SURVIVAL DATA ANALYSIS

[생존자료분석]

2013 FALL
MIDTERM EXAM SOLUTION

[2013년 2학기 중간고사 답안]

INSTRUCTOR : SEOKHO LEE

[강사 : 이석호]

- You may use calculators. Formula sheets are not allowed in this exam. [계산기 사용 가능합니다. 어떠한 참고자료도 허용되지 않습니다.]
- Carry out tests at level $\alpha = 0.05$ unless otherwise stated. [특별히 언급하지 않는한, 검정은 유의수준 5%하에서 수행합니다.]

Some Chi-Squared Percentiles

	Right-Tail Probability			
df	0.100	0.050	0.025	0.010
1	2.71	3.84	5.02	6.63
2	4.61	5.99	7.38	9.21
3	6.25	7.81	9.35	11.34
4	7.78	9.49	11.14	13.28
5	9.24	11.07	12.83	15.09
6	10.64	12.59	14.45	16.81
7	12.02	14.07	16.01	18.48
8	13.36	15.51	17.53	20.09
9	14.68	16.92	19.02	21.67
10	15.99	18.31	20.48	23.21

Some Normal Percentiles

Right-Tail Probability			
0.001	0.050	0.025	0.010
1.282	1.645	1.960	2.326

NAME (이름) :

UIN (학번) :

SIGNATURE (서명) :

1. (20 points) To justify Kaplan-Meier estimate of survivor function $S(t) = P(T > t)$, prove the following questions: Suppose $t_{(1)} < t_{(2)} < \cdots < t_{(j)}$ are the distinct, ordered failure times and let $t_{(0)} = 0$. [생존함수에 대한 카플란-마이어 추정량을 정당화하기 위해 다음의 질문에 답하시오. $t_{(1)} < t_{(2)} < \cdots < t_{(j)}$를 서로다른 순서화된 failure time이라 가정하자.]

   (a) (10 points) Show $P(T > t_{(j)}) = P(T > t_{(j-1)})P(T > t_{(j)}|T \geq t_{(j)})$

   (b) (10 points) Show $S(t_{(j)}) = \prod_{i=1}^{j} P(T > t_{(i)}|T \geq t_{(i)})$

2. (55 points) The data presented in the below table are preliminary results from a clinical trial to evaluate the efficacy of maintenance chemotherapy for acute myelogenous leukemia (AML). After reaching a status of remission through treatment by chemotherapy, the patients who entered the study were assigned randomly to two groups. The first group received maintenance chemotherapy; the second, or control, group did not. The objective of the trial was to see if maintenance chemotherapy prolonged the time until relapse. [아래 테이블에 주어진 자료는 급성골수성백혈병(AML)에 대한 화학요법을 지속함에 따른 효과를 평가하기 위해 실시한 임상실험으로부터의 초기결과이다. 실험대상은 화학요법을 통해 회복단계에 들어선 환자들로써, 실험을 위해 임의적으로 두 그룹으로 나누었다. 첫번째 그룹은 화학요법을 지속적으로 받게 하고, 두번째 그룹(control)은 회복단계에서 화학요법을 받지 않게 하였다. 이 실험의 목적은 화학요법을 지속적으로 받는 경우 재발할 때까지 걸리는 시간을 늘리는지 여부를 판단하는 것이다.]

Group	Complete remission time (in weeks) [재발시간 (단위:주)]
1. Maintained [유지]	9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+
2. Nonmaintained [중단]	5, 5, 8, 8, 12, 16+, 23, 27, 30, 33, 43, 45

(a) (10 points) Compute the average survival time ($\bar{T}_i$) and the average hazard rate ($\bar{h}_i$) for each group. Interpret these results. [평균생존시간 및 평균위험률을 각 그룹에 대해 계산하고 이를 해석하시오]

$\bar{T}_1 = 423/11 = 38.45$

$\bar{T}_2 = 255/12 = 21.25$

$\bar{h}_1 = 7/423 = 0.0165$

$\bar{h}_2 = 11/255 = 0.0431$

Interpretation:

(b) (10 points) Fill the missing information in the following tables of ordered failure time:
[정렬된 failure time에 대한 아래의 표에 빈칸을 채우시오]

