

의학연구자료의 생존분석

김호

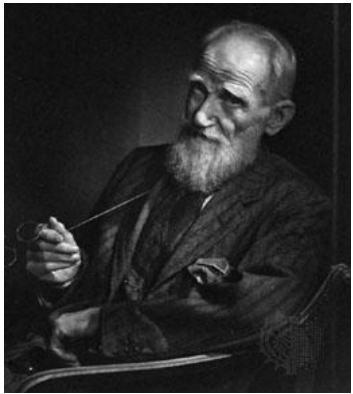
서울대학교 보건대학원

들어가며



“세상에는 세가지 거짓말이 있다. 거짓말, 새빨간 거짓말, 그리고 통계.(There are three kinds of lies: lies, damned lies, and statistics)”

– 벤자민 디즈레일리(Benjamin Disraeli)



“진정으로 교육받은 사람의 징표는 그가 통계에 의해 깊은 감동은 받았다는 것이다. (It is the mark of a truly intelligent person to be moved by statistics.)”

✦ 조지 버나드 쇼 (George Bernard Shaw)

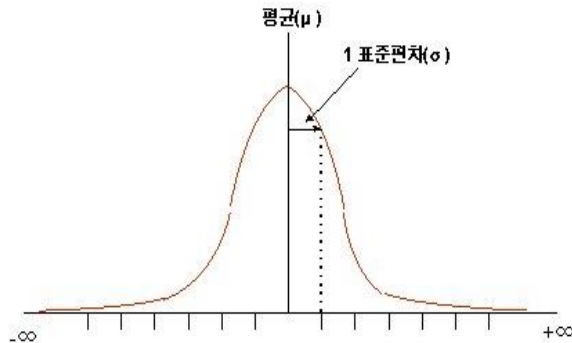
- “너의 자료를 고문하라. 그러면 어떠한 결과라도 얻을 수 있을 것이다.(Torture numbers, and they'll confess to anything): ~Gregg Easterbrook
- “만약 당신의 한쪽 발이 얼음 속에 다른 발이 끓는 솥 속에 있다고 하면, 통계적으로는 당신은 아주 평안한 상태이다. (Say you were standing with one foot in the oven and one foot in an ice bucket. According to the percentage people, you should be perfectly comfortable.)”: Bobby Bragan, 1963
-

- Then there is the man who drowned crossing a stream with an average depth of six inches. ~W.I.E. Gates
- The average human has one breast and one testicle. ~Des McHale
- The death of one man is a tragedy. The death of millions is a statistic. ~Joe Stalin, comment to Churchill at Potsdam, 1945
- I abhor averages. I like the individual case. A man may have six meals one day and none the next, making an average of three meals per day, but that is not a good way to live. ~Louis D. Brandeis

모집단과 표본

- 모집단
- 모수

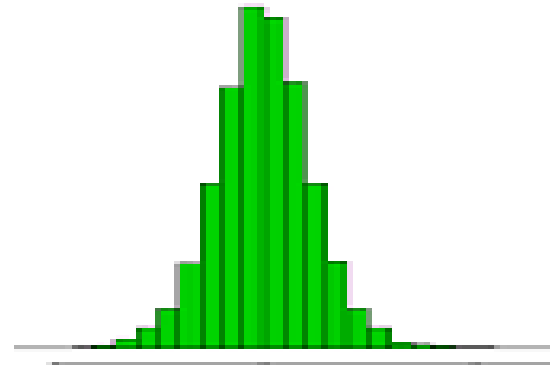
$$N(\mu, \sigma^2)$$



- 표본 $Y_1, \dots, Y_n$
- 추정치

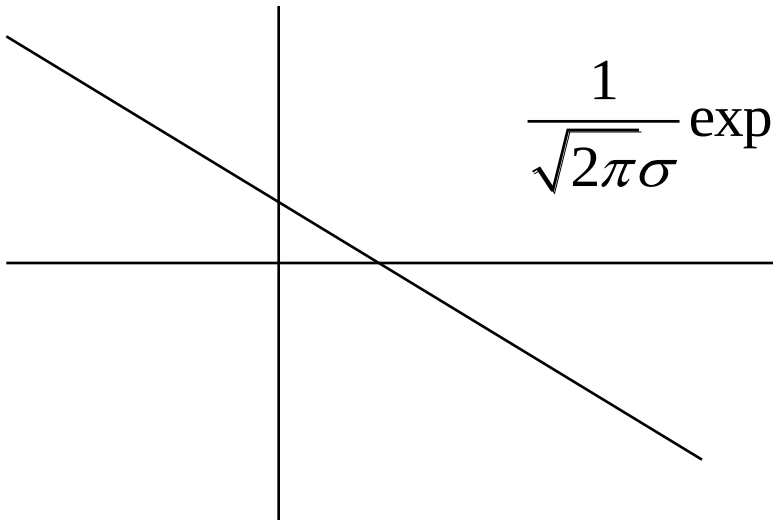
$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$



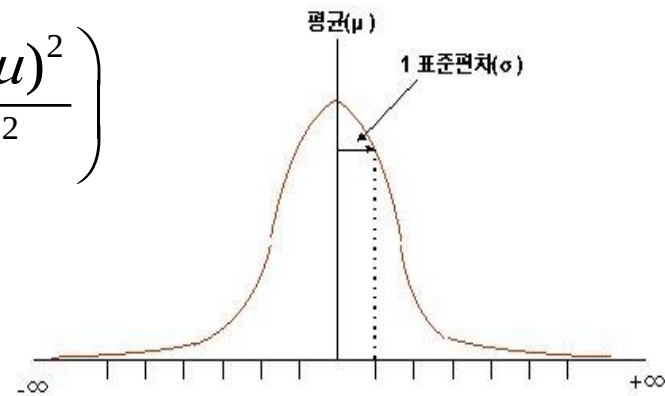
- 모수: 가정한 모형의 통계적 성질을 완전히 결정하는 상수(들)

$$Y = a + b x$$



$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$N(\mu, \sigma^2)$$



$$Y_1, Y_2, \dots, Y_n$$

$$\bar{Y}_n, S_n^2$$

- 관심모수 : 연구의 가설을 수학적인 모수로 표시해야 함

- 두 집단에서 평균비교 $d = \mu_1 - \mu_2$

- 두 집단의 비율비교 $r = \frac{p_1}{p_2}$

$$OR = \frac{p_1 / (1 - p_1)}{p_2 / (1 - p_2)}$$

생존분석이란?

- Time to an event
 - 예: 사망, 질병 발생 혹은 재발, 기업도산, 재범시간, 등 다양함
- 2 가지 특성:
 - 시간은 대부분 정규분포가 아님
 - 중도절단 (censoring)을 고려해야함
 - 연구의 종료
 - 추적의 실패
 - Withdraw from the study (drop out)
 - Death from unrelated causes

생존분석 자료를

- 회귀분석을 이용해서 분석
 - 시간, 혹은 $\log(\text{시간})$ 을 종속변수로
 - 중도절단(사건이 그 시점 이후에 일어남)에 대한 고려가 없음
- 로지스틱 회귀분석을 이용해서 분석
 - 특정시점(연구조종료일)에서의 생존여부(1,0)를 종속변수로
 - 생존기간이 다른 자료들이 동일한 방법으로 취급됨
- 생존분석에서는 (시간, 절단) 두 변수가 종속변수로 필요함

Make-up data

Test: 1, 1, 1, 1+, 4+, 5+

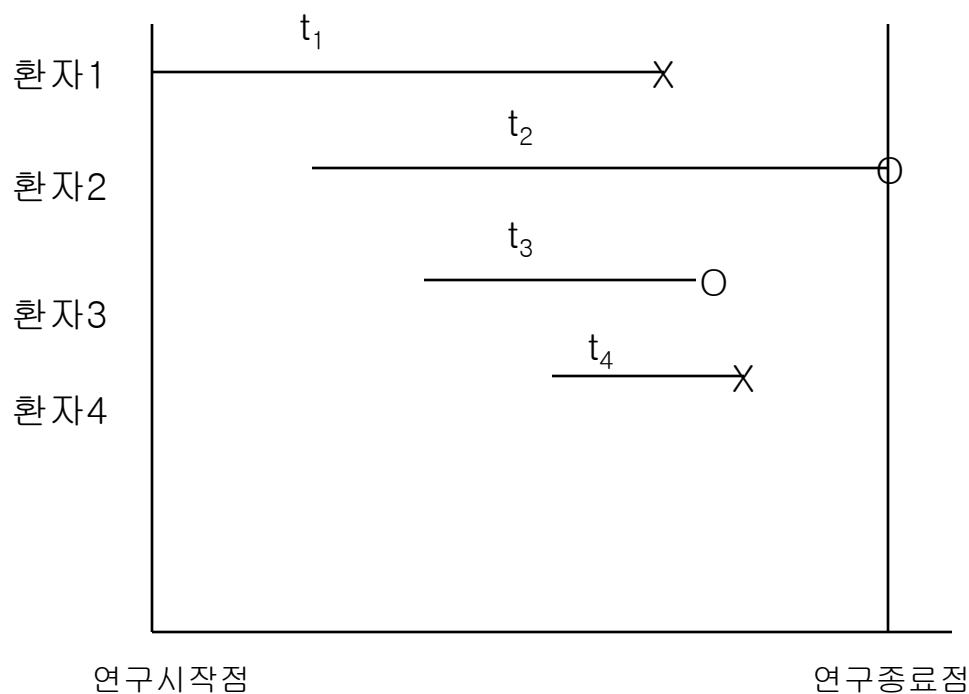
Placebo: 1+, 2+, 3, 3+

(1) Ignore +

$$\frac{1+1+1+1+4+5}{6} = \frac{13}{6}, \frac{1+2+3+3}{4} = \frac{9}{4}$$

(2) Delete +'s : ???

