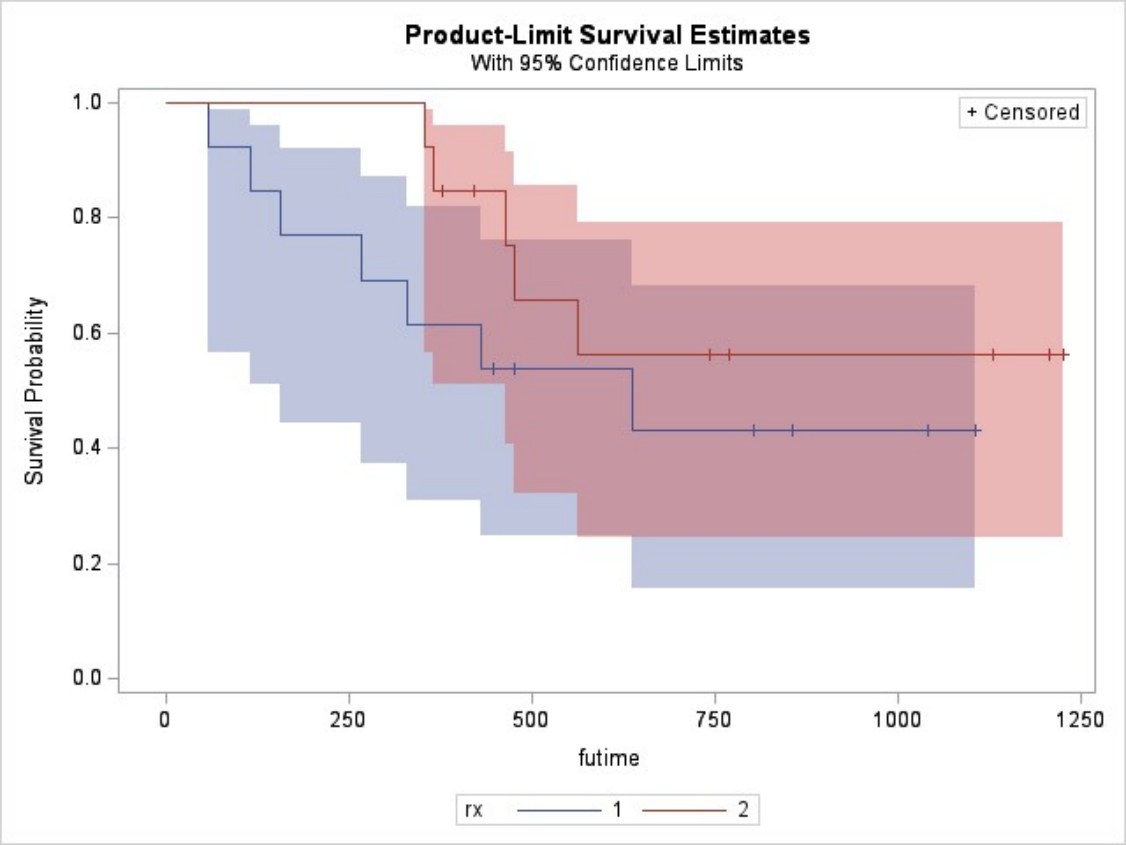
	futime	fustat	age	resid.ds	rx	ecog.ps
1	59	1	72.3315	2	1	1
2	115	1	74.4932	2	1	1
3	156	1	66.4658	2	1	2
4	421	0	53.3644	2	2	1
5	431	1	50.3397	2	1	1
6	448	0	56.4301	1	1	2
7	464	1	56.9370	2	2	2
8	475	1	59.8548	2	2	2
9	477	0	64.1753	2	1	1
10	563	1	55.1781	1	2	2
11	638	1	56.7562	1	1	2
12	744	0	50.1096	1	2	1
13	769	0	59.6301	2	2	2
14	770	0	57.0521	2	2	1
15	803	0	39.2712	1	1	1
16	855	0	43.1233	1	1	2
17	1040	0	38.8932	2	1	2
18	1106	0	44.6000	1	1	1
19	1129	0	53.9068	1	2	1
20	1206	0	44.2055	2	2	1
21	1227	0	59.5890	1	2	2
22	268	1	74.5041	2	1	2
23	329	1	43.1370	2	1	1
24	353	1	63.2192	1	2	2
25	365	1	64.4247	2	2	1
26	377	0	58.3096	1	2	1

