

고급 생존분석

1장 서론

2020년 가을학기

전북대학교 통계학과

생존 분석

관심있는 반응변수가 생존시간(survival time)인 통계분석을 총체적으로 일컬음

- 생존시간 (survival time): 어떤 정의된 시점부터 특정한 사건(event)이 관측될 때까지의 시간 (time-to-event)

분석의 주요 관심 문제

- 생존함수 (survival function) 추정 : 예) 1년 생존율, 3년 생존율, 환자의 절반이 생존하는 시간, 두 처리군의 생존분포 비교
- 생존함수 또는 위험함수 (hazard function)에 영향을 주는 공변량 (covariate) 또는 예측변수를 구하고 연관 정도 규명

강의시간: 월 19:00-22:00 **담당교수:** 최 혜 미 (자연대본관 206호)

질문 및 상담: hchoi@jbnu.ac.kr, LMS 이용 또는 연구실 방문

통계 소프트웨어: R, Jupyter Notebook, SAS

교재 및 참고문헌

1. R을 이용한 생존분석 기초 (김재희 저) 자유아카데미, 2016
2. 생존분석 (김양진 저) 자유아카데미, 2017

평가계획:

| | | |
|------|-----|-----------------------|
| 출석 | 10% | 실시간 수업 출석과 녹화영상 시청 반영 |
| 중간고사 | 30% | 10/26, 월 |
| 기말고사 | 30% | 12/14, 월 |
| 과제 | 30% | 2주마다 과제물 제출 및 발표 예정 |

- R 과 SAS 프로그램 소개 및 서론
- 생존분석의 기초개념과 통계량
- 중도절단 및 우도함수
- 개수과정과 마팅게일
- 비모수 방법에 의한 생존함수 추정
- 생존함수 동일성 검정
- Cox 비례위험모형
- 프레일티 모형
- 경쟁위험모형

생존 데이터(time-to-event data)

의학, 생물학, 보건학, 역학, 공학, 경제학 등 여러 분야에서 발생

생존 시간의 예

- 수술 후 질병의 재발 시점 또는 사망시간, 다리 골절 완치까지의 시간 (의약학)
- 부품의 수명 (공학 또는 산업계)
- 결혼의 지속시간, 실업자들의 구직까지의 실업기간 (사회과학)
- 수감자들의 출소 후 재범까지 걸리는 시간 (범죄학)
- 잡지구독기간, 보험유지기간, 상품 재구매까지의 시간 (마케팅 분야)

생존 데이터(time-to-event data)

특성: 중도 절단 (censored)

- 연구기간 내에 사건이 발생하지 않을 수 있음
- 연구기간 내 추적 (follow up) 실패할 수 있음

3장에서 자세히..

기본 가정

- 추적에 실패한 개체도 연구 내에 있는 다른 개체와 같은 분포를 따름
- 중도절단과 사건발생은 서로 독립

Korean J Med 2019;94(2):182-190.

DOI: <https://doi.org/10.3904/kjm.2019.94.2.182>

비호지킨림프종에서 골수흡인 검사 결과에 따른 전체 생존율의 분석

최정윤¹, 강가원¹, 이병현¹, 유은상¹, 김대식¹, 이세련¹, 성화정¹, 김석진², 최철원¹, 김병수¹, 박용¹

<https://www.ekjm.org/journal/view.php?number=25601>

Korean J Med 2017;92(3):269-276.

DOI: <https://doi.org/10.3904/kjm.2017.92.3.269>

혈액투석 환자에서 혈관접근로의 동맥 미세석회화가 관상동맥 석회화와 심혈관 사망률에 끼치는 영향

이영복¹, 최보미¹, 황현석¹, 박훈석¹, 박순철², 김영옥¹

<https://www.ekjm.org/journal/view.php?number=25439>

www.prism.go.kr > downloadResearchAttachFile ▾ HWP

재범방지 프로그램의 효과성 평가 및 평가도구 개발

추가적으로 범죄발생 시기에 초점을 두어, 어떤 교정정책이 재범발생을 가장 지연시키는지 생존 분석(survival analysis)을 통해 관련 요인들을 검토한다. 달리 말하면 ...