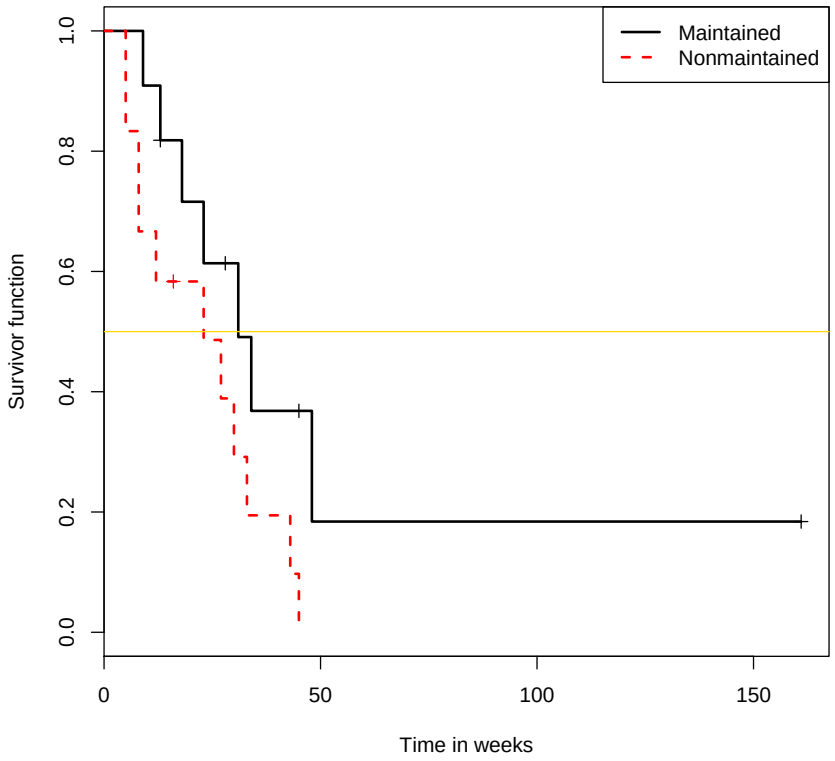
For 'Maintained' group :

$t_{1(j)}$	n_{1j}	m_{1j}	q_{1j}	$S_1(t_{(j)})$
0	11	0	0	1
9	11	1	0	$1 \times (11-1)/11 = 0.909$
13	10	1	1	$0.909 \times (10-1)/10 = 0.818$
18	8	1	0	$0.818 \times (8-1)/8 = 0.716$
23	7	1	1	$0.716 \times (7-1)/7 = 0.614$
31	5	1	0	$0.614 \times (5-1)/5 = 0.491$
34	4	1	1	$0.491 \times (4-1)/4 = 0.368$
48	2	1	1	$0.368 \times (2-1)/2 = 0.184$

For 'Nonmaintained' group :

$t_{2(j)}$	n_{2j}	m_{2j}	q_{2j}	$S_2(t_{(j)})$
0	12	0	0	1
5	12	2	0	$1 \times (12-2)/12 = 0.833$
8	10	2	0	$0.833 \times (10-2)/10 = 0.666$
12	8	1	1	$0.666 \times (8-1)/8 = 0.583$
23	6	1	0	$0.583 \times (6-1)/6 = 0.486$
27	5	1	0	$0.486 \times (5-1)/5 = 0.389$
30	4	1	0	$0.389 \times (4-1)/4 = 0.292$
33	3	1	0	$0.292 \times (3-1)/3 = 0.195$
43	2	1	0	$0.195 \times (2-1)/2 = 0.098$
45	1	1	0	$0.098 \times (1-1)/1 = 0$

(c) (10 points) Based on your results in part (b), draw the Kaplan-Meier curves for 2 groups on the below blank graph. [(b)의 결과에 기초하여 두 그룹에 대한 카플란-마이어 생존곡선을 아래의 빈 그래프에 그리시오.]



(d) (5 points) Compute the median survival time for each group. Interpret them. [생존시간의 중간값을 각 그룹에 대해 계산하고 이를 해석하시오.]
31 for 'Maintained' group; 23 for 'Nonmaintained' group

(e) (10 points) Compute the log rank test statistic for testing the equality of two survivor functions and complete the test. To do this, you should use the table in the next page. [두 생존함수의 동등성에 대한 검정을 log-rank 검정통계량을 계산하고 검정을 수행하시오. 이를 위해 다음 페이지에 주어진 테이블을 이용하시오.]
From the table in the next page, log rank statistic is $(3.71)^2/4.0075 = 3.4346$. Since $3.4336 < \chi_{0.05}(1) = 3.84$, we fail to reject the null hypothesis that two survivor functions are the same.