생존분석 자료



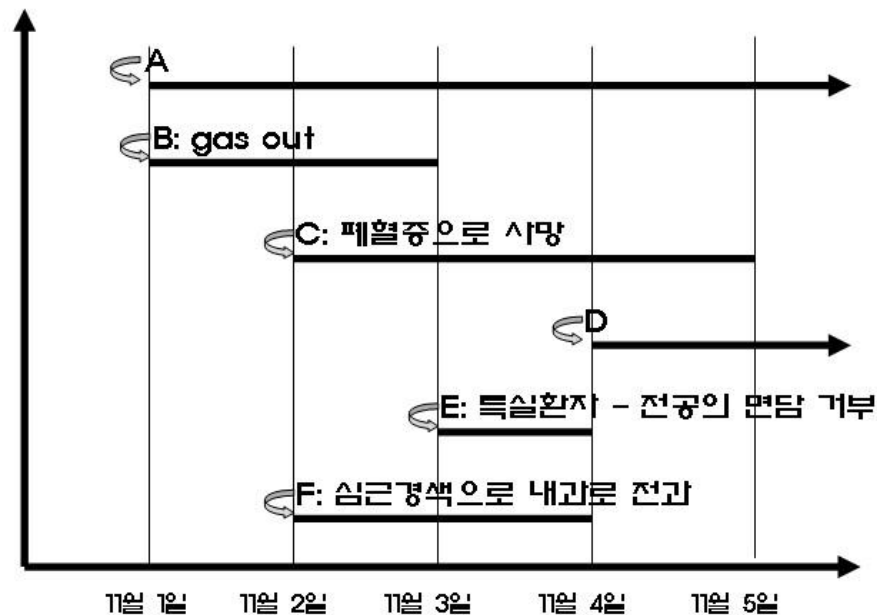
X: 사망, O:중도절단

Pt time delta sex age ...

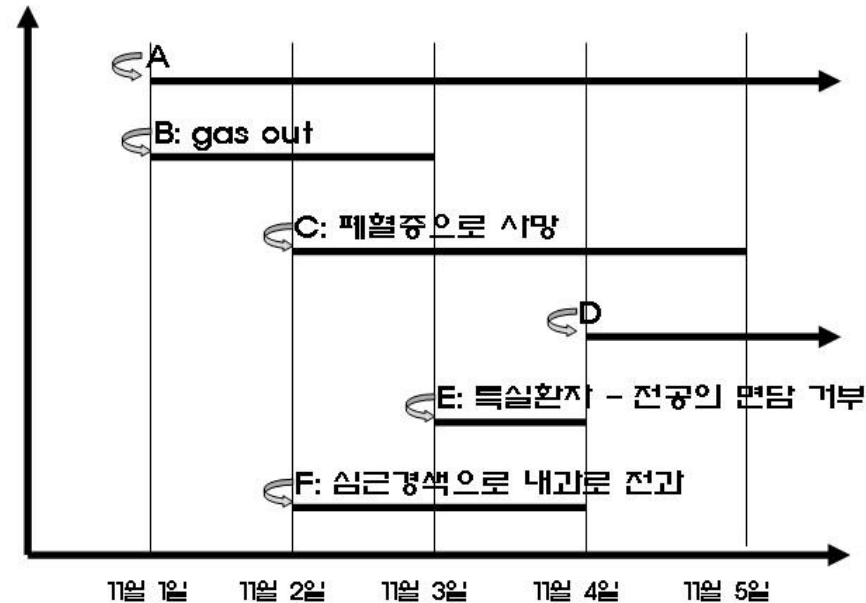
1	10	1	1	65
2	12	0	2	70
3	7	0	1	67
4	5	1	2	45
5				
6				

문제풀이

외과의사 봉달취군은 11월 1일부터 11월 5일까지 막창자꼬리 절제술을 받은 환자들에서 매일 아침 gas out이 있었는지 문진으로 확인하여 gas out을 장폐색의 종료로 보고 수술 후 기능적 장폐색이 지속되는 기간을 기록하였다. 아래 그림은 그 결과의 일부를 도해한 것이다. 생존분석을 이용하여 본 자료를 분석하고자 할 때 각각의 자료가 complete data인지, censored data인지 구별하고 그 이유를 쓰시오. 또한 각각의 장폐색 기간(duration of ileus)을 계산하십시오.



문제 풀이



답)

A: censored data, Termination of the study, 4일

B: complete data (uncensored data), End with Event, 2일

C: censored data, Loss to follow up, 3일

D: censored data, Termination of the study, 1일

E: censored data, Drop-out, 1일

F: censored data, Loss to follow up, 2일

생존 함수

T 생존시간을 나타내는 확률변수

t 특정 시간

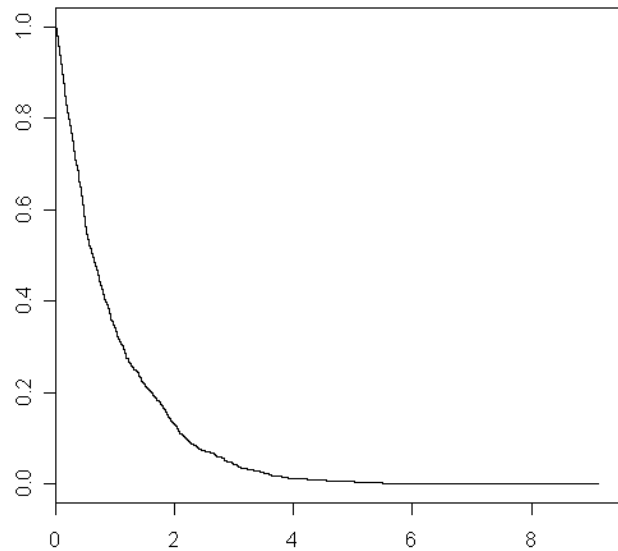
$S(t) = P(T > t)$ 생존함수

$h(t) = \lim_{dt \rightarrow 0} P(t < T \leq t + dt | T > t) / dt$ 위험함수

$= f(t) / S(t)$

생존함수 : 환자가 t 시간 이상 생존할 확률

위험함수: t 시점까지는 생존했다고 가정하고 바로 직후 사망할 확률

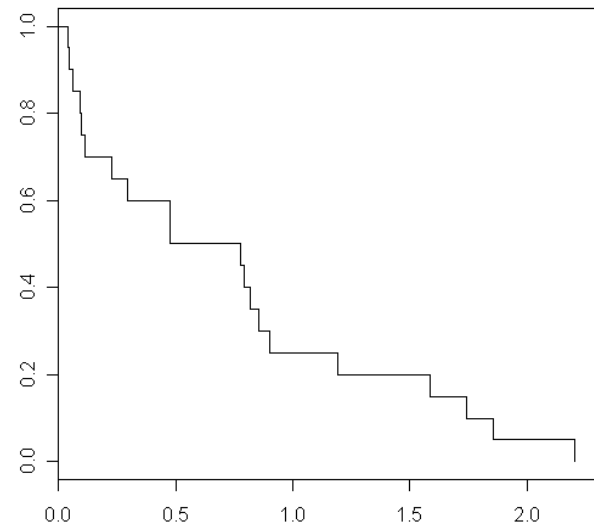


생존함수의 이론적 형태

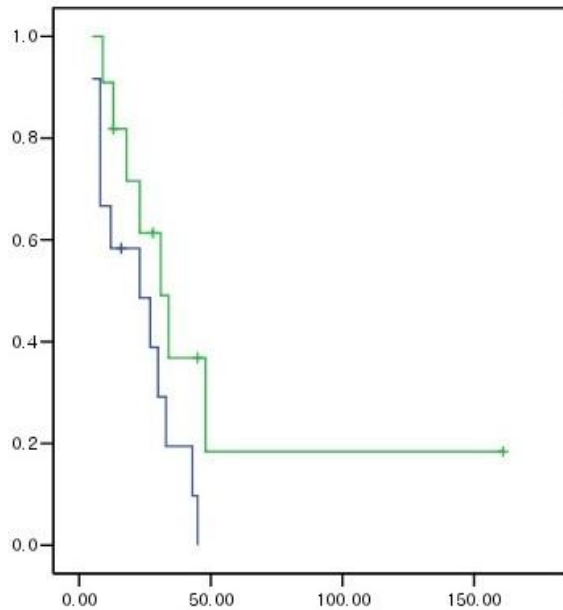
$$S(0) = 1, S(\infty) = 0$$

실제 자료에서의 생존함수의 추정형태

마지막 시점에서 0이 아닐 수도 있음 : 중도절단의 효과



생존함수



생존함수들이 이와 같은 경우,
한 집단은 연구종료까지 피험자
의 20%에서는 사건이 발생하지
않았음을 의미

중도절단율이 너무 높은 경우에
대비해야함

생명표를 이용한 생존함수 추정

	사망자수	중도절단수
(0 - 1]	85	0
(1 - 2]	32	8
(2 - 3]	31	17
(3 - 4]	30	17
(4 - 5]	26	16
(5 - 6]	28	28
(6 - 7]	26	31
(7 - 8]	22	32
(8 - 9]	17	25
(9 - 10]	12	26
(10- 11]	9	18
(11- 12]	5	10
(12- 13]	5	7
(13-	3	11

(0-1] 위험그룹인원수 : 총사망자수+총중도절단수 =331+246

유효인원수 = $577 - 0/2 = 577$

사망률=사망자수/유효인원수= $85/577=0.1473$

생존율=1-사망률=0.8527

생존함수=1

사망확률=1- 생존함수

지금의 중도절단수

이전의 사망자수

(1-2] 위험그룹인원수 : $577 - 85 - 0 = 492$

유효인원수 = $492 - 8/2 = 488$

사망률=사망자수/유효인원수= $32/488=0.0656$

생존율=1-사망률=0.9344

생존함수= $1 * 0.8527 = 0.8527$

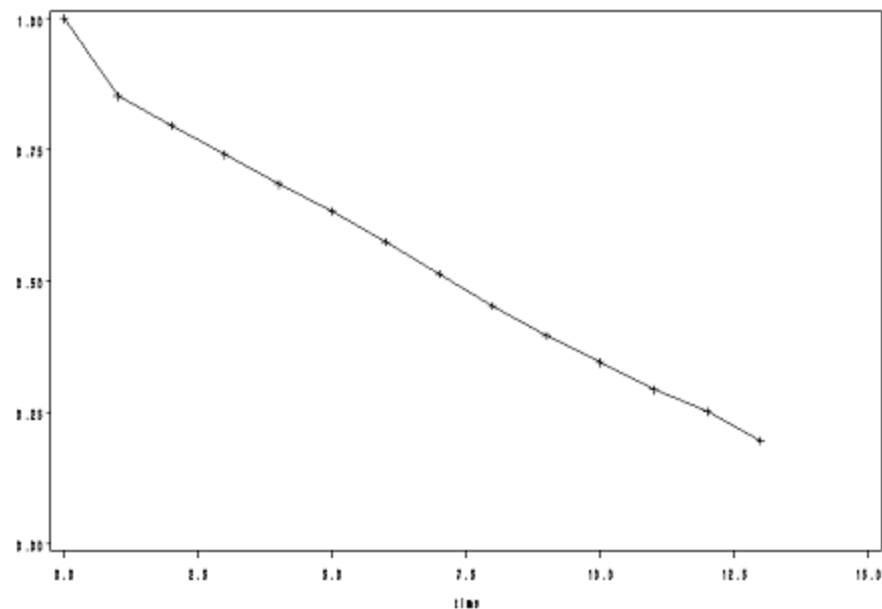
사망확률=1- 생존함수= $1 - 0.8527 = 0.1473$

이전의 중도절단수

생존함수는 누적적으로 곱해준다.