- Dataset available in R survival package
  - futime: survival or censoring time (day)
  - fustat: censoring status (censor=0)
  - age: in years
  - resid.ds: residual disease present (1=no,2=yes)
  - rx: treatment group
  - ecog.ps: ECOG performance status (1 is better, see reference)

Examples

```
ods graphics on;
proc lifetest data=example method=KM plot=survival(cl) ;
  time futime*fustat(0);
  strata rx/test=all;
  run;
ods graphics off;
```

Example- Ovarian Data (Non-stratified)



Example- Ovarian Data (Non-stratified)

Rank Statistics		
rx	Log-Rank	Wilcoxon
1	1.7665	47.000
2	-1.7665	-47.000

Covariance Matrix for the Log-Rank Statistics		
rx	1	2
1	2.93620	-2.93620
2	-2.93620	2.93620

Covariance Matrix for the Wilcoxon Statistics		
rx	1	2
1	1154.00	-1154.00
2	-1154.00	1154.00

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	1.0627	1	0.3026
Wilcoxon	1.9142	1	0.1665
-2Log(LR)	1.1149	1	0.2910

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	1.0627	1	0.3026
Wilcoxon	1.9142	1	0.1665
Tarone	1.4852	1	0.2230
Peto	1.6990	1	0.1924
Modified Peto	1.7431	1	0.1867
Fleming(1)	1.6849	1	0.1943

Stratified Log-rank Test

- When comparing 2 groups, it is desirable to test between group within certain subgroups formed by factors such as
  - Disease severity
- The reason is that the background risk (in the control group) among the different factors are different
  - The risk in the mild patient population is different from that in the severe population
- The stratification factor for disease severity
  - Mild
  - Moderate
  - Severe

Stratified Log-rank Test

- Let $s = 1, 2, \ldots, S$ be the stratum for a factor
- $h_{0s}(t)$ is the hazard of the s stratum in the control group

$$h_{0i}(t) \neq h_{0j}(t) \quad \text{for } i \neq j \; i, j = 1,2, \ldots, s$$

- Assume the same treatment effect

$$\lambda = \frac{h_{1s}(t)}{h_{0s}(t)}$$

- The stratified log-rank test
  - Compares treatment effect between groups within each stratum first
  - Combine the comparisons of all strata

2x2 Table at the i^{th} Event Time in Stratum s

Group	Events occurred at $t_{(i)}$	Number of subjects Survival at $t_{(i)}^{+}$	Number of subject at risk at $t_{(i)}^{-}$
0	d_{s0i}	$n_{s0i} - d_{s0i}$	n_{s0i}
1	d_{s1i}	$n_{s1i} - d_{s1i}$	n_{s1i}
Total	d_{si}	$n_{si} - d_{si}$	n_{si}

Stratified Log-rank Test

- Log-rank statistics in Stratum s, r'_s distinct survival time in Stratum s

$$L_s = \sum_{i=1}^{r'_s} (d_{s0i} - e_{s0i})$$

$$\mathrm{var}(L_s) = \sum_{i=1}^{r'_s} \frac{n_{0i} n_{1i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)}$$