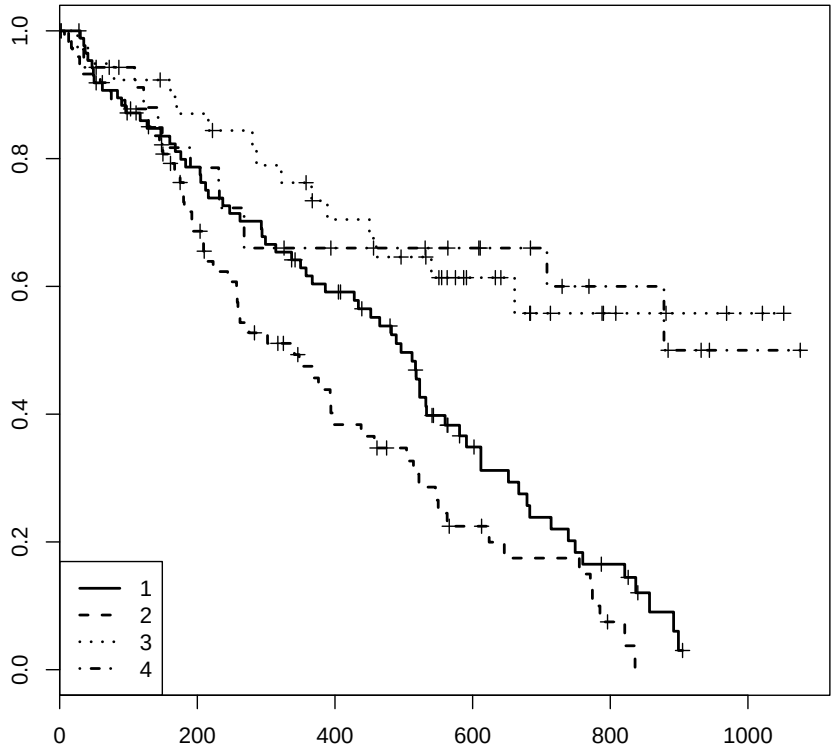
	# failures		# risk set		# expected		Observed-Expected		
$t_{(j)}$	m_{1j}	m_{2j}	n_{1j}	n_{2j}	e_{1j}	e_{2j}	$m_{1j}-e_{1j}$	$m_{2j}-e_{2j}$	variance
5	0	2	11	12	$(11/23)\times 2 = 0.96$	$(12/23)\times 2 = 1.04$	-0.96	0.96	$(11\times 12\times 2\times 21)/(23^2\times 22) = 0.4764$
8	0	2	11	10	$(11/21)\times 2 = 1.05$	$(10/21)\times 2 = 0.95$	-1.05	1.05	$(11\times 10\times 2\times 19)/(21^2\times 20) = 0.4739$
9	1	0	11	8	$(11/19)\times 1 = 0.58$	$(8/19)\times 1 = 0.42$	0.42	-0.42	$(11\times 8\times 1\times 18)/(19^2\times 18) = 0.2438$
12	0	1	10	8	$(10/18)\times 1 = 0.56$	$(8/18)\times 1 = 0.44$	-0.56	0.56	$(10\times 8\times 1\times 17)/(18^2\times 17) = 0.2469$
13	1	0	10	7	$(10/17)\times 1 = 0.59$	$(7/17)\times 1 = 0.41$	0.41	-0.41	$(10\times 7\times 1\times 16)/(17^2\times 16) = 0.2422$
18	1	0	8	6	$(8/14)\times 1 = 0.57$	$(6/14)\times 1 = 0.43$	0.43	-0.43	$(8\times 6\times 1\times 13)/(14^2\times 13) = 0.2449$
23	1	1	7	6	$(7/13)\times 2 = 1.08$	$(6/13)\times 2 = 0.92$	-0.08	0.08	$(7\times 6\times 2\times 11)/(13^2\times 12) = 0.4556$
27	0	1	6	5	$(6/11)\times 1 = 0.55$	$(5/11)\times 1 = 0.45$	-0.55	0.55	$(6\times 5\times 1\times 10)/(11^2\times 10) = 0.2479$
30	0	1	5	4	$(5/9)\times 1 = 0.56$	$(4/9)\times 1 = 0.44$	-0.56	0.56	$(5\times 4\times 1\times 8)/(9^2\times 8) = 0.2469$
31	1	0	5	3	$(5/8)\times 1 = 0.62$	$(3/8)\times 1 = 0.38$	0.38	-0.38	$(5\times 3\times 1\times 7)/(8^2\times 7) = 0.2344$
33	0	1	4	3	$(4/7)\times 1 = 0.57$	$(3/7)\times 1 = 0.43$	-0.57	0.57	$(4\times 3\times 1\times 6)/(7^2\times 6) = 0.2449$
34	1	0	4	2	$(4/6)\times 1 = 0.67$	$(2/6)\times 1 = 0.33$	0.33	-0.33	$(4\times 2\times 1\times 5)/(6^2\times 5) = 0.2222$
43	0	1	3	2	$(3/5)\times 1 = 0.60$	$(2/5)\times 1 = 0.40$	-0.60	0.60	$(3\times 2\times 1\times 4)/(5^2\times 4) = 0.2400$
45	0	1	3	1	$(3/4)\times 1 = 0.75$	$(1/4)\times 1 = 0.25$	-0.75	0.75	$(3\times 1\times 1\times 3)/(4^2\times 3) = 0.1875$
48	1	0	2	0	$(2/2)\times 1 = 1.00$	$(0/2)\times 1 = 0.00$	0.00	0.00	$(2\times 0\times 1\times 1)/(2^2\times 1) = 0.0000$
Total	7	11			10.71	7.29	-3.71	3.71	4.0075