생명표를 이용한 생존함수추정

	사망자수	중도절단수
(0 - 1]	85	0
(1 - 2]	32	8
(2 - 3]	31	17
(3 - 4]	30	17
(4 - 5]	26	16
(5 - 6]	28	28
(6 - 7]	26	31
(7 - 8]	22	32
(8 - 9]	17	25
(9 - 10]	12	26
(10- 11]	9	18
(11- 12]	5	10
(12- 13]	5	7
(13-	3	11



SPSS를 이용한 생명표 작성

예. PBC 자료

- Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984.
- 주요변수 : 생존시간, trt(D-penicillamine and placebo), age, sex 외에 임상, 생화학 지표

	n	time	delta	trt	age	sex	ascites	hepa	spider	edema	billi	chol
1	1.00	400.00	1.00	1.00	58.77	1.00	1.00	1.00	1.00	1.00	14.50	261.00
2	2.00	4500.00	.00	1.00	56.45	1.00	.00	1.00	1.00	.00	1.10	302.00
3	3.00	1012.00	1.00	1.00	70.07	.00	.00	.00	.00	.50	1.40	176.00
4	4.00	1925.00	1.00	1.00	54.74	1.00	.00	1.00	1.00	.50	1.80	244.00
5	5.00	1504.00	.00	.00	38.11	1.00	.00	1.00	1.00	.00	3.40	279.00
6	6.00	2503.00	1.00	.00	66.26	1.00	.00	1.00	.00	.00	.80	248.00
7	7.00	1832.00	.00	.00	55.53	1.00	.00	1.00	.00	.00	1.00	322.00
8	8.00	2466.00	1.00	.00	53.06	1.00	.00	.00	.00	.00	.30	280.00
9	9.00	2400.00	1.00	1.00	42.51	1.00	.00	.00	1.00	.00	3.20	562.00
10	10.00	51.00	1.00	.00	70.56	1.00	1.00	.00	1.00	1.00	12.60	200.00
11	11.00	3762.00	1.00	.00	53.71	1.00	.00	1.00	1.00	.00	1.40	259.00
12	12.00	304.00	1.00	.00	59.14	1.00	.00	.00	1.00	.00	3.60	236.00
13	13.00	3577.00	.00	.00	45.69	1.00	.00	.00	.00	.00	.70	281.00
14	14.00	1217.00	1.00	.00	56.22	.00	1.00	1.00	.00	1.00	.80	.
15	15.00	3584.00	1.00	1.00	64.65	1.00	.00	.00	.00	.00	.80	231.00
16	16.00	3672.00	.00	.00	40.44	1.00	.00	.00	.00	.00	.70	204.00
17	17.00	769.00	1.00	.00	52.18	1.00	.00	1.00	.00	.00	2.70	274.00
18	18.00	131.00	1.00	1.00	53.93	1.00	.00	1.00	1.00	1.00	11.40	178.00
19	19.00	4232.00	.00	1.00	49.56	1.00	.00	1.00	.00	.50	.70	235.00
20	20.00	1356.00	1.00	.00	59.95	1.00	.00	1.00	.00	.00	5.10	374.00
21	21.00	3445.00	.00	.00	64.19	.00	.00	1.00	1.00	.00	.60	252.00

자료 : A.xls

분석 > 생존분석 > 생명표 > 시간변수, 시간 간격, 상태변수

A.sav - SPSS 데이터 편집기

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

1 : n 1

	n	time	delta	sex	ascites	hepa	spider	edema	billi
1	1.00	400.00	1.0	58.77	1.00	1.00	1.00	1.00	14.50
2	2.00	4500.00	.0	56.45	1.00	.00	1.00	.00	1.10
3	3.00	1012.00	1.0	70.07	.00	.00	.00	.50	1.40
4	4.00	1925.00	1.0	54.74	1.00	.00	1.00	.50	1.80
5	5.00	1504.00	.0	68.11	1.00	.00	1.00	.00	3.40
6	6.00	2503.00	1.0	56.26	1.00	.00	1.00	.00	.80
7	7.00	1832.00	.0	55.53	1.00	.00	1.00	.00	1.00
8	8.00	2466.00	1.0	53.06	1.00	.00	.00	.00	.30
9	9.00	2400.00	1.0	42.51	1.00	.00	.00	1.00	3.20
10	10.00	51.00	1.0	70.56	1.00	1.00	.00	1.00	12.60
11	11.00	3762.00	1.0	64.65	1.00	.00	1.00	.00	1.40
12	12.00	304.00	1.0	40.44	1.00	.00	.00	.00	3.60
13	13.00	3577.00	.0	52.18	1.00	.00	1.00	.00	.70
14	14.00	1217.00	1.0	53.93	1.00	.00	1.00	1.00	11.40
15	15.00	3584.00	1.00	49.56	1.00	.00	1.00	.50	.70
16	16.00	3672.00	.00	59.95	1.00	.00	1.00	.00	5.10
17	17.00	769.00	1.00	64.19	.00	.00	1.00	.00	.60
18	18.00	131.00	1.00	56.28	1.00	.00	.00	1.00	3.40
19	19.00	4232.00	.00	55.97	1.00	1.00	1.00	1.00	17.40
20	20.00	1356.00	1.00						
21	21.00	3445.00	.00						
22	22.00	673.00	1.00						
23	23.00	264.00	1.00						

데이터 보기 / 변수 보기 /

생명표 SPSS 프로세서 준비 완료

1 1.00 2 2.00 3 3.00 4 4.00 5 5.00 6 6.00 7 7.00 8 8.00 9 9.00 10 10.00 11 11.00 12 12.00 13 13.00 14 14.00 15 15.00 16 16.00 17 17.00 18 18.00 19 19.00 20 20.00 21 21.00 22 22.00 23 23.00

This subfile contains: 312 observations

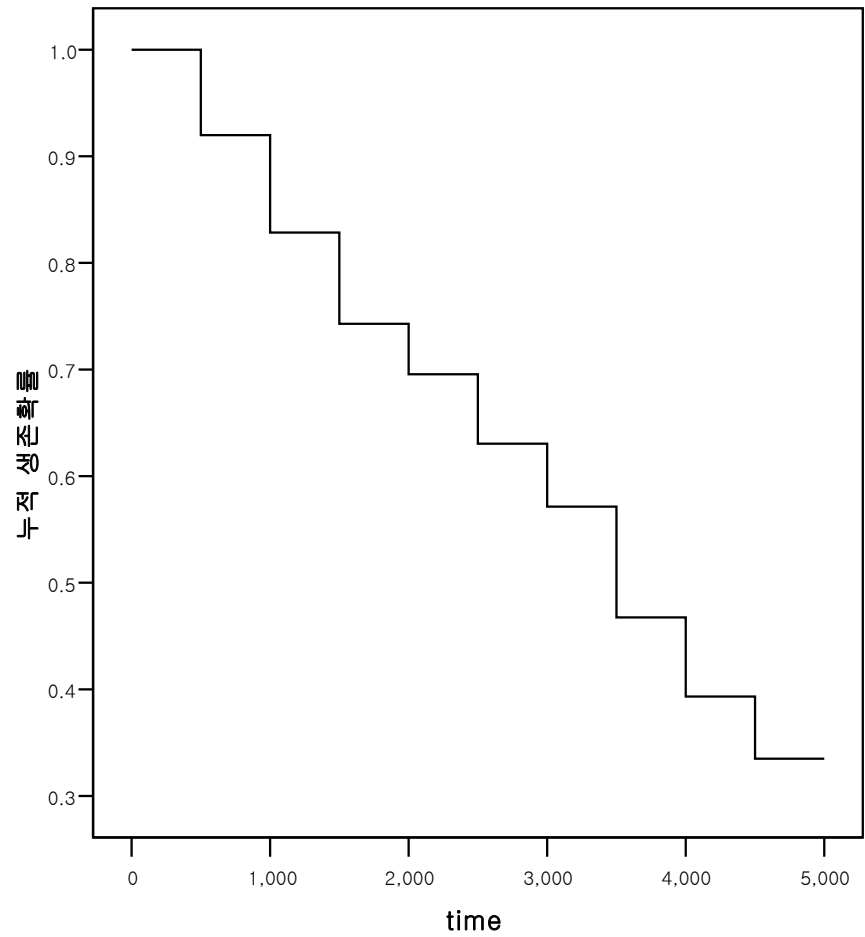
Life Table

Survival Variable		time	time						
		Number	Number	Number	Number	Cumul			
Intrvl	Entrng	Wdrawn	Exposd	of	Propn	Propn	Propn	Proba-	
Start	this	During	to	Termnl	Termi-	Sur-	Surv	bility	Hazard
Time	Intrvl	Intrvl	Risk	Events	nating	viving	at End	Densty	Rate
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
.0	312.0	.0	312.0	25.0	.0801	.9199	.9199	.0002	.0002
500.0	287.0	10.0	282.0	28.0	.0993	.9007	.8285	.0002	.0002
1000.0	249.0	34.0	232.0	24.0	.1034	.8966	.7428	.0002	.0002
1500.0	191.0	36.0	173.0	11.0	.0636	.9364	.6956	.0001	.0001
2000.0	144.0	32.0	128.0	12.0	.0938	.9063	.6304	.0001	.0002
2500.0	100.0	29.0	85.5	8.0	.0936	.9064	.5714	.0001	.0002
3000.0	63.0	16.0	55.0	10.0	.1818	.8182	.4675	.0002	.0004
3500.0	37.0	11.0	31.5	5.0	.1587	.8413	.3933	.0001	.0003
4000.0	21.0	15.0	13.5	2.0	.1481	.8519	.3350	.0001	.0003
4500.0	4.0	4.0	2.0	.0	.0000	1.0000	.3350	.0000	.0000

The median survival time for these data is 3343.6

생존함수

Intrvl Start Time	SE of Cumul Surviving	SE of Proba- bility Densy	SE of Hazard Rate
-----	-----	-----	-----
.0	.0154	.0000	.0000
500.0	.0214	.0000	.0000
1000.0	.0254	.0000	.0000
1500.0	.0275	.0000	.0000
2000.0	.0307	.0000	.0001
2500.0	.0342	.0000	.0001
3000.0	.0408	.0001	.0001
3500.0	.0459	.0001	.0002
4000.0	.0545	.0001	.0002
4500.0	.0545	.0000	.0000



문제풀이

2. 다음 표는 171명의 자궁암환자에 대한 생존자료이다. 이 자료의 생명표를 완성하고, 생존함수그래프를 그리시오.

