Chapter 11: Survival Analysis

Dr. Daniel B. Rowe
Professor of Computational Statistics
Department of Mathematical and Statistical Sciences
Marquette University

Time to Event

Survival analysis is the statistical analysis of **time-to-event** variables.

The event could be a heart attack, cancer remission, or death.

What is the probability that a participant survives 5 years?

Are there differences in survival between groups?

How do certain characteristics affect participants chances of survival?

Time to Event

Not all participants enroll when a study begins, so when a study ends, not all participants were enrolled for the same amount of time.

True survival time (failure time) is not known because the study ended before the event or participants dropped out.

The last observed follow-up is called the censored or censoring time.

Right censoring is when a participant does not have the event of interest during the study, last observed follow-up is less than the time to event.

Time to Event

Patients experiences with Myocardial Infarction over 10 years

Some join up to 2 years after start, all are followed for 10 years from start.
- 3 Myocardial Infarction
- 1 death
- 2 drop out
- 4 completions

All enrolled at the same time, all are followed for 10 years from start.
- 3 Myocardial Infarction
- 1 death
- 2 drop out
- 4 completions

All enrolled at the same time, all are followed for 10 years from start.
- 3 Myocardial Infarction
- 1 death
- 2 drop out
- 4 completions

Survival Analysis analyzes not only the number of MI events, but also times.

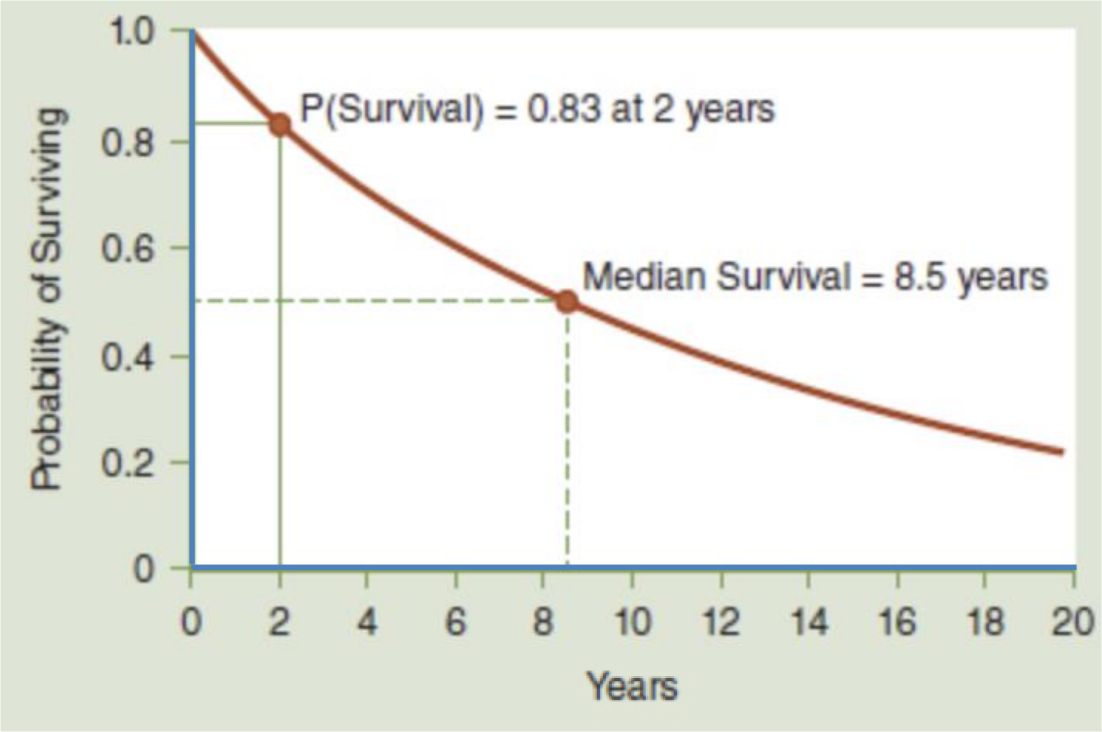
11.1 Introduction to Survival Data

Survival analysis measures two pieces of information

1) Whether the event occurred, 1=yes, 0=no
2) Last follow-up time, from enrollment.

The **survival function** is the probability a person survives past a time t.

Myocardial Infarction Study



t=0.0 : survival probability=1.00

t=2.0 : survival probability=0.83

t=8.5 : survival probability=0.50 (Median)

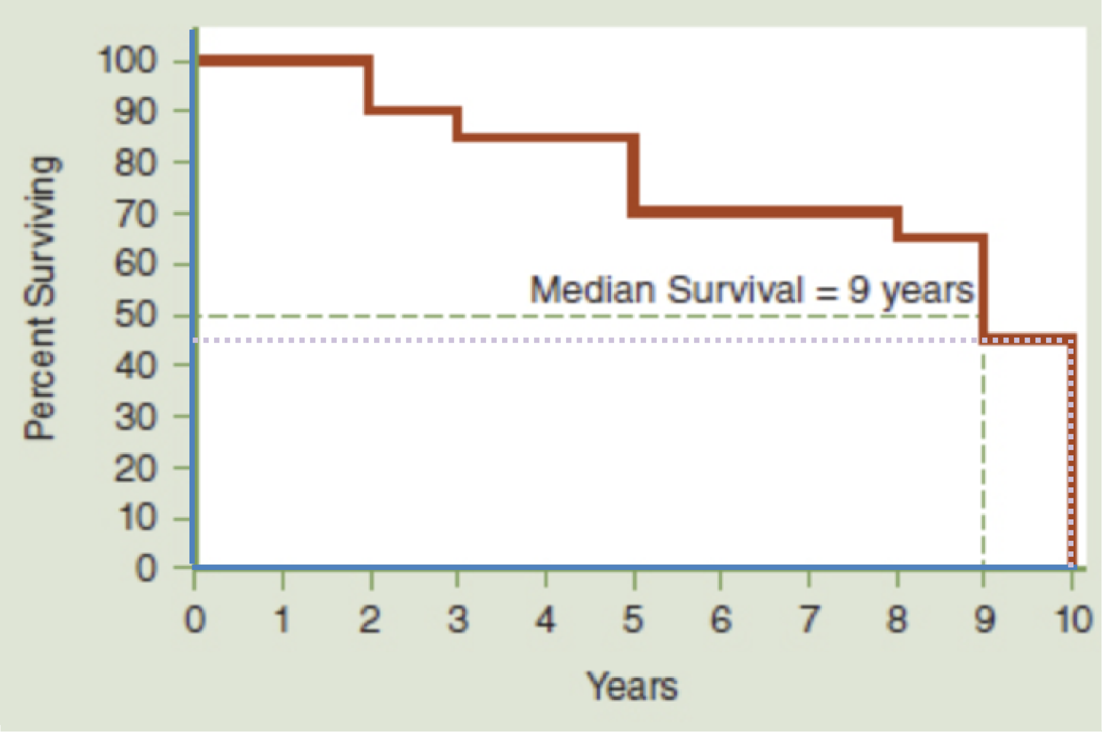
t=10.0: survival probability=0.47

11.2 Estimating the Survival Function

There are several parametric and nonparametric ways to estimate survival curves.