(f) (10 points) Compute the Pearson's chi-square statistic and conduct the test for testing the null hypothesis. Is this result consistent with (e)? [피어슨 카이제곱통계량을 계산하고 귀무가설에 대한 검정을 수행하시오. 이 결과는 (e)에서 주어진 결과와 동일한가?]
$\frac{(7-10.71)^2}{10.71} + \frac{(11-7.29)^2}{7.29} = 3.1732$, which is, again, less than $\chi_{0.05}(1) = 3.84$. Thus, we fail to reject the null hypothesis that two survivor functions are the same. This result is consistent with the result in (e)

3. (25 points) The case study here explores an epidemiological study on the treatment of heroin addicts. The primary purpose of this study was to compare the retention time in four methadone treatment clinics for heroin addicts. A patient's survival time T was determined as the time in days until s/he dropped out of the clinic or was censored at the end of the study. [헤로인 중독의 치료에 대한 역학연구의 사례이다. 이 연구의 주목적은 메타돈 유치법(헤로인 중독의 치료 중에 대용 마약으로 메타돈을 줌)을 수행하는 4개의 병원에서 헤로인 중독자들이 포기하지 않고 치료를 계속 받는 시간을 병원별로 비교하는 것이다. 생존시간은 병원에서 치료를 지속하는 일수이고, censoring은 연구기간이 종료될 때까지 병원치료를 거부하지 않는 경우에 일어난다.)]

`Days.survival`	Survival time in days
`Status`	Censoring status (1='dropped out[도망]', 0='censored[끝까지 치료]')
`x`	Clinics (numbered 1∼4 [1∼4까지 4개 병원])

The below is the computer results from survival analysis using R. [다음은 R을 이용한 생존분석의 컴퓨터 결과이다.]




```
> survfit(Surv(Days.survival,Status)~x,data=addicts)
Call: survfit(formula = Surv(Days.survival, Status) ~ x, data = addicts)

    records n.max n.start events median 0.95LCL 0.95UCL
x=1      88    88      88     66    496     386     581
x=2      75    75      75     56    341     257     457
x=3      39    39      39     15     NA     540      NA
x=4      36    36      36     13    878     708      NA

> survdiff(Surv(Days.survival,Status)~x,data=addicts)
Call:
survdiff(formula = Surv(Days.survival, Status) ~ x, data = addicts)

     N Observed Expected (O-E)^2/E (O-E)^2/V
x=1 88       66     56.3      1.66      2.68
x=2 75       56     34.6     13.26     17.77
x=3 39       15     31.4      8.59     10.94
x=4 36       13     27.7      7.77      9.79

 Chisq= 32.6  on xxx degrees of freedom, p= xxxx
```

Based on the above results, answer the followings [위의 결과에 근거하여 다음에 답하라.]

(a) (5 points) How many individuals were dropped out for each group? And how many individuals are censored for each group? [각 그룹에서 얼마나 많은 중독자가 도망갔는가? 각 그룹에서 얼마나 많은 사람들이 censored 되었는가?]

	# of dead	# of censoring	total
Group 1	66	22	88
Group 2	56	19	75
Group 3	15	24	39
Group 4	13	23	36

(b) (10 points) Are survivor functions for 4 groups the same? [4개의 그룹의 생존함수는 동일한가?]
log rank statistic is given as 32.6, which is greater than $\chi_{0.05}(3) = 7.81$. We reject the null hypothesis that 4 survivor functions are the same. We conclude that 4 survivor functions are significantly different.

(c) (10 points) Interpret the plot of survivor function estimates. [생존함수 그래프 결과를 해석하라.]