표. 자궁암 환자의 생존자료

구간	사망자수	중도절단수
0-1	32	31
1-2	31	22
2-3	10	14
3-4	5	8
4-5	4	3
5-6	4	2
6-7	1	0
7-8	0	0
8-9	0	2
9-10	0	1

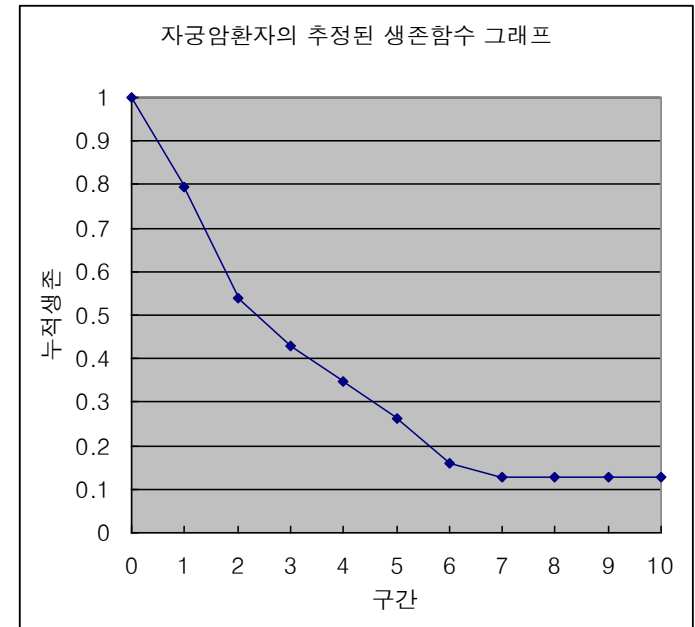
자궁암 환자의 생존표

구간	생존환자수	사망자수	중도절단수	유효인원수	구간사망률	구간생존률	생존합수
0-1	171	32	31	155.5	0.205788	0.794212	0.794212
1-2	108	31	22	97	0.319588	0.680412	0.540392
2-3	55	10	14	48	0.208333	0.791667	0.42781
3-4	31	5	8	27	0.185185	0.814815	0.348586
4-5	18	4	3	16.5	0.242424	0.757576	0.26408
5-6	11	4	2	10	0.4	0.6	0.158448
6-7	5	1	0	5	0.2	0.8	0.126759
7-8	4	0	0	4	0	1	0.126759
8-9	4	0	2	3	0	1	0.126759
9-10	2	0	1	1.5	0	1	0.126759

유효인원수: 구간에서의 위험그룹 인원수에서 중도 절단된 인원수의 1/2을 뺀 수

생존합수는 일정 기간까지의 구간 생존 확률의 누적치이다. 즉 그 구간까지의 각 구간 생존 확률의 곱으로 구할 수 있다.

처음 두 구간의 생존확률을 계산해 보자. 첫번째 구간의 위험 그룹 인원수와 유효인원수는 각각 171, 155.5명이므로 이때의 생존율은 $1 - 32/155.5 = 0.7942$ 이다. 두번째 구간의 위험그룹 인원수는 $171 - 32 - 31 = 108$ 이고, 유효인원수는 $108 - 22/2 = 97$ 이 된다. 이 구간에서의 사망률은 $31/97 = 0.3195$ 이므로 생존율은 $1 - 0.3195 = 0.6804$ 이다. 두번째 구간까지 생존할 확률은 첫번째 구간의 생존율이 0.7942이므로 $0.7942 \times 0.6804 = 0.5404$ 이다.



Kaplan-Meier 곡선

(누적한계추정법)

- 대표적인 생존함수 추정법
- 신장이식환자 호전기간 자료

3.0	4.0+	4.5	4.5	5.5	6.0	6.4	6.5
7.0	7.5	8.4+	10.0	10.0+	12.0	15.0	

사망수/위험집단수

$S(3.0) = 1 - 1/15 = 0.933$, $S(4.0) = 0.933$ 으로 불변
 $S(4.5) = 0.933 * (1 - 2/13) = 0.846$

문제풀이

- 3. 폐암으로 새로이 진단 받고 항암화학치료를 받은 10명 환자들의 생존 기간이 다음 표와 같았다. 여기서 +로 표시된 것은 중도 절단된 자료를 가리킨다. 각 시점에서 생존 함수를 누적한계추정법을 이용하여 추정하여 보시오.

<표> 폐암환자에서 항암화학치료 시행 후 생존기간 (단위: 월)

3.5, 4.0+, 4.5, 4.5, 5.0, 6.0, 6.4+, 6.7, 7.0, 7.5
--

답)

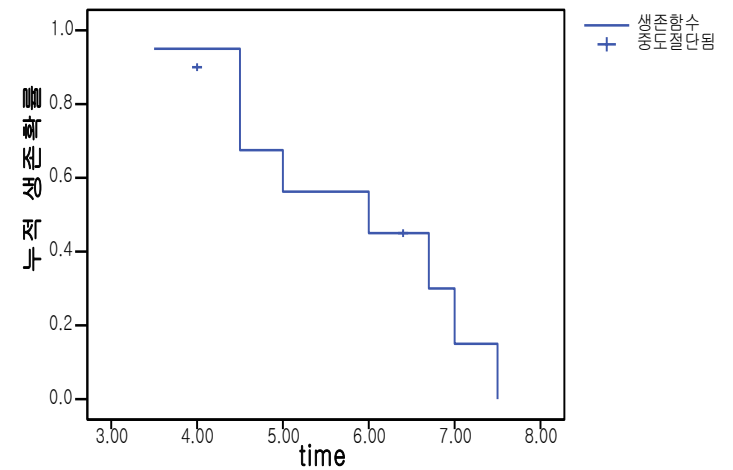
우선 가장 짧은 시간인 3.5에서의 생존확률, 즉 $S(3.5)=p(t>3.5)=p_1$ 의 추정치를 생각해 보자. 이 시점에서 위험그룹의 인원수는 10명이고 한 명이 이때 사망 q_1 (구간생존확률) $=1-1/10=0.9$ 이다. 다음 데이터는 4.0에서 중도절단되었으므로 이 환자의 생존 시간은 4.0보다는 크다. 따라서 이 시점에서의 사망자수는 0명이 되어 $S(4.0)=0.9$ 이다.

4.5에서는 이때까지 생존했던 8명 중 두 명이 사망하였으므로 구간 생존 확률은 $1-2/8=0.75$ 가 되고 이때의 생존함수 추정치는 $S(4.5)=0.9*0.75=0.675$ 가 된다. 이를 반복하면 생존함수를 구할 수 있게 된다.

폐암환자에서 항암화학치료 시행 후 누적한계 추정치

시점	생존분포의 추정치	사망률	누적사망자수	남은 인원수
0	1	0	0	10
3.5	0.9	0.1	1	9
4.0*	.	.	1	8
4.5	.	.	2	7
4.5	0.675	0.325	3	6
5	0.5625	0.4375	4	5
6	0.45	0.55	5	4
6.4*	.	.	5	3
6.7	0.3	0.7	6	2
7	0.15	0.85	7	1
7.5	0	1	8	0

생존함수



Rem 자료

제목없음 - SPSS 데이터 편집기

파일(F) 편집(E) 보기(V) 데이터(D) ...

1:t 3

	t	censor
1	3.00	1.00
2	4.00	.00
3	4.50	1.00
4	4.50	1.00
5	5.50	1.00
6	6.00	1.00
7	6.40	1.00
8	6.50	1.00
9	7.00	1.00
10	7.50	1.00
11	8.40	.00
12	10.00	1.00
13	10.00	.00
14	12.00	1.00
15	15.00	1.00
16		
17		

rem.xls

파일 -> 열기 -> 데이터 ->
파일형식 (엑셀)

분석 > 생존분석 >
Kaplan-Meier 생존분석 >
사건정의 (1) > 옵션 (그래프선택)