Let’s examine nonparametric step survival curves.

Time on x axis and survival (percentage) at risk on y axis.



t=0.0 : survival probability=1.00

t=2.0 : survival probability=0.90

t=9.0 : survival probability=0.50 (Median)

t=10.0: survival probability=0.45

11.2 Estimating the Survival Function

Example of 24 year study with 20 participants.
Some die, many drop out, few finish.

Participant	Year of Death	Last Contact
1		24
2	3	
3		11
4		19
5		24
6		13
7	14	
8		2
9		18
10		17
11		24
12		21
13		12
14	1	
15		10
16	23	
17		6
18	5	
19		9
20	17	

Original Data

11.2 Estimating the Survival Function - Actuarial

We can organize the data into a simple table.
Divide the 24 year study into 5 year intervals.

0- 4 years
5- 9 years
10-14 years
15-19 years
20-24 years.

Participant	Year of Death	Last Contact
1		24
2	3	
3		11
4		19
5		24
6		13
7	14	
8		2
9		18
10		17
11		24
12		21
13		12
14	1	
15		10
16	23	
17		6
18	5	
19		9
20	17	

Original Data

11.2 Estimating the Survival Function - Actuarial

We can organize the data into a simple table.
Divide the 24 year study into 5 year intervals.

Interval	Participant	Year of Death	Last Contact
0- 4 years	14	1	
	8		2
	2	3	
5- 9 years	18	5	
	17		6
	19		9
10-14 years	15		10
	3		11
	13		12
	6		13
	7	14	
15-19 years	10		17
	20	17	
	9		18
	4		19
20-24 years	12		21
	16	23	
	1		24
	5		24
	11		24

Time Data

11.2 Estimating the Survival Function - Actuarial

We can organize the data into a simple table.
Divide the 24 year study into 5 year intervals.

Count *Alive* at beginning of each interval.
Count how many *Deaths* during interval.
Number *censored* (dropped out) in each interval.

Interval	Participant	Year of Death	Last Contact
0- 4 years	14	1	
	8		2
	2	3	
5- 9 years	18	5	
	17		6
	19		9
10-14 years	15		10
	3		11
	13		12
	6		13
	7	14	
15-19 years	10		17
	20	17	
	9		18
	4		19
20-24 years	12		21
	16	23	
	1		24
	5		24
	11		24

Time Data

11.2 Estimating the Survival Function - Actuarial

We can organize the data into a simple table.
Divide the 24 year study into 5 year intervals.

Count *Alive* at beginning of each interval.
Count how many *Deaths* during interval.
Number *censored* (dropped out) in each interval.

Interval in Years	Number Alive at Beginning of Interval	Number of Deaths During Interval	Number Censored
0–4	20	2	1
5–9	17	1	2
10–14	14	1	4
15–19	9	1	3
20–24	5	1	4

Form Table

	Participant	Year of Death	Last Contact
0- 4 years	14	1	
	8		2
	2	3	
5- 9 years	18	5	
	17		6
	19		9
10-14 years	15		10
	3		11
	13		12
	6		13
	7	14	
15-19 years	10		17
	20	17	
	9		18
	4		19
20-24 years	12		21
	16	23	
	1		24
	5		24
	11		24

Time Data

11.2 Estimating the Survival Function - Actuarial

Life Tables (actuarial tables)

N_t = number event free during interval t
(Number at risk)

D_t = number who die during interval t

C_t = number censored during interval t

N_{t*}= average number at risk during interval t

Deaths assumed to occur at end of the interval.

Censored events assumed occur evenly in interval.

$N_{t*}=N_t - C_t/2$

	Participant	Year of Death	Last Contact
0- 4 years	14	1	
	8		2
	2	3	
5- 9 years	18	5	
	17		6
	19		9
10-14 years	15		10
	3		11
	13		12
	6		13
	7	14	
15-19 years	10		17
	20	17	
	9		18
	4		19
20-24 years	12		21
	16	23	
	1		24
	5		24
	11		24

Time Data

11.2 Estimating the Survival Function - Actuarial

Life Tables (actuarial tables)

N_t = number event free during interval t
(Number at risk)

D_t = number who die during interval t

C_t = number censored during interval t

N_{t*}= average number at risk during interval t

$$N_{t*} = N_t - C_t/2$$

q_t = prop. die in interval t, $q_t = D_t/N_{t*}$

p_t = prop. survive in interval t, $p_t = 1 - q_t$

S_t = prop. survive past interval t, $S_{t+1} = p_{t+1} S_t$

Interval	Participant	Year of Death	Last Contact
0- 4 years	14	1	
	8		2
	2	3	
5- 9 years	18	5	
	17		6
	19		9
10-14 years	15		10
	3		11
	13		12
	6		13
	7	14	
15-19 years	10		17
	20	17	
	9		18
	4		19
20-24 years	12		21
	16	23	
	1		24
	5		24
	11		24

Time Data

11.2 Estimating the Survival Function - Actuarial

Interval in Years	Number at Risk During Interval, N_t	Average Number at Risk During Interval, N_{t*}	Number of Deaths During Interval, D_t	Lost to Follow-Up, C_t	Proportion Dying During Interval, q_t	Among Those at Risk, Proportion Surviving Interval, p_t	Survival Probability, S_t
0–4	20	20 – (1/2) = 19.5	2	1	2/19.5 = 0.103	1 – 0.103 = 0.897	1(0.897) = 0.897

$N_{t*} = N_t – C_t/2$ $q_t=D_t/N_{t*}$ $p_t=1–q_t$ $S_{t+1}=p_{t+1}S_t$

Interval in Years	Number Alive at Beginning of Interval	Number of Deaths During Interval	Number Censored
0–4	20	2	1
5–9	17	1	2
10–14	14	1	4
15–19	9	1	3
20–24	5	1	4

N_t = # event free during interval t (Number at risk)

D_t = # who die during interval t

C_t = # censored during interval t

N_{t*} = avg. # at risk during interval t, $N_{t*} = N_t – C_t/2$

q_t = prop. die in interval t, $q_t=D_t/N_{t*}$

p_t = prop. survive in interval t, $p_t=1–q_t$

S_t = prop. survive past interval t, $S_{t+1}=p_{t+1}S_t$

	Participant	Year of Death	Last Contact
0- 4 years	14	1	
	8		2
	2	3	
5- 9 years	18	5	
	17		6
	19		9
10-14 years	15		10
	3		11
	13		12
	6		13
	7	14	
15-19 years	10		17
	20	17	
	9		18
	4		19
20-24 years	12		21
	16	23	
	1		24
	5		24
	11		24