- The stratified log-rank test is

$$LS = \frac{\sum_{s=1}^{S} L_s}{\sqrt{\sum_{s=1}^{S} var(L_s)}}$$

Example – Ovarian Data

	futime	fustat	age	resid.ds	rx	ecog.ps
1	59	1	72.3315	2	1	1
2	115	1	74.4932	2	1	1
3	156	1	66.4658	2	1	2
4	421	0	53.3644	2	2	1
5	431	1	50.3397	2	1	1
6	448	0	56.4301	1	1	2
7	464	1	56.9370	2	2	2
8	475	1	59.8548	2	2	2
9	477	0	64.1753	2	1	1
10	563	1	55.1781	1	2	2
11	638	1	56.7562	1	1	2
12	744	0	50.1096	1	2	1
13	769	0	59.6301	2	2	2
14	770	0	57.0521	2	2	1
15	803	0	39.2712	1	1	1
16	855	0	43.1233	1	1	2
17	1040	0	38.8932	2	1	2
18	1106	0	44.6000	1	1	1
19	1129	0	53.9068	1	2	1
20	1206	0	44.2055	2	2	1
21	1227	0	59.5890	1	2	2
22	268	1	74.5041	2	1	2
23	329	1	43.1370	2	1	1
24	353	1	63.2192	1	2	2
25	365	1	64.4247	2	2	1
26	377	0	58.3096	1	2	1

- Dataset available in R survival package
  - futime: survival or censoring time (day)
  - fustat: censoring status (censor=0)
  - age: in years
  - resid.ds: residual disease present (1=no,2=yes)
  - rx: treatment group
  - ecog.ps: ECOG performance status (1 is better, see reference)
- Ecog performance status as strata