Survival Analysis for t t

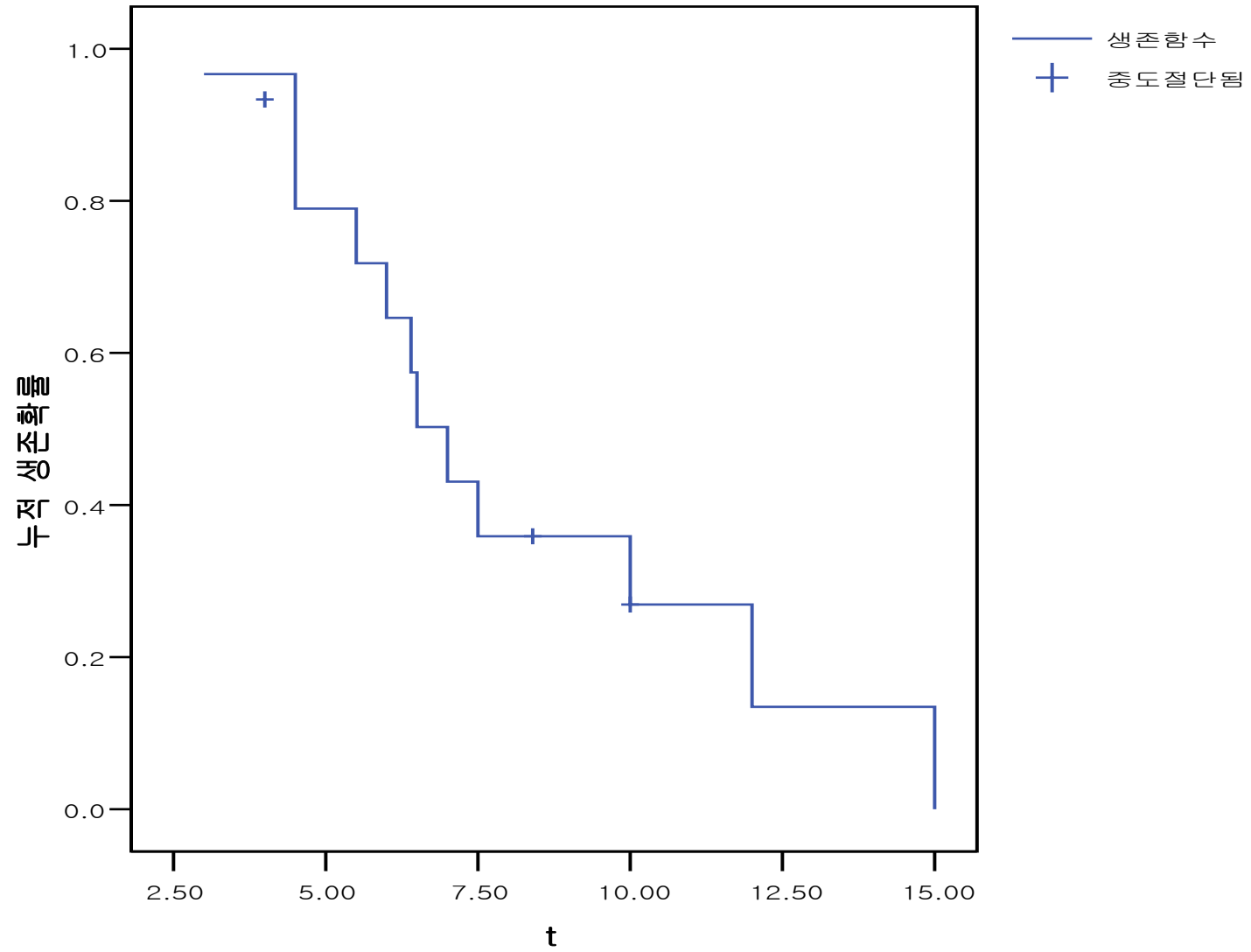
Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
3.00	1.00	.9333	.0644	1	14
4.00	.00			1	13
4.50	1.00			2	12
4.50	1.00	.7897	.1081	3	11
5.50	1.00	.7179	.1198	4	10
6.00	1.00	.6462	.1275	5	9
6.40	1.00	.5744	.1320	6	8
6.50	1.00	.5026	.1336	7	7
7.00	1.00	.4308	.1324	8	6
7.50	1.00	.3590	.1283	9	5
8.40	.00			9	4
10.00	1.00	.2692	.1237	10	3
10.00	.00			10	2
12.00	1.00	.1346	.1135	11	1
15.00	1.00	.0000	.0000	12	0

Number of Cases: 15 Censored: 3 (20.00%) Events: 12

	Survival Time	Standard Error	95% Confidence Interval
Mean:	8.17	1.06	(6.10, 10.24)
Median:	7.00	.55	(5.92, 8.08)

	Percentiles		
	25.00	50.00	75.00
Value	12.00	7.00	5.50
Standard Error	2.28	.55	1.25

생존함수



문제풀이

5. 다음은 40명의 급성심근경색 환자들의 1년 생존자료로 생존기간을 주수로 표현하였다.

2, 3, 3, 5,6, 7+, 8, 8, 9, 10, 10, 11, 11+, 12, 12, 13, 14, 14, 16,
17+, 20, 21, 23, 30, 39, 40, 40+, 42, 42, 44, 45, 46, 46, ,48,
48+, 52, 52+, 52+, 52+

(+ 는 중도절단 자료를 나타냄)

이 자료를 바탕으로 SPSS를 사용하여 구한 결과이다.

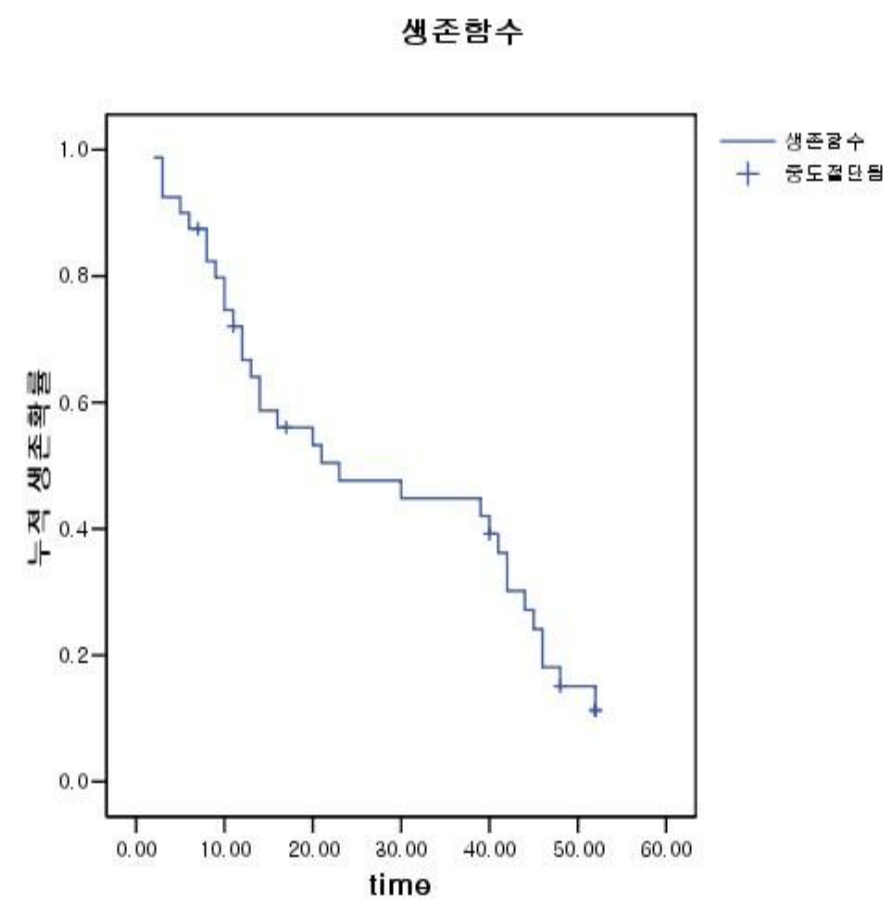
분석결과를 보고 다음 물음에 답하시오.

Survival Analysis for time

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
2.00	.00	.9750	.0247	1	39
3.00	.00			2	38
3.00	.00	.9250	.0416	3	37
5.00	.00	.9000	.0474	4	36
6.00	.00	.8750	.0523	5	35
7.00	1.00			5	34
8.00	.00			6	33
8.00	.00	.8235	.0606	7	32
9.00	.00	.7978	.0639	8	31
10.00	.00			9	30
10.00	.00	.7463	.0694	10	29
11.00	.00	.7206	.0716	11	28
11.00	1.00			11	27
12.00	.00			12	26
12.00	.00	.6672	.0756	13	25
13.00	.00	.6405	.0771	14	24
14.00	.00			15	23
14.00	.00	.5871	.0794	16	22
16.00	.00	.5605	.0802	17	21
17.00	1.00			17	20
20.00	.00	.5324	.0809	18	19
21.00	.00	.5044	.0814	19	18
23.00	.00	.4764	.0815	20	17
30.00	.00	.4484	.0814	21	16
39.00	.00	.4203	.0810	22	15
40.00	.00	.3923	.0803	23	14
40.00	1.00			23	13
41.00	.00	.3621	.0796	24	12
42.00	.00			25	11
42.00	.00	.3018	.0769	26	10
44.00	.00	.2716	.0749	27	9
45.00	.00	.2414	.0724	28	8
46.00	.00			29	7
46.00	.00	.1811	.0657	30	6
48.00	.00	.1509	.0613	31	5
48.00	1.00			31	4
52.00	.00	.1132	.0564	32	3
52.00	1.00			32	2
52.00	1.00			32	1
52.00	1.00			32	0

Number of Cases: 40 Censored: 8 (20.00%) Events: 32

	Survival Time	Standard Error	95% Confidence Interval
Mean:	27.46	2.97	(21.64, 33.27)
(Limited to 52.00)			
Median:	23.00	10.18	(3.04, 42.96)



a) 중도절단의 수는 얼마인가?

8명

b) 심근경색 환자의 50%가 생존하고 있는 기간은 몇 주인가?

23주, 95% CI (3.04, 42.96)

c) 자료에서 40명의 환자 중 20명의 환자가 사망하지 않고 추적 관찰되고 있는 시간은 17주이다. 생존함수에서 구한 환자의 50%가 사망하지 않고 생존하고 있을 기간과 다른 이유는 무엇인가?

누적한계추정치는 중도절단 시간에서는 변하지 않고 사망이 관찰되는 시점 t_i 에서 $(n_i - d_i)/n_i$ 를 곱해 준 만큼씩 감소하는 단계함수이다. 중도 절단의 효과는 각 단계에서의 크기 n_i 의 값에서 나타나게 된다.