Time Data

11.2 Estimating the Survival Function - Actuarial

Interval in Years	Number at Risk During Interval, N_t	Average Number at Risk During Interval, N_{t*}	Number of Deaths During Interval, D_t	Lost to Follow-Up, C_t	Proportion Dying During Interval, q_t	Among Those at Risk, Proportion Surviving Interval, p_t	Survival Probability, S_t
0–4	20	20 – (1/2) = 19.5	2	1	2/19.5 = 0.103	1 – 0.103 = 0.897	1(0.897) = 0.897
5–9	17	17 – (2/2) = 16.0	1	2	1/16 = 0.063	1 – 0.063 = 0.937	(0.937)(0.897) = 0.840

$N_{t*} = N_t – C_t/2$ $\quad q_t = D_t/N_{t*}$ $\quad p_t = 1 – q_t$ $\quad S_{t+1} = p_{t+1}S_t$

Interval in Years	Number Alive at Beginning of Interval	Number of Deaths During Interval	Number Censored
0–4	20	2	1
5–9	17	1	2
10–14	14	1	4
15–19	9	1	3
20–24	5	1	4

N_t = # event free during interval t
(Number at risk)

D_t = # who die during interval t

C_t = # censored during interval t

N_{t*} = avg. # at risk during interval t, $N_{t*} = N_t – C_t/2$

q_t = prop. die in interval t, $q_t = D_t/N_{t*}$

p_t = prop. survive in interval t, $p_t = 1 – q_t$

S_t = prop. survive past interval t, $S_{t+1} = p_{t+1}S_t$

Interval	Participant	Year of Death	Last Contact
0- 4 years	14	1	
	8		2
	2	3	
5- 9 years	18	5	
	17		6
	19		9
10-14 years	15		10
	3		11
	13		12
	6		13
	7	14	
15-19 years	10		17
	20	17	
	9		18
	4		19
20-24 years	12		21
	16	23	
	1		24
	5		24
	11		24

Time Data

11.2 Estimating the Survival Function - Actuarial

Interval in Years	Number at Risk During Interval, N_t	Average Number at Risk During Interval, N_{t^*}	Number of Deaths During Interval, D_t	Lost to Follow-Up, C_t	Proportion Dying During Interval, q_t	Among Those at Risk, Proportion Surviving Interval, p_t	Survival Probability, S_t
0–4	20	19.5	2	1	0.103	0.897	0.897
5–9	17	16.0	1	2	0.063	0.937	0.840
10–14	14	12.0	1	4	0.083	0.917	0.770
15–19	9	7.5	1	3	0.133	0.867	0.688
20–24	5	3.0	1	4	0.333	0.667	0.446

$N_{t^*} = N_t - C_t/2$ $\quad q_t = D_t/N_{t^*}$ $\quad p_t = 1 - q_t$ $\quad S_{t+1} = p_{t+1} S_t$

Interval in Years	Number Alive at Beginning of Interval	Number of Deaths During Interval	Number Censored
0–4	20	2	1
5–9	17	1	2
10–14	14	1	4
15–19	9	1	3
20–24	5	1	4

N_t = # event free during interval t
(Number at risk)

D_t = # who die during interval t

C_t = # censored during interval t

N_{t^*} = avg. # at risk during interval t, $N_{t^*} = N_t - C_t/2$

q_t = prop. die in interval t, $q_t = D_t/N_{t^*}$

p_t = prop. survive in interval t, $p_t = 1 - q_t$

S_t = prop. survive past interval t, $S_{t+1} = p_{t+1} S_t$

	Participant	Year of Death	Last Contact
0- 4 years	14	1	
	8		2
	2	3	
5- 9 years	18	5	
	17		6
	19		9
10-14 years	15		10
	3		11
	13		12
	6		13
	7	14	
15-19 years	10		17
	20	17	
	9		18
	4		19
20-24 years	12		21
	16	23	
	1		24
	5		24
	11		24

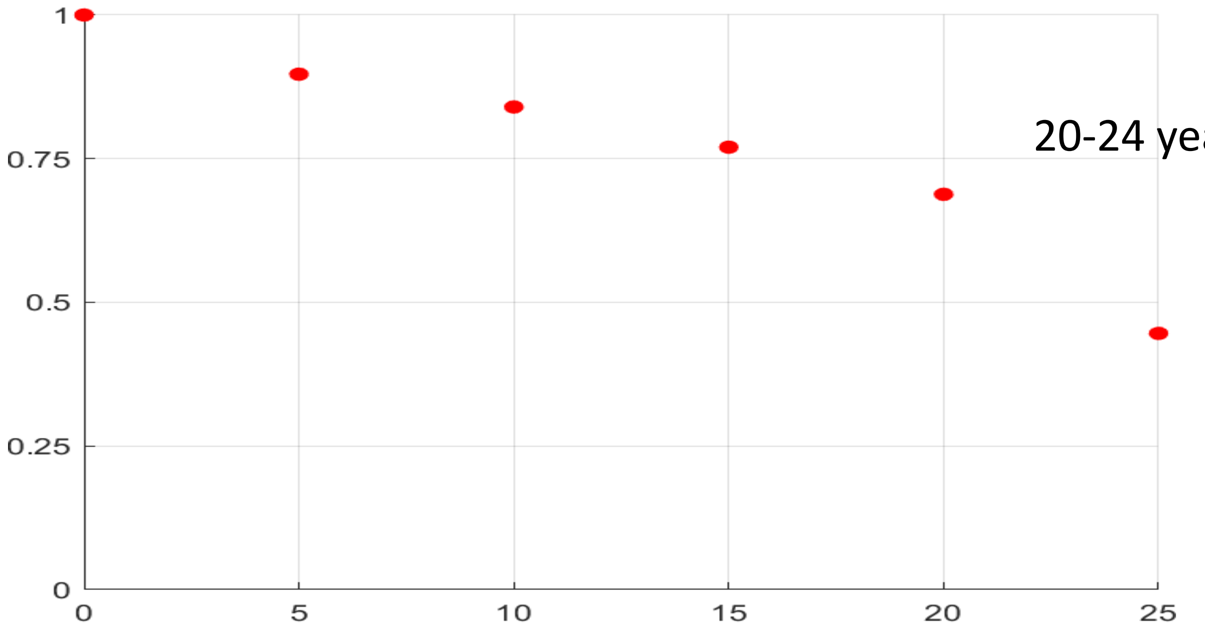
Time Data

11.2 Estimating the Survival Function - Actuarial

Interval in Years	Number at Risk During Interval, N_t	Average Number at Risk During Interval, N_{t*}	Number of Deaths During Interval, D_t	Lost to Follow-Up, C_t	Proportion Dying During Interval, q_t	Among Those at Risk, Proportion Surviving Interval, p_t	Survival Probability, S_t
0–4	20	19.5	2	1	0.103	0.897	0.897
5–9	17	16.0	1	2	0.063	0.937	0.840
10–14	14	12.0	1	4	0.083	0.917	0.770
15–19	9	7.5	1	3	0.133	0.867	0.688
20–24	5	3.0	1	4	0.333	0.667	0.446