SAS Code

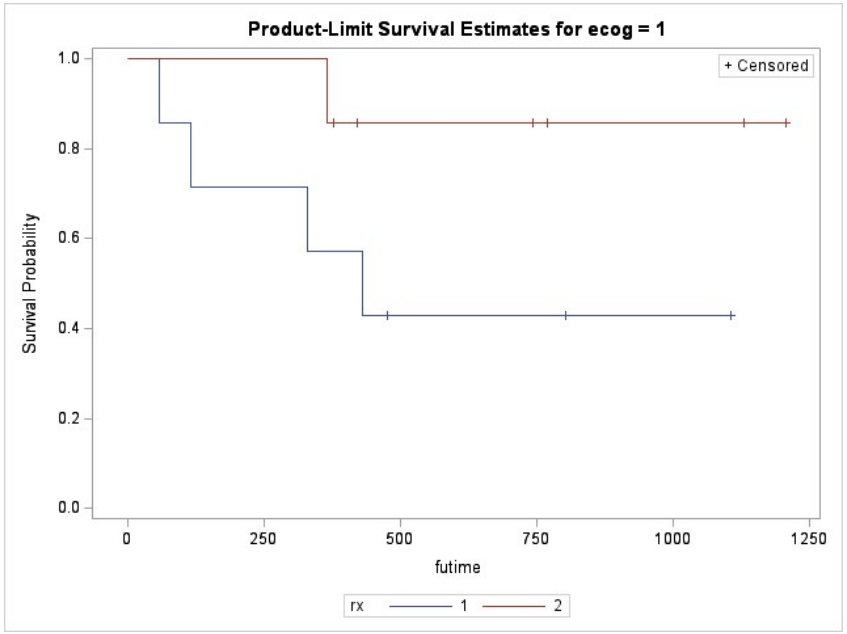
Stratified Logrank

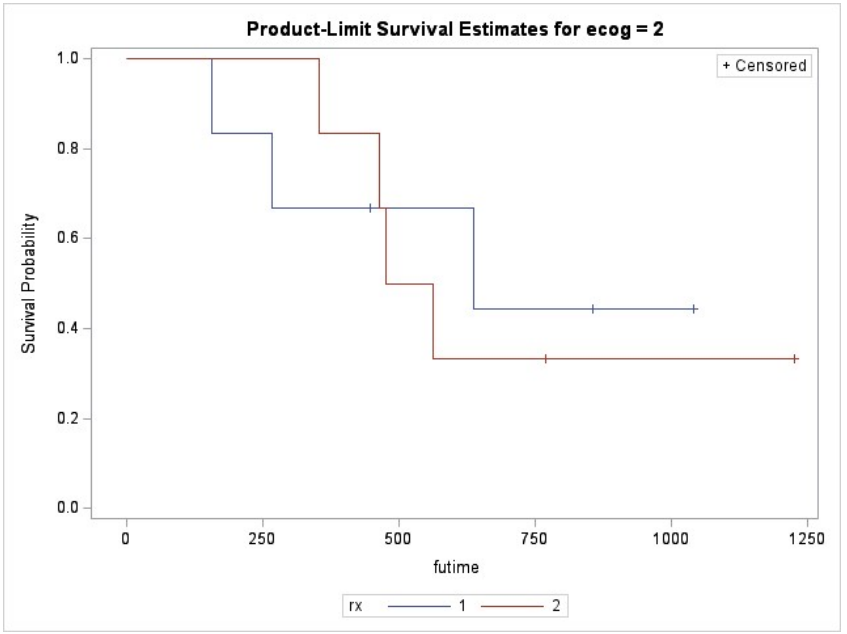
```
ods graphics on;
proc lifetest data=example method=KM
/*plot=survival(cl)*/ ;
  time futime*fustat(0);
  strata ecog /group=rx;
  run;
ods graphics off;
```

Non-stratified Logrank

```
ods graphics on;
proc lifetest data=example method=KM
plot=survival(cl) ;
  time futime*fustat(0);
  strata rx;
  run;
ods graphics off;
```

Example- Ovarian Data





Example- Ovarian Data

Summary of the Number of Censored and Uncensored Values						
Stratum	ecog	rx	Total	Failed	Censored	Percent Censored
1	1	1	7	4	3	42.86
1	1	2	7	1	6	85.71
Subtotal			14	5	9	64.29
2	2	1	6	3	3	50.00
2	2	2	6	4	2	33.33
Subtotal			12	7	5	41.67
Total			26	12	14	53.85

Example- Ovarian Data (Stratified)

Stratified Test of Equality over Group			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	0.7679	1	0.3809
Wilcoxon	1.6026	1	0.2055
Tarone	1.1728	1	0.2788
Peto	1.3372	1	0.2475
Modified Peto	1.4180	1	0.2337
Fleming(1)	1.3119	1	0.2521

Stratified Comparison of Survival Curves for futime over Group

Rank Statistics		
rx	Log-Rank	Wilcoxon
1	1.5000	22.000
2	-1.5000	-22.000

Covariance Matrix for the Log-Rank Statistics		
rx	1	2
1	2.93019	-2.93019
2	-2.93019	2.93019

Covariance Matrix for the Wilcoxon Statistics		
rx	1	2
1	302.000	-302.000
2	-302.000	302.000

Stratified Test of Equality over Group			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	0.7679	1	0.3809
Wilcoxon	1.6026	1	0.2055

Comparison of Three or More Groups

- May consider evaluate survival functions of K independent samples: $S_0(t), S_1(t), \cdots, S_K(t)$
  - $K-1$ groups in comparison with one control group
  - From K country

- The null hypothesis is that all the K survival functions are the same

$$H_0: \ S_1(t) = S_2(t) = \cdots = S_K(t)$$
$$H_A: \ not(\ S_1(t) = S_2(t) = \cdots = S_K(t))$$

- Group k
  - n_k - number of subjects in Group k
  - $(T_{k1}, \delta_{k1}), (T, \delta_{k2}), \ldots, (T_{kn_k}, \delta_{kn_k})$ are observed survival data in Group k
  - $k = 1, \ldots, K$

- Number of events in *K* groups: r
  - Ordered distinct event time: $t_{(1)} < t_{(2)} < \cdots < t_{(r')}$
  - $r' \leq r$

K x2 Table at the i^{th} Event Time

Group	Events occurred at $t_{(i)}$	Number of subjects Survival at $t_{(i)}^{+}$	Number of subject at risk at $t_{(i)}^{-}$
1	d_{1i}	$n_{1i} - d_{1i}$	n_{1i}
2	d_{2i}	$n_{2i} - d_{2i}$	n_{2i}
…			
K	d_{Ki}	$n_{Ki} - d_{Ki}$	n_{Ki}
Total	d_i	$n_i - d_i$	n_i