따라서 중도 절단의 효과를 고려하지 않고 단순히 추적 관찰 중인 피수검자의 수가 50%남아 있는 것과 중도 절단의 효과를 고려한 누적한계추정치에 의한 중위수 값은 다르게 된다.

log-rank 검정: 생존함수의 비교 (비모수적 방법)

흑색종 환자들의 생존기간 (단위:월)

BCG 처리그룹	33.7+ 3.9 10.5 5.4 19.5 23.8+ 7.9 16.9+ 16.6+ 33.7+ 17.1+
CP 처리그룹	8.0 26.9+ 21.4+ 18.1+ 16.0+ 6.9 11.0+ 24.8+ 23.0+ 8.3 10.8+ 12.2+ 12.5+ 24.4 7.7 14.8+ 8.2+ 8.2+ 7.8+

- 두 그룹의 생존 함수를 총괄적으로 비교하기 위한 가설은

$H_0: S_1(t) = S_2(t)$ for all t

$H_a: S_1(t) \neq S_2(t)$ for some t

두 그룹을 섞은 후 $t_1 < t_2 < \dots t_k$ 으로 정리한 후 t_i 에서

	사망	생존	계
그룹 1	D_{1i}	$N_{1i} - D_{1i}$	N_{1i}
그룹 2	D_{2i}	$N_{2i} - D_{2i}$	N_{2i}
계	D_i	$N_i - D_i$	N_i

- N_{1i} , N_{2i} , D_i 가 고정되어 있다고 가정하면 D_{1i} 는 초기하분포를 따르게 되고 그 평균과 분산은

- $E(D_{1i}) = N_{1i} D_i / N_i = E_{1i}$

- $V(D_{1i}) = V_{1i}$

$$T = \frac{\{\sum_{i=1}^k (D_{1i} - E_{1i})\}^2}{\sum_{i=1}^k V_{1i}}$$

- 귀무가설하에서 T 는 자유도 1인 카이제곱 분포를 가지게 되고, T 의 값이 임계치보다 크게 되면 두 그룹의 생존함수가 같다는 귀무가설을 기각하게 된다.

- 로그-순위 검정 법은 각 시점에서 같은 가중치를 준다.

- 대안

$$T = \frac{\left\{ \sum_{i=1}^k w_i (D_{1i} - E_{1i}) \right\}^2}{\sum_{i=1}^k w_i^2 V_{1i}}$$

Gehan의 방법 $w_i = N_i / (N + 1)$ 인원수에 비례

Heavy censoring 이 있으면 급격히 감소

Peto/Prentice의 방법 $w_i = \tilde{S}(t_i)$

천천히 감소한다.

Tarone-Ware $w_i = \sqrt{N_i / (N + 1)}$

	t	censor	group
1	33.70	.00	1.00
2	3.90	1.00	1.00
3	10.50	1.00	1.00
4	5.40	1.00	1.00
5	19.50	1.00	1.00
6	23.80	.00	1.00
7	7.90	1.00	1.00
8	16.90	.00	1.00
9	16.60	.00	1.00
10	33.70	.00	1.00
11	17.10	.00	1.00
12	8.00	1.00	2.00
13	26.90	.00	2.00
14	21.40	.00	2.00
15	18.10	.00	2.00
16	16.00	.00	2.00
17	6.90	1.00	2.00
18	11.00	.00	2.00
19	24.80	.00	2.00
20	23.00	.00	2.00
21	8.30	1.00	2.00
22	10.80	.00	2.00
23	12.20	.00	2.00
24	12.50	.00	2.00
25	24.40	1.00	2.00
26	7.70	1.00	2.00
27	14.80	.00	2.00
28	8.20	.00	2.00
29	8.20	.00	2.00
30	7.80	.00	2.00
31			

- 자료: m1.xls
- 분석 > 생존분석 > Kaplan-Meier 생존분석 > 사건정의(1) , 요인에 group 입력 , 요인비교에서 로그-순위검정법 입력> 옵션 (그래프선택)

Survival Analysis for t t

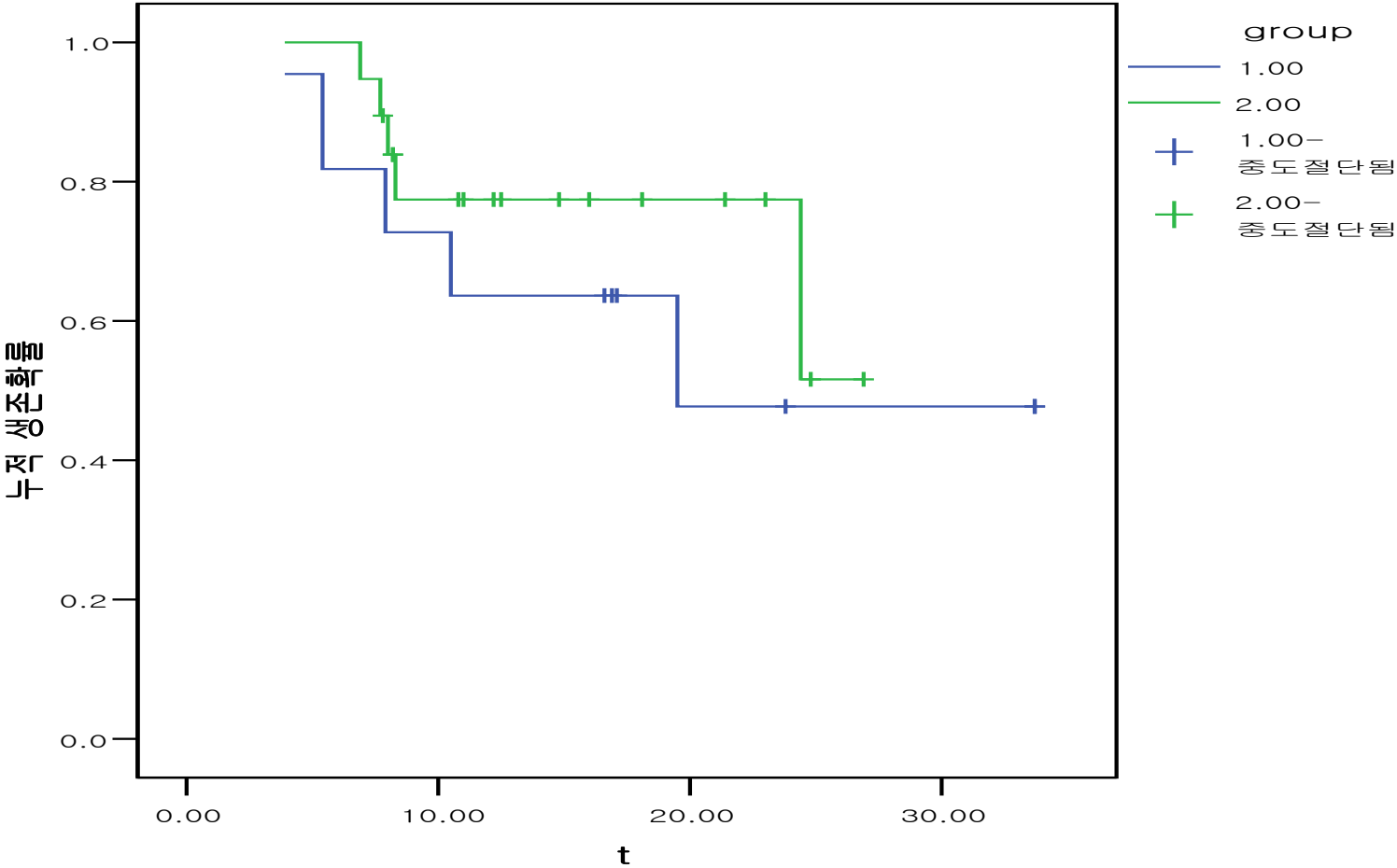
		Total	Number Events	Number Censored	Percent Censored
group	1.00	11	5	6	54.55
group	2.00	19	5	14	73.68
Overall		30	10	20	66.67

Test Statistics for Equality of Survival Distributions for group

	Statistic	df	Significance
Log Rank	.75	1	.3873
Breslow	1.00	1	.3183
Tarone-Ware	.94	1	.3328

↓
Gehan 또는
Wilcoxon

생존함수



4. 급성 골수구성 백혈병(AML; acute myelocytic leukemia)을 치료하기 위한 새로운 항암제가 개발되어 1차 임상시험을 마치고 2차 임상시험을 시행하고자 한다. 환자 23명을 두 군으로 나누어 한 군에는 신약을, 다른 한 군에는 기존의 항암제를 투여하고 일정 기간 관찰 후 생존율을 이용하여 항암제의 효과를 비교하고자 한다.

생존기간을 조사한 결과는 다음과 같다

신약 투여군(11명) 9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161

기존 항암제 투여군(12명) 5, 5, 8, 8, 12, 16+, 23, 27, 30, 33, 43, 45

(+ 는 중도절단 자료를 나타냄)

다음은 spss를 이용하여 생존분석을 시행한 결과이다. 치료 효과의 차이를 설명하시오.

Factor 화학요법 = 기존항암제

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
5.00	질병악화			11	
5.00	질병악화	.8333	.1076	2	10
8.00	질병악화			9	
8.00	질병악화	.6667	.1361	4	8
12.00	질병악화	.5833	.1423	5	7
16.00	중도절단			6	
23.00	질병악화	.4861	.1481	6	5
27.00	질병악화	.3889	.1470	7	4
30.00	질병악화	.2917	.1387	8	3
33.00	질병악화	.1944	.1219	9	2
43.00	질병악화	.0972	.0919	10	1
45.00	질병악화	.0000	.0000	11	0
Number of Cases: 12		Censored: 1	(8.33%)	Events: 11	
Survival Time		Standard Error	95% Confidence Interval		
Mean:	22.71	4.39	(14.11, 31.30)		
Median:	23.00	11.43	(.60, 45.40)		

Factor 화학요법 = 신약

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
9.00	질병악화	.9091	.0867	1	10
13.00	질병악화	.8182	.1163	2	9
13.00	중도절단			8	
18.00	질병악화	.7159	.1397	3	7
23.00	질병악화	.6136	.1526	4	6
28.00	중도절단			5	
31.00	질병악화	.4909	.1642	5	4
34.00	질병악화	.3682	.1627	6	3
45.00	중도절단			2	
48.00	질병악화	.1841	.1535	7	1
161.00	중도절단			0	
Number of Cases: 11		Censored: 4	(36.36%)	Events: 7	
Survival Time		Standard Error	95% Confidence Interval		
Mean:	52.65	19.83	(13.78, 91.51)		
(Limited to 161.00)					
Median:	31.00	7.36	(16.58, 45.42)		