$N_{t*} = N_t - C_t/2$ $\quad q_t = D_t/N_{t*}$ $\quad p_t = 1 - q_t$ $\quad S_{t+1} = p_{t+1} S_t$

Interval in Years	Number Alive at Beginning of Interval	Number of Deaths During Interval	Number Censored
0–4	20	2	1
5–9	17	1	2
10–14	14	1	4
15–19	9	1	3
20–24	5	1	4



Interval	Participant	Year of Death	Last Contact
0- 4 years	14	1	
	8		2
	2	3	
5- 9 years	18	5	
	17		6
	19		9
10-14 years	15		10
	3		11
	13		12
	6		13
	7	14	
15-19 years	10		17
	20	17	
	9		18
	4		19
20-24 years	12		21
	16	23	
	1		24
	5		24
	11		24

Time Data

At ends of intervals, depends on intervals.

11.2 Estimating the Survival Function - Kaplan-Meier

Kaplan-Meier Survival Curve approach re-estimates the probability each time an event occurs. Re-estimates every death or censoring.

Assumes censoring is independent of the likelihood of developing the event of interest. You don’t drop out because you don’t think you will ever get the event or because you know you will get it. You drop out because you are too busy or move.

Survival probabilities are comparable in participants who are recruited earlier as well as later. How participants are recruited doesn’t change.

11.2 Estimating the Survival Function - Kaplan-Meier

Estimating Survival Curve

Time, years	Number at Risk, N_t	Number of Deaths, D_t	Number Censored, C_t	Survival Probability, $S_{t+1} = S_t \times ((N_{t+1} - D_{t+1})/N_{t+1})$
0	20			1†
1	20	1		1 × [(20 – 1)/20] = 0.950
2	19		1	0.950 × [(19 – 0)/19] = 0.950
3	18	1		0.950 × [(18 – 1)/18] = 0.897
5	17	1		0.897 × [(17 – 1)/17] = 0.844
6	16		1	0.844
9	15		1	0.844
10	14		1	0.844
11	13		1	0.844
12	12		1	0.844
13	11		1	0.844
14	10	1		0.760
17	9	1	1	0.676
18	7		1	0.676
19	6		1	0.676
21	5		1	0.676
23	4	1		0.507
24	3		3	0.507

Participant	Year of Death	Last Contact
14	1	
8		2
2	3	
18	5	
17		6
19		9
15		10
3		11
13		12
6		13
7	14	
10		17
20	17	
9		18
4		19
12		21
16	23	
1		24
5		24
11		24

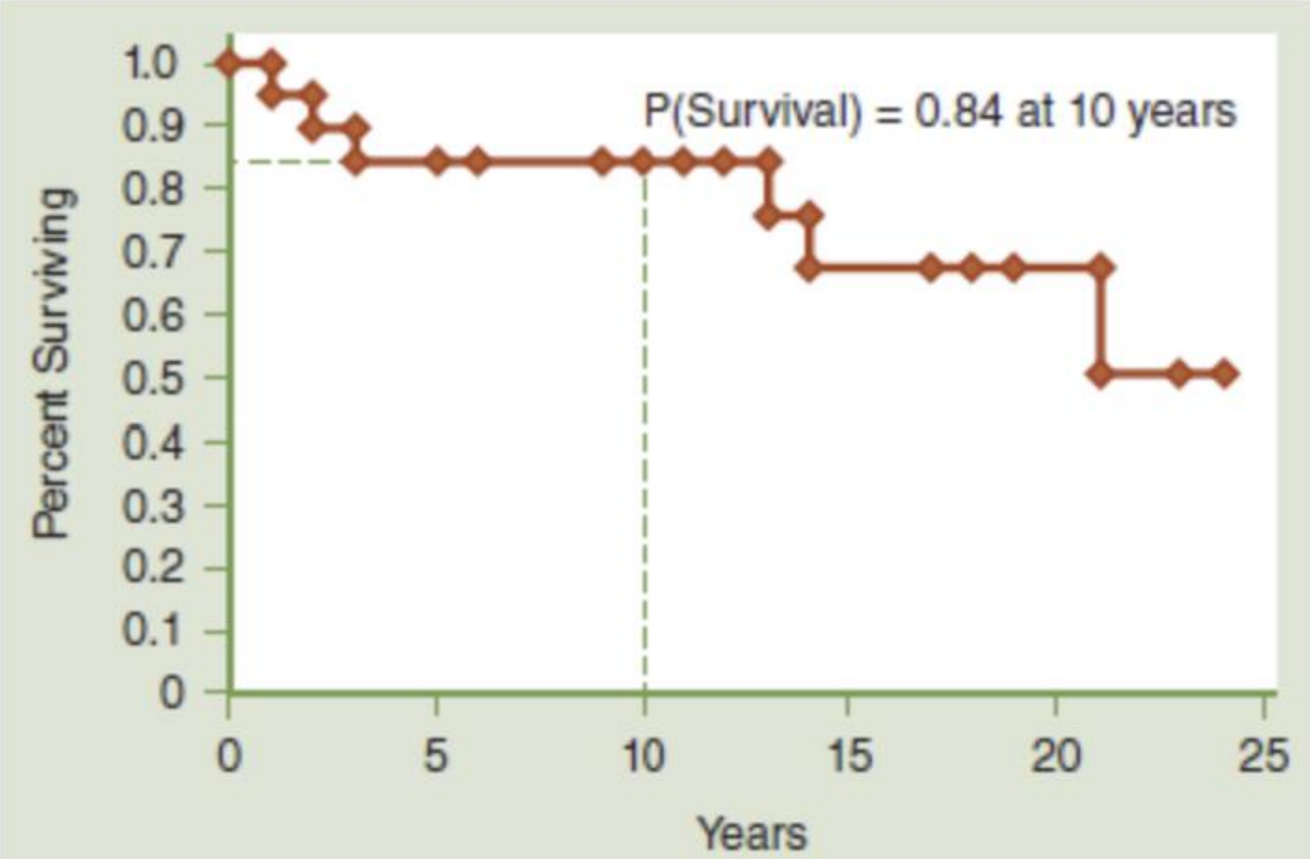
Time Data

$$S_{t+1} = S_t \frac{N_t - D_t}{N_t}$$

11.2 Estimating the Survival Function - Kaplan-Meier

Estimating Survival Curve

Time, years	Survival Probability, $S_{t+1} = S_t \times [(N_{t+1} - D_{t+1})/N_{t+1}]$
0	1†
1	1 × [(20 – 1)/20] = 0.950
2	0.950 × [(19 – 0)/19] = 0.950
3	0.950 × [(18 – 1)/18] = 0.897
5	0.897 × [(17 – 1)/17] = 0.844
6	0.844
9	0.844
10	0.844
11	0.844
12	0.844
13	0.844
14	0.760
17	0.676
18	0.676
19	0.676
21	0.676
23	0.507
24	0.507