At The i^{th} Event Time

- Let

$$O_i = \begin{pmatrix} d_{1i} \\ \vdots \\ d_{Ki} \end{pmatrix}, \qquad E_i = \begin{pmatrix} n_{1i} \\ \vdots \\ n_{Ki} \end{pmatrix} \frac{d_i}{n_i}, \quad V_i = \begin{pmatrix} v_{11i} & v_{12i} & \cdots & v_{1Ki} \\ & v_{22i} & \cdots & v_{2Ki} \\ & & \cdots & \\ & & & v_{2Ki} \end{pmatrix}$$

- The test statistics can be written as

$$(O_i - E_i)' V_i^{-1} (O_i - E_i) \sim \chi^2(K)$$

- Omnibus test
  - To test if there is any difference among the survival functions
  - Not for pairwise comparisons
  - Not powerful to detect trend

Trend Log-rank Tests

- Let us assume there is a trend among the $(K+1)$ independent samples
  - K different doses
  - Possible to have monotone dose-response relationship

- A trend test is to test the hypothesis against a specific alternatives

$$H_0: \ S_1(t) = S_2(t) = \cdots = S_K(t)$$

$$H_A: \ S_1(t) \leq S_2(t) \leq \cdots \leq S_K(t))$$

Trend Log-rank Tests

- Log-rank statistics Kth dose

$$L_k = \sum_{i=1}^{r'_k} (d_{0i} - e_{ki})$$

$$\mathrm{var}(L_k) = \sum_{i=1}^{r'_k} \frac{n_{0i} n_{ki} d_{ki} (n_{0ki} - d_{0ki})}{n_{0ki}^2 (n_{0ki} - 1)}$$

- The trend log-rank test is

$$LS = \frac{\sum_{k=1}^{K} \omega_k L_k}{\sqrt{\sum_{k=1}^{K} (\omega_k - \bar{\omega})^2 var(L_k)}} \sim \chi_1^2$$

$$\bar{\omega} = \frac{\sum_{k=1}^{K} \omega_k e_k}{\sum_{k=1}^{K} e_k}$$

where ω_k is a weight for dose *k*

Trend Tests

- How to choose weight ω_k
- For monotone relationship

$$\omega_k = k \; for \; k = 1, \ldots, K$$

- When $K = 3$
  - $\omega_1 = 1, \omega_2 = 2, \omega_3 = 3$
  - $\omega_1 = -1, \omega_2 = 0, \omega_3 = 1$

Homework 4

1. Using the Leukemia data
   - Group 0: 6+,6,6,6,7,9+,10+,10,11+,13,16,17+,19+,20+,22,23,25+,32+,32+,34+,35+
   - Group 1: 1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23
   - 17 Distinct event time: 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 15, 16, 17, 22, 23

   a) Construct the 16 2x2 tables for the log-rank test

   b) Verify the rank statistics by calculating the difference of the observed and expected in each 2x2 tables

$$L = \sum_{i=1}^{k} (d_{0i} - e_{0i}) = -10.25$$

   c) Include the variance calculation for each of the 2x2 tables as well and verify that

$$\mathrm{var}(L) = \sum_{i=1}^{k} \frac{n_{0i} n_{1i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)} = 6.26$$

2. Outline steps how you would show the following using simulation (no actual simulation is required)
   a) The Wilcoxon has the optimal power when the failure times are log-normally distributed, with equal variance in both groups but different means.
   b) With weights $\omega_i = S(t_i)$, the test is most powerful under the alternative hypothesis of log-logistic model

3. Results from a trial evaluating a new treatment in comparison with a standard of care (SOC) indicate that biomark+ subgroup will have positive survival benefit with the new treatment and biomarker- subgroup actually gets harm.
   a) Please discuss what you think of an analysis by including all subject? Overall analyses including all subjects can be stratified or non-stratified analyses.
   b) How would you recommend the analysis?
   c) Do you think if FDA should approve this drug?