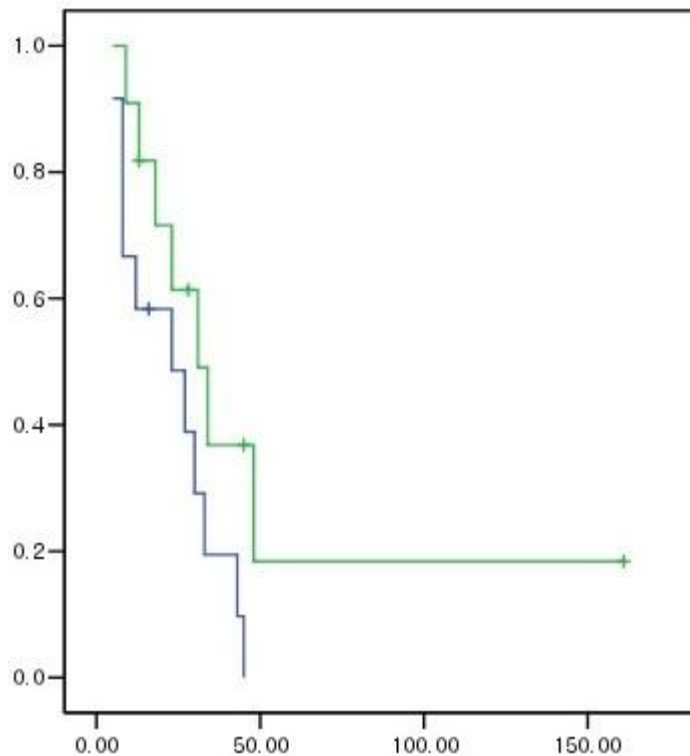
Survival Analysis for 시간 생존기간

Number	Percent	Total	Number Events
Censored	Censored		
화학요법	기존약	12	11
8.33			1
화학요법	신약	11	7
36.36			4
Overall		23	18
5	21.74		

Test Statistics for Equality of Survival Distributions for 화학요법

	Statistic	df	Significance
Log Rank	3.40	1	.0653
Breslow	2.72	1	.0989

생존함수



답)

기존 항암제를 사용하던 군에서는 한 건의 중도절단이 있었고, 새 항암제 군에서는 네 건의 중도절단이 있었다. Survival time이 신약군에서 더 긴 것으로 보이나 sample size가 작아서 95% 신뢰구간이 넓고 log rank test상 $p=0.065$ 으로 확인되었다. 더 많은 사례를 확보하면 p 값이 0.05떨어질 수도 있으므로 중도절단의 사유가 심각한 부작용에 의한 drop out이 아니라면 추가적인 연구를 생각해 볼 수 있다.

Cox의 비례위험 회귀모형 (준모수적 방법)

- 생존시간 T에 영향을 주는 변수(공변량)들 $x_1, x_2, \dots, x_k$ 가 있을 때

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \dots + \beta_k x_k)$$

$$\log h(t) = \log h_0(t) + \beta_1 x_1 + \dots + \beta_k x_k$$

- $h_0(t)$ 기저 위험함수 : 모든 x들이 0일 때의 위험함수
- i번째 환자와 j번째 환자의 위험비가 시간과 무관하게 상수가 된다.

$$h_i(t) / h_j(t) = \exp(\beta_1 (x_{i1} - x_{j1}) + \dots + \beta_k (x_{ik} - x_{jk}))$$

Cox모형의 계수 해석 (이산형)

$$h_i(t) / h_j(t) = \exp(\beta_1(x_{i1} - x_{j1}) + \dots + \beta_k(x_{ik} - x_{jk}))$$

만약 $x_1 = 1$ for male, 0 for female

$$\log h(t | \text{Male}) = \log h_0(t) + \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k \text{ for male}$$

$$\log h(t | \text{female}) = \log h_0(t) + \beta_2 x_2 + \dots + \beta_k x_k \text{ for female}$$

$$\log h(t | \text{male}) - \log h(t | \text{female}) = \log \frac{h(t | \text{male})}{h(t | \text{female})} = \beta_1$$

$$\frac{h(t | \text{male})}{h(t | \text{female})} = \exp(\beta_1) \quad \text{Hazard Ratio}$$

예: Beta=0.057, HR=1.059 : 다른 변수들의 값들이 일정할 때 성별만이 여에서 남으로 변한다고 하면 HR이 5.9% 증가한다.

Cox모형의 계수 해석(연속형)

$$h_i(t) / h_j(t) = \exp(\beta_1(x_{i1} - x_{j1}) + \dots + \beta_k(x_{ik} - x_{jk}))$$

만약 $x_1 = \text{연령}$

$$\log h(t \mid \text{Age} = x+1) = \log h_0(t) + \beta_1(x+1) + \beta_2x_2 + \dots + \beta_kx_k$$

$$\log h(t \mid \text{Age} = x) = \log h_0(t) + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k$$

$$\log h(t \mid \text{Age} = x+1) - \log h(t \mid \text{Age} = x) = \log \frac{h(t \mid \text{Age} = x+1)}{h(t \mid \text{Age} = x)} = \beta_1$$

$$\frac{h(t \mid \text{Age} = x+1)}{h(t \mid \text{Age} = x)} = \exp(\beta_1) \quad \text{Hazard Ratio}$$

예: Beta=0.027, HR=1.027 : 다른 변수들의 값들이 일정할 때 연령이 1 세 증가할 때마다 HR이 2.7% 증가한다.

PBC 자료

- Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984.
- 주요변수 : 생존시간, trt(D-penicillamine and placebo), age, sex 외에 임상, 생화학 지표

	n	time	delta	trt	age	sex	ascites	hepa	spider	edema	billi	chol
1	1.00	400.00	1.00	1.00	58.77	1.00	1.00	1.00	1.00	1.00	14.50	261.00
2	2.00	4500.00	.00	1.00	56.45	1.00	.00	1.00	1.00	.00	1.10	302.00
3	3.00	1012.00	1.00	1.00	70.07	.00	.00	.00	.00	.50	1.40	176.00
4	4.00	1925.00	1.00	1.00	54.74	1.00	.00	1.00	1.00	.50	1.80	244.00
5	5.00	1504.00	.00	.00	38.11	1.00	.00	1.00	1.00	.00	3.40	279.00
6	6.00	2503.00	1.00	.00	66.26	1.00	.00	1.00	.00	.00	.80	248.00
7	7.00	1832.00	.00	.00	55.53	1.00	.00	1.00	.00	.00	1.00	322.00
8	8.00	2466.00	1.00	.00	53.06	1.00	.00	.00	.00	.00	.30	280.00
9	9.00	2400.00	1.00	1.00	42.51	1.00	.00	.00	1.00	.00	3.20	562.00
10	10.00	51.00	1.00	.00	70.56	1.00	1.00	.00	1.00	1.00	12.60	200.00
11	11.00	3762.00	1.00	.00	53.71	1.00	.00	1.00	1.00	.00	1.40	259.00
12	12.00	304.00	1.00	.00	59.14	1.00	.00	.00	1.00	.00	3.60	236.00
13	13.00	3577.00	.00	.00	45.69	1.00	.00	.00	.00	.00	.70	281.00
14	14.00	1217.00	1.00	.00	56.22	.00	1.00	1.00	.00	1.00	.80	.
15	15.00	3584.00	1.00	1.00	64.65	1.00	.00	.00	.00	.00	.80	231.00
16	16.00	3672.00	.00	.00	40.44	1.00	.00	.00	.00	.00	.70	204.00
17	17.00	769.00	1.00	.00	52.18	1.00	.00	1.00	.00	.00	2.70	274.00
18	18.00	131.00	1.00	1.00	53.93	1.00	.00	1.00	1.00	1.00	11.40	178.00
19	19.00	4232.00	.00	1.00	49.56	1.00	.00	1.00	.00	.50	.70	235.00
20	20.00	1356.00	1.00	.00	59.95	1.00	.00	1.00	.00	.00	5.10	374.00
21	21.00	3445.00	.00	.00	64.19	.00	.00	1.00	1.00	.00	.60	252.00

자료 : A.xls

분석 > 생존분석 > Cox 회귀모형 > 시간변수, 상태변수, 공변량

케이스 처리 요약

	N	퍼센트
분석가능한 케이스	사건(a)	125 40.1%
	중도절단	187 59.9%
	전체	312 100.0%
삭제 케이스	결측 케이스	0 .0%
	음의 시간을 갖는 케이스	0 .0%
계층에서 가장 최근 사건 이전까지의	중도절단 케이스	0 .0%
	전체	0 .0%
	전체	312 100.0%

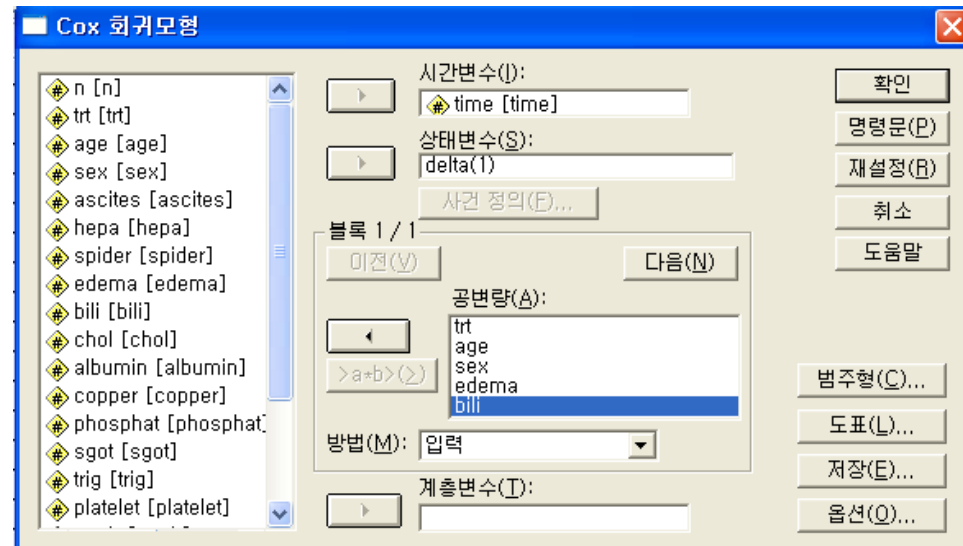
Trt=1(D-penicillamine)이 0에 비해서 위험율이 높지만 통계적인 유의성은 없다

a 종속변수 time: time

방정식의 변수

	B	표준오차	Wald	자유도	유의확률	Exp(B)	Exp(B)에 대한 95.0% CI	
							하한	상한
trt	.057	.179	.102	1	.750	1.059	.745	1.504