Participant	Year of Death	Last Contact
14	1	
8		2
2	3	
18	5	
17		6
19		9
15		10
3		11
13		12
6		13
7	14	
10		17
20	17	
9		18
4		19
12		21
16	23	
1		24
5		24
11		24

Time Data

$$S_{t+1} = S_t \frac{N_t - D_t}{N_t}$$

11.2 Estimating the Survival Function - Kaplan-Meier

Forming Confidence Interval

Time, years	Number at Risk, N_t	Number of Deaths, D_t	Survival Probability, S_t	$\frac{D_t}{N_t(N_t-D_t)}$	$\sum\frac{D_t}{N_t(N_t-D_t)}$	$S_t\sqrt{\sum\frac{D_t}{N_t(N_t-D_t)}}$	$1.96 \times SE(S_t)$
0	20		1				
1	20	1	0.950	0.003	0.003	0.049	0.096
2	19		0.950	0.000	0.003	0.049	0.096
3	18	1	0.897	0.003	0.006	0.069	0.135
5	17	1	0.844	0.004	0.010	0.083	0.162
6	16		0.844	0.000	0.010	0.083	0.162
9	15		0.844	0.000	0.010	0.083	0.162
10	14		0.844	0.000	0.010	0.083	0.162
11	13		0.844	0.000	0.010	0.083	0.162
12	12		0.844	0.000	0.010	0.083	0.162
13	11		0.844	0.000	0.010	0.083	0.162
14	10	1	0.760	0.011	0.021	0.109	0.214
17	9	1	0.676	0.014	0.035	0.126	0.246
18	7		0.676	0.000	0.035	0.126	0.246
19	6		0.676	0.000	0.035	0.126	0.246
21	5		0.676	0.000	0.035	0.126	0.246
23	4	1	0.507	0.083	0.118	0.174	0.341
24	3		0.507	0.000	0.118	0.174	0.341

Participant	Year of Death	Last Contact
14	1	
8		2
2	3	
18	5	
17		6
19		9
15		10
3		11
13		12
6		13
7	14	
10		17
20	17	
9		18
4		19
12		21
16	23	
1		24
5		24
11		24

Time Data

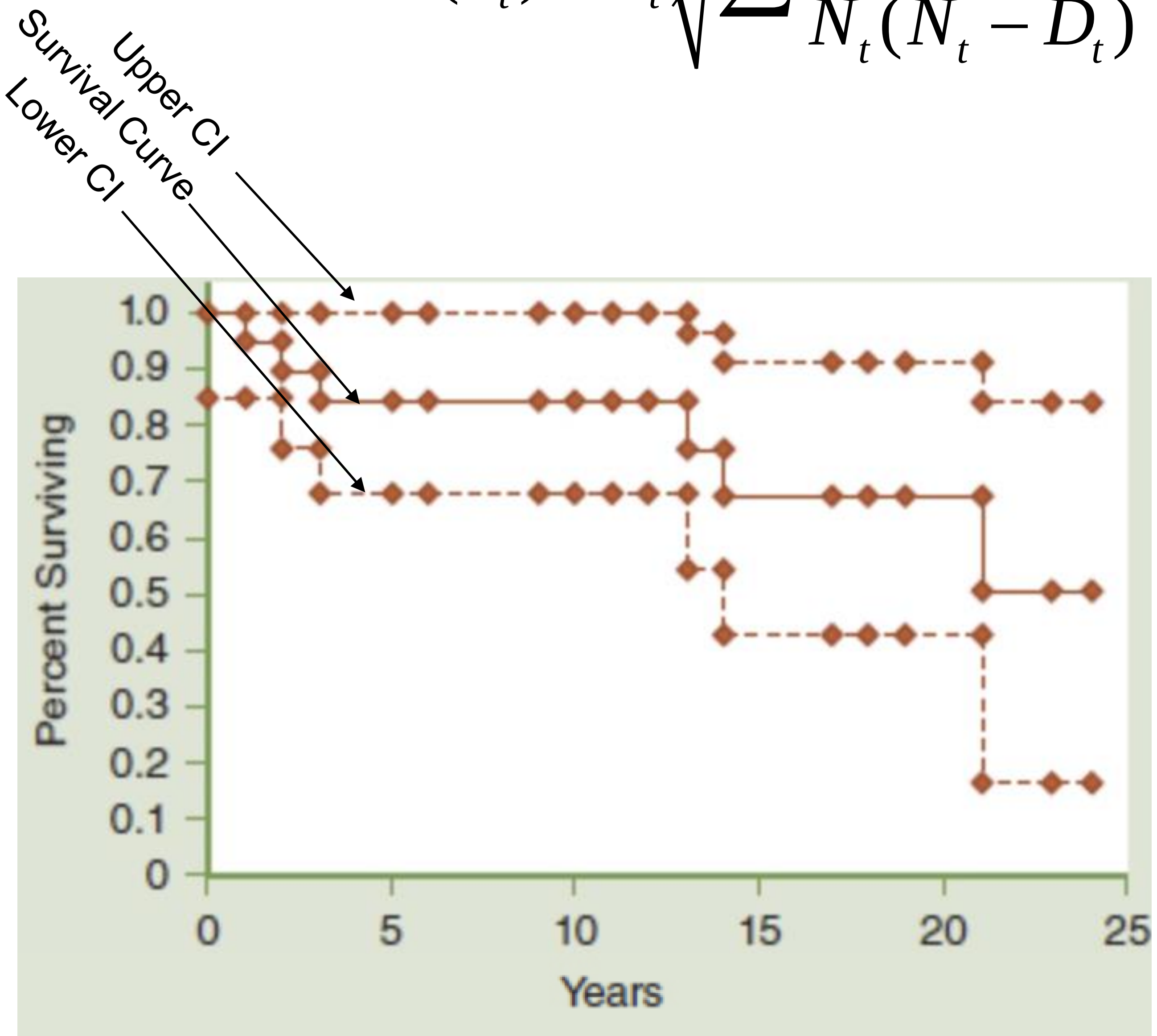
$$S_{t+1} = S_t \frac{N_t - D_t}{N_t}$$

11.2 Estimating the Survival Function - Kaplan-Meier

Forming Confidence Interval

Time, years	Survival Probability, S_t	$1.96 \times SE(S_t)$
0	1	
1	0.950	0.096
2	0.950	0.096
3	0.897	0.135
5	0.844	0.162
6	0.844	0.162
9	0.844	0.162
10	0.844	0.162
11	0.844	0.162
12	0.844	0.162
13	0.844	0.162
14	0.760	0.214
17	0.676	0.246
18	0.676	0.246
19	0.676	0.246
21	0.676	0.246
23	0.507	0.341
24	0.507	0.341

$$SE(S_t) = S_t \sqrt{\sum \frac{D_t}{N_t(N_t - D_t)}}$$



Participant	Year of Death	Last Contact
14	1	
8		2
2	3	
18	5	
17		6
19		9
15		10
3		11
13		12
6		13
7	14	
10		17
20	17	
9		18
4		19
12		21
16	23	
1		24
5		24
11		24

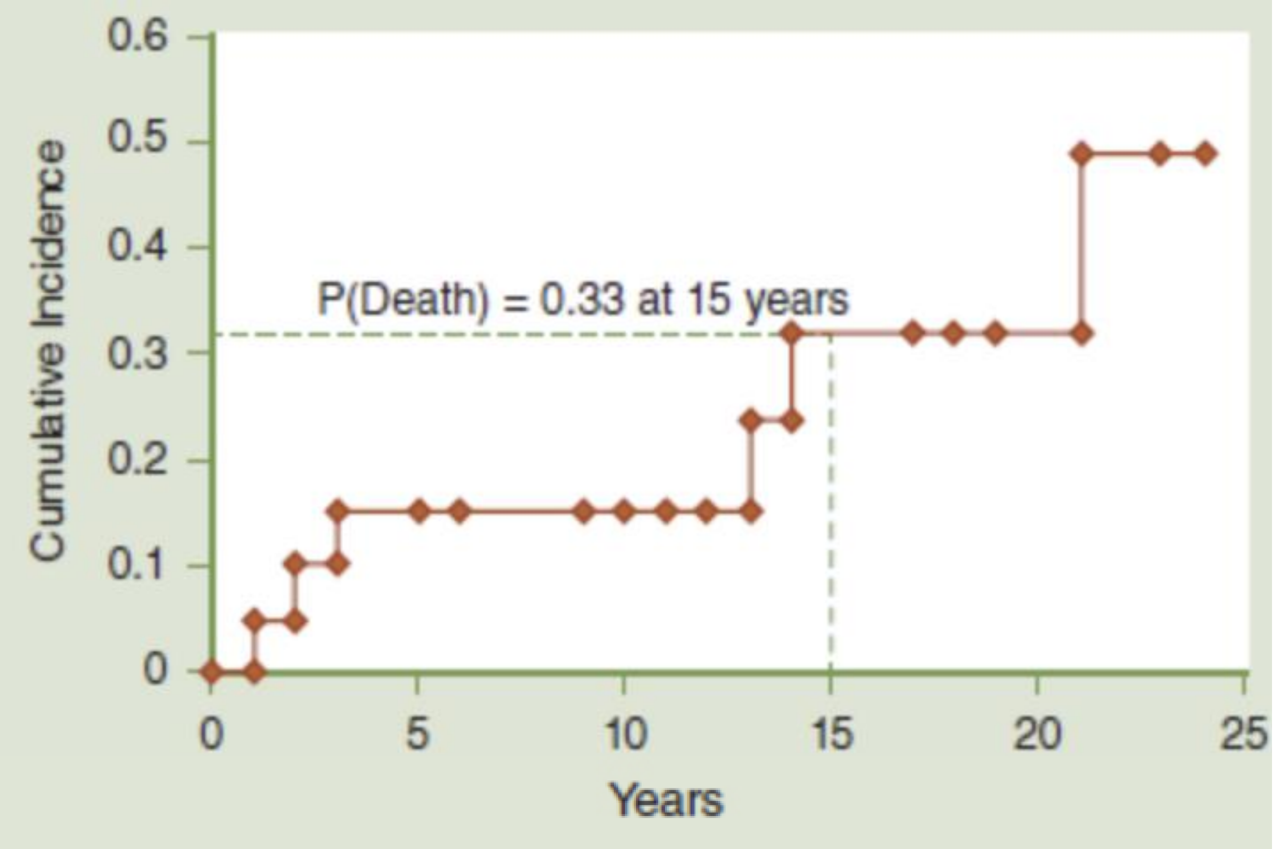
Time Data

$$S_{t+1} = S_t \frac{N_t - D_t}{N_t}$$

11.2 Estimating the Survival Function - Kaplan-Meier

Time, years	Survival Probability, S_t	Failure Probability, $1 - S_t$
0	1	0
1	0.950	0.050
2	0.950	0.050
3	0.897	0.103
5	0.844	0.156
6	0.844	0.156
9	0.844	0.156
10	0.844	0.156
11	0.844	0.156
12	0.844	0.156
13	0.844	0.156
14	0.760	0.240
17	0.676	0.324
18	0.676	0.324
19	0.676	0.324
21	0.676	0.324
23	0.507	0.493
24	0.507	0.493

Some prefer cumulative incidence



Participant	Year of Death	Last Contact
14	1	
8		2
2	3	
18	5	
17		6
19		9
15		10
3		11
13		12
6		13
7	14	
10		17
20	17	
9		18
4		19
12		21
16	23	
1		24
5		24
11		24

Time Data

$$S_{t+1} = S_t \frac{N_t - D_t}{N_t}$$

11.3 Comparing Survival Curves

There are methods for comparing equivalence of survival curves.

An example is one survival curve for a group receiving a medication and another survival curve for another group receiving a placebo.

We might be comparing survival curves for men vs. women or between two demographic groups.

Here present version of log-rank test statistic linked to χ^2 test.
Compares observed events to expected events at each time point.

11.3 Comparing Survival Curves

Example: Small clinical trial to compare chemo Before vs. After surgery.

Chemotherapy Before Surgery		Chemotherapy After Surgery	
Month of Death	Month of Last Contact	Month of Death	Month of Last Contact
8	8	33	48
12	32	28	48
26	20	41	25
14	40		37
21			48
27			25
			43

We can perform a hypothesis test to see if the two treatments result in equivalent outcomes.

11.3 Comparing Survival Curves

Example: We can perform a hypothesis test for equivalence.