방정식의 변수



	B	표준오차	Wald	자유도	유의확률	Exp(B)	Exp(B)에 대한 95.0% CI 하한	상한
trt	.045	.188	.058	1	.810	1.046	.723	1.514
age	.036	.009	15.101	1	.000	1.036	1.018	1.055
sex	-.538	.245	4.815	1	.028	.584	.361	.944
edema	1.642	.296	30.680	1	.000	5.164	2.889	9.233
bili	.129	.015	78.532	1	.000	1.137	1.105	1.170

변환>변수계산 로그 변환 (bili, proth alb)

분석 > 생존분석 > Cox 회귀모형 > 시간변수, 상태변수, 공변량 stage
(범주형으로 표시)

방정식의 변수

B	표준오차	Wald	자유도	유의확률	Exp(B)	Exp(B)에 대한 95.0% CI		
						하한	상한	
trt	-.096	.186	.264	1	.607	.909	.631	1.308
age	.027	.009	7.904	1	.005	1.027	1.008	1.046
sex	-.430	.257	2.798	1	.094	.651	.393	1.077
edema	.821	.301	7.427	1	.006	2.274	1.259	4.105
logalb	-2.789	.770	13.128	1	.000	.062	.014	.278
logproth	3.129	1.213	6.653	1	.010	22.856	2.120	246.406
stage			4.163	3	.244			
변수 이름 stage(1)	-1.834	1.029	3.181	1	.074	.160	.021	1.199
변수 이름 stage(2)	-.341	.315	1.172	1	.279	.711	.383	1.319
변수 이름 stage(3)	-.188	.222	.714	1	.398	.829	.536	1.281
logbili	.839	.102	67.947	1	.000	2.313	1.895	2.824

변환>변수계산 로그 변환 (bili, proth alb)

분석 > 생존분석 > Cox 회귀모형 > 시간변수, 상태변수, 공변량 stage
(범주형으로 표시), 방법: 전진 Wald 선택

방정식의 변수

	B	표준오차	Wald	자유도	유의확률	Exp(B)	Exp(B)에 대한 95.0% CI		
							하한	상한	
단계 1	logbili	1.085	.093	135.092	1	.000	2.958	2.464	3.552
단계 2	logalb	-4.005	.676	35.061	1	.000	.018	.005	.069
단계 3	logbili	.983	.098	101.174	1	.000	2.673	2.207	3.238
	age	.036	.008	18.007	1	.000	1.036	1.019	1.053
	logalb	-3.675	.670	30.103	1	.000	.025	.007	.094
단계 4	logbili	.999	.096	107.520	1	.000	2.715	2.248	3.280
	age	.033	.009	14.648	1	.000	1.033	1.016	1.051
	logalb	-3.506	.671	27.330	1	.000	.030	.008	.112
단계 5	logproth	3.541	.968	13.395	1	.000	34.510	5.180	229.903
	logbili	.915	.097	88.621	1	.000	2.498	2.064	3.022
	age	.033	.009	14.757	1	.000	1.034	1.016	1.052
	edema	.785	.299	6.881	1	.009	2.192	1.219	3.939
	logalb	-3.053	.724	17.781	1	.000	.047	.011	.195
	logproth	3.016	1.024	8.676	1	.003	20.403	2.743	151.757
	logbili	.879	.099	79.299	1	.000	2.409	1.985	2.923

Cox 모형에서의 비례성 검정

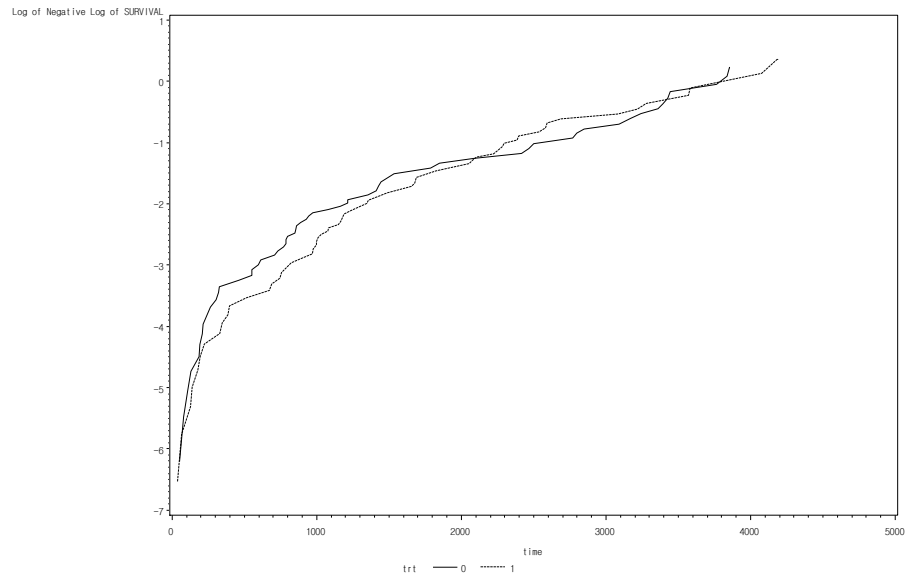
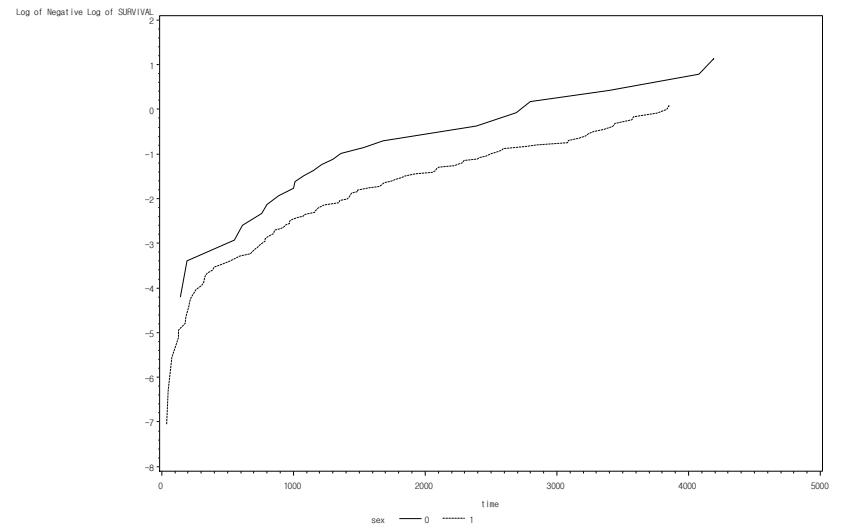
Plot $\log(-\log \hat{S}_{0q}(t))$ vs $t, q = 1, 2, \dots, Q$

If Parallel: OK

Parallel but not equally spaced : transformation needed

Schoenfeld residuals should be a random walk !

Cox 모형에서의 비례성 검정



Cox 모형

- 생존자료에서 다른 변수들의 효과를 보정한 후 trt효과를 볼 수 있는 가장 대표적인 통계모형
- Cox 모형의 결과는 HR (위험율)로 해석함
- 비례가정은 반드시 확인하는 것이 원칙임 -> 비례가정이 만족되지 않을 경우 time dependent covariate approach (예.Extended Cox 모형)

생존자료분석 요약

1. 생존율 산출

1. 생명표법: 표본수가 많을 때 (>50)
2. Kaplan-Meier method: 표본수가 적을 때 (< 50)

2. 생존율비교Mantel-Haenszel method

1. Mantel-Haenszel method
2. Log-rank method
3. Gehan's generalized Wilcoxon

3. 생존기간에 영향을 주는 인자에 대한 HR 추정

Cox proportional Hazard model

1. 단변량 분석 (평균, 표준편차, 비율 등 계산)

2. t-test , chi-square test (두 집단 비교)

3. 회귀분석, 로지스틱회귀 분석 (다른 변인들의 효과를 보정한 후의 주변수 효과)

피험자수 계산

시간형 결과의 비교 (생존분석 이용) 모수적 방법 (지수분포 이용)

- 총 사건수 $D = 4 \frac{(Z_{\alpha} + Z_{\beta})^2}{(\log(\theta))^2}$, $\Delta =$ 위험비
- 예) 검정력 90%를 가지고 위험비 2.0을 95% 유의수준의 양쪽 검정으로 유의하게 감지할 수 있는 검정을 위해서는

$$4 \frac{(1.96 + 1.282)^2}{(\log(2.0))^2} = 87.50 \approx 88 \text{ event가 필요하다.}$$

즉 실험군 비교군 각각 44개의사건이 필요하게 된다.
표7 과 동일한 결과

시간형 결과의 비교 (생존분석 이용) 비모수적 방법 (로그순위 검정 이용)

- 단형할당의 경우 총 사건수

$$D = 4 \frac{(Z_{\alpha} + Z_{\beta})^2 (\theta + 1)^2}{(\theta - 1)^2}$$

- 예) 검정력 90%를 가지고 위험비 1.75를 95% 유의수준의 양쪽 검정으로 위험비 1로부터 유의하게 감지할 수 있는 검정을 위해서는

$$D = \frac{(1.96 + 1.282)^2 (1.75 + 1)^2}{(1.75 - 1)^2} = 141 \text{ event} < \text{표 7} >$$

- 만약 30%가 실험 종료까지 사건이 발생하지 않는다면 (event free) (중도절단비, censoring rate=30%), 총표본수는 $141/0.30=202$ 로 주어진다.

- 표7. 로그 순위 검정을 이용한 비교를 위해 필요한 (두 집단) 총표본수 표 (위:모수적, 아래:비모수적 방법)

Δ	양측	$\beta=0.1$		$\beta=0.2$	
		$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.10$
1.25		844	688	630	496
		852	694	636	500
1.50		256	208	192	150
		262	214	196	154
1.75		134	110	100	80
		<u>142</u>	116	106	84
2.00		<u>88</u>	72	66	52
		94	78	70	56
2.25		64	52	48	38
		72	58	54	42
2.50		50	42	28	30
		58	48	42	34

김호

서울대학교 보건대학원

hokim@snu.ac.kr

[http:// plaza.snu. ac.kr/ ~hokim](http://plaza.snu.ac.kr/~hokim)

-> 열린강의실 -> 수업외 자료방 ->
의학연구자료의 생존분석법