Chemo Before Surgery

Time, months	Number at Risk, N_t	Number of Deaths, D_t	Number Censored, C_t	Survival Probability, $S_{t+1} = S_t \times ((N_{t+1} - D_{t+1})/N_{t+1})$
0	10			1.000
8	10	1	1	0.900
12	8	1		0.788
14	7	1		0.675
20	6		1	0.675
21	5	1		0.540
26	4	1		0.405
27	3	1		0.270
32	2		1	0.270
40	1		1	0.270

Chemo After Surgery

Time, months	Number at Risk, N_t	Number of Deaths, D_t	Number Censored, C_t	Survival Probability, $S_{t+1} = S_t \times ((N_{t+1} \times D_{t+1})/N_{t+1})$
0	10			1.000
25	10		2	1.000
28	8	1		0.875
33	7	1		0.750
37	6		1	0.750
41	5	1		0.600
43	4		1	0.600
48	3		3	0.600

Plot the survival curves.

11.3 Comparing Survival Curves

Example: We can perform a hypothesis test for equivalence.

Step 1: Hypotheses and significance. α=0.05

H_0: The two survival curves are identical.

H_1: The two survival curves are not identical.

$\sum_{t=1}^{T} O_{ij}$ =Observed Deaths in Group *j*

$\sum_{t=1}^{T} E_{ij}$ =Expected Deaths in Group *j*

Step 2: Test Statistic (log-rank test)

$$\chi^2 = \sum_{j=1}^{2} \left(\sum_{t=1}^{T} O_{ij} - \sum_{t=1}^{T} E_{ij} \right)^2 \Big/ \sum_{t=1}^{T} E_{ij}$$

$$df = k - 1$$

Step 3: Decision Rule

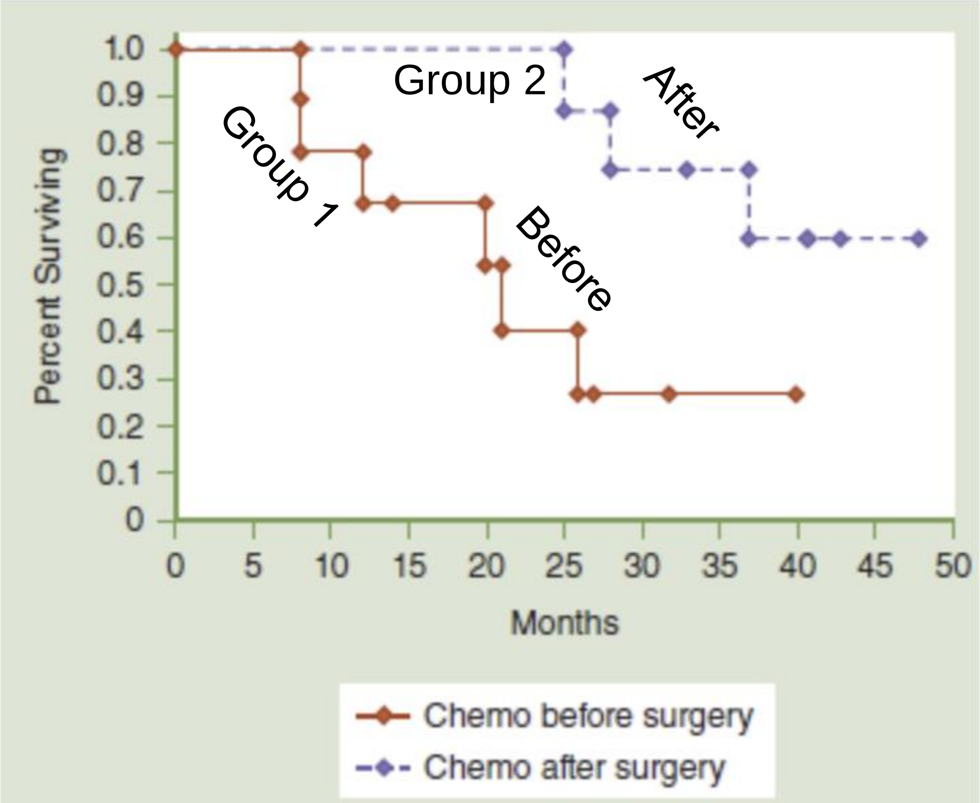
Reject if $\chi^2 > \chi^2_{\alpha,df}$=3.84.

χ^2 Table

df	.10	.05
1	2.71	3.84

Step 4: Compute Test Statistic

Next slide.



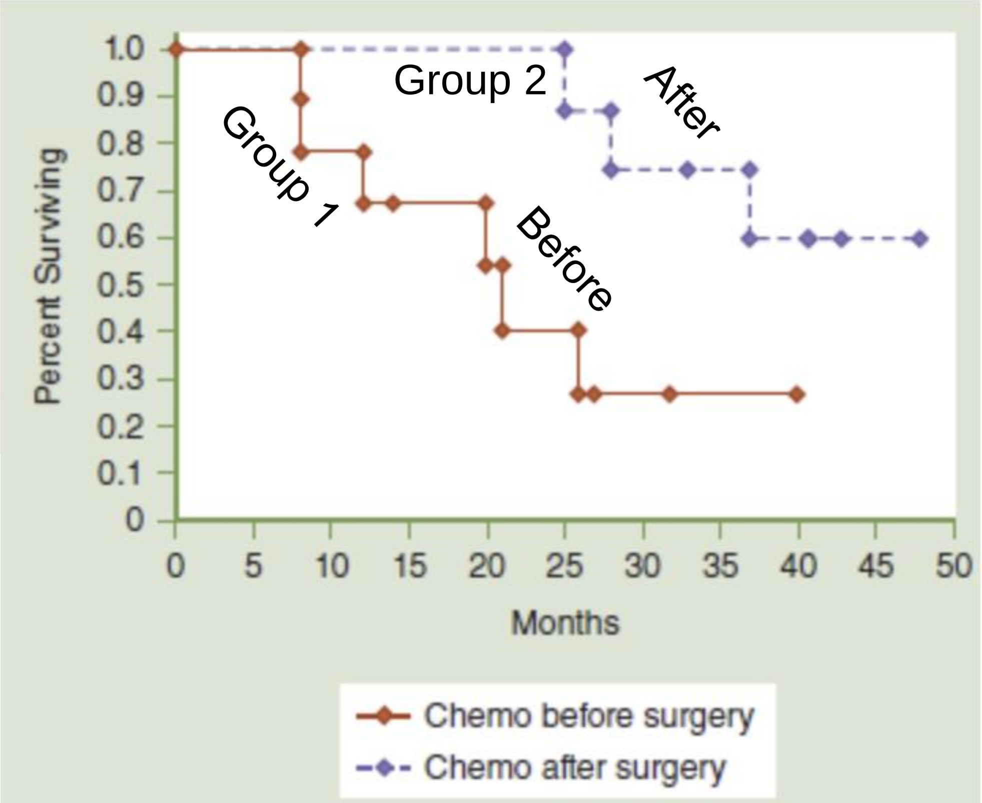
11.3 Comparing Survival Curves

Step 4: Compute Test Statistic

Death Time, months	Number at Risk in Group 1, N_{1t}	Number at Risk in Group 2, N_{2t}	Total Number at Risk, N_t	Number of Events in Group 1, O_{1t}	Number of Events in Group 2, O_{2t}	Total Number of Events, O_t	Expected Number of Events in Group 1, $E_{1t} = N_{1t} \times (O_t/N_t)$	Expected Number of Events in Group 2, $E_{2t} = N_{2t} \times (O_t/N_t)$
8	10	10	20	1	0	1	0.500	0.500
12	8	10	18	1	0	1	0.444	0.556
14	7	10	17	1	0	1	0.412	0.588
21	5	10	15	1	0	1	0.333	0.667
26	4	8	12	1	0	1	0.333	0.667
27	3	8	11	1	0	1	0.273	0.727
28	2	8	10	0	1	1	0.200	0.800
33	1	7	8	0	1	1	0.125	0.875
41	0	5	5	0	1	1	0.000	1.000
				6	3		2.620	6.380

$\sum_{t=1}^{T} O_{ij}$ =Observed Deaths in Group j

$\sum_{t=1}^{T} E_{ij}$ =Expected Deaths in Group j



$$\chi^2 = \sum_{j=1}^{2} \frac{\left(\sum_{t=1}^{T} O_{ij} - \sum_{t=1}^{T} E_{ij}\right)^2}{\sum_{t=1}^{T} E_{ij}} = \frac{(6-2.620)^2}{2.620} + \frac{(3-6.380)^2}{6.380}$$

$$\chi^2 = 4.360 + 1.791 = 6.151$$

11.3 Comparing Survival Curves

Example: We can perform a hypothesis test for equivalence.

Step 1: Hypotheses and significance. $\alpha=0.05$

H_0: The two survival curves are identical.

H_1: The two survival curves are not identical.

$\sum_{t=1}^{T} O_{ij}$ =Observed Deaths in Group *j*

$\sum_{t=1}^{T} E_{ij}$ =Expected Deaths in Group *j*

Step 2: Test Statistic (log-rank test)

$$\chi^2 = \sum_{j=1}^{2}\left(\sum_{t=1}^{T} O_{ij} - \sum_{t=1}^{T} E_{ij}\right)^2 \Big/ \sum_{t=1}^{T} E_{ij}$$

$$df = k - 1$$

Step 3: Decision Rule

Reject if $\chi^2 > \chi^2_{\alpha,df}=3.84$.

χ^2 Table

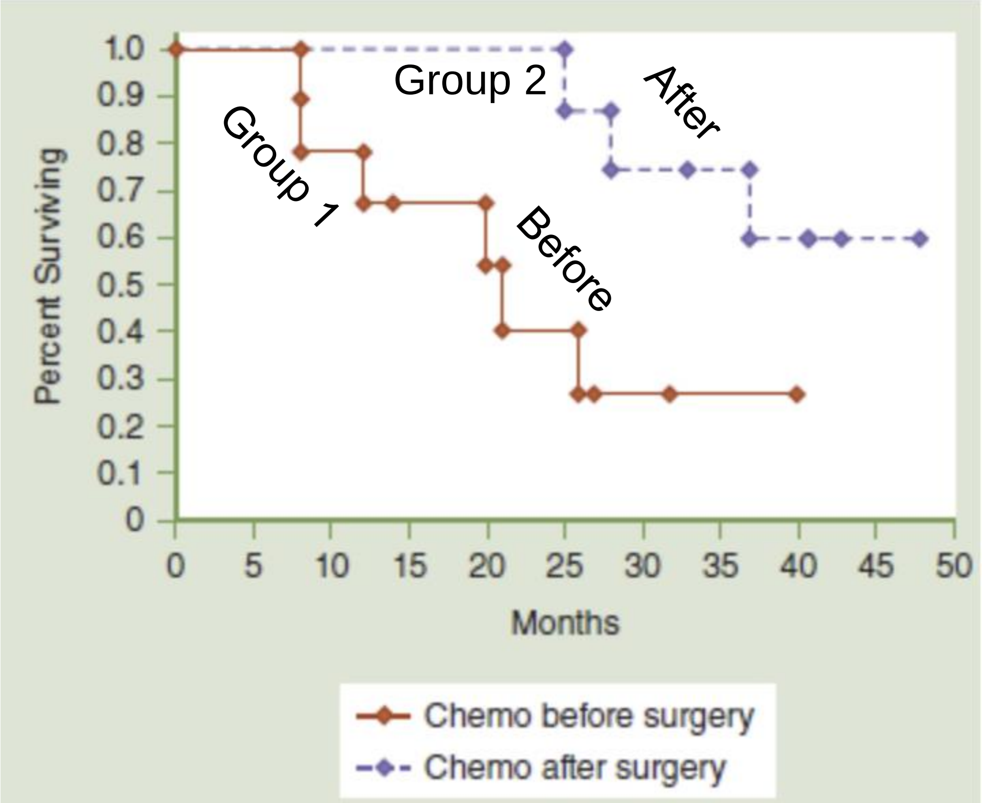
df	.10	.05
1	2.71	3.84

Step 4: Compute Test Statistic

$\chi^2=6.15$

Step 5: Conclusion

Reject H_0 because 6.16>3.84.



11.6 Summary

The **survival function** is the probability a person survives past a time t.

Actuarial Life Table

N_t = # event free during interval t
(Number at risk)

D_t = # who die in interval t

C_t = # censored in interval t

N_{t*}= avg. # at risk in interval t, $N_{t*}=N_t-C_t/2$

q_t = prop. die in interval t, $q_t=D_t/N_{t*}$

p_t = prop. survive in interval t, $p_t=1-q_t$

S_t = prop. survive past interval t

Can plot S_t vs. t.

Kaplan-Meier Life Table

$$S_{t+1} = S_t \frac{N_t - D_t}{N_t}$$

$$SE(S_t) = S_t \sqrt{\sum \frac{D_t}{N_t(N_t - D_t)}}$$

Chi-Square Test

$$\chi^2 = \sum_{j=1}^{2} \frac{\left(\sum_{t=1}^{T} O_{ij} - \sum_{t=1}^{T} E_{ij}\right)^2}{\sum_{t=1}^{T} E_{ij}} \quad df = k-1$$

Cox Proportional Hazards Model

$$h(t) = h_0(t)\exp(b_1 x_1 + b_2 x_2 + \ldots + b_p x_p)$$

Questions?

Homework 11

Read Chapter 11.

Problems 12, 14. (Both interpreting graphs.)