설암 (tongue cancer) : `tonguecancer.csv`

- 구강암 환자의 배수성 (polyploidy: 염색체 수의 증감을 볼 수 있는 현상) 효과에 대한 예측연구를 위해 설암환자 데이터 수집
 - 조직샘플을 세포계산기를 사용하여 종양의 DNA 프로파일의 종류 파악 (이수성, 이배성)
1. 연구대상자 설암환자 80명 : 이수성 종양 52명, 이배성 종양 28명
 2. 데이터 사망까지의 시간 데이터
 3. 분석문제 그룹별 생존함수 추정, 그룹간 생존함수의 차이

생존 데이터 사례와 분석 문제

[표 1.2](a) 설암 데이터

| Aneuploid Tumors (이수성 종양) | | Diploid Tumors (이배성 종양) | |
|------------------------------|--------------------------|----------------------------|--------------------------|
| Death times | Censored Observations | Death times | Censored Observations |
| 1 | 61 | 1 | 8 |
| 3 | 74 | 3 | 67 |
| 3 | 79 | 4 | 76 |
| 4 | 80 | 5 | 104 |
| 10 | 81 | 5 | 176 |
| 13 | 87 | 8 | 231 |
| 13 | 87 | 12 | |
| 16 | 88 | 13 | |
| 16 | 89 | 18 | |
| 24 | 93 | 23 | |
| 26 | 97 | 26 | |
| 27 | 101 | 27 | |
| 28 | 104 | 30 | |
| 30 | 108 | 42 | |
| 30 | 109 | 56 | |
| 32 | 120 | 62 | |
| 41 | 131 | 69 | |
| 51 | 150 | 104 | |

(b) 설암 엑셀 데이터셋

| group | time | status |
|-------|------|--------|
| 1 | 1 | 1 |
| 1 | 3 | 1 |
| 1 | 3 | 1 |
| 1 | 4 | 1 |
| 1 | 10 | 1 |
| 1 | 13 | 1 |
| 1 | 13 | 1 |
| 1 | 16 | 1 |
| 1 | 16 | 1 |
| 1 | 24 | 1 |
| 1 | 26 | 1 |
| 1 | 27 | 1 |
| 1 | 28 | 1 |
| 1 | ⋮ | ⋮ |
| 1 | 104 | 1 |
| 1 | 157 | 1 |
| 1 | 167 | 1 |
| 1 | 61 | 0 |
| 1 | 74 | 0 |
| 1 | 79 | 0 |
| 1 | 80 | 0 |
| 1 | 81 | 0 |
| 1 | 87 | 0 |
| 1 | 87 | 0 |
| 1 | 88 | 0 |
| 1 | 89 | 0 |
| 1 | ⋮ | ⋮ |
| 2 | 1 | 1 |

생존 데이터 사례와 분석 문제

남성 후두암 (laryngeal cancer) : `table1.3b_laryngealcancer.csv`

1. 연구대상자 네덜란드 A 병원에서 1970년부터 1978년까지 후두암 진단받은 90명 남자환자의 첫번째 치료부터 사망사건까지의 시간 데이터 수집

2. 데이터

- 첫번째 치료이후 죽음 또는 연구의 끝 사이의 시간
- 생존시간, 중도절단 여부, 병의 단계, 진단 시 나이, 진단받은 년도
- 병의 단계(stage)별 환자 수
 - Stage (I): $T_1 N_0 M_0$ (33명)
 - Stage (II): $T_2 N_0 M_0$ (17명)
 - Stage (III): $T_3 N_0 M_0$, $T_x N_1 M_0$ ($x = 1, 2, 3$) (27명)
 - Stage (IV): TIS를 제외한 다른 모든 TNM조합 (13명)

3. 분석문제

- 병의 단계가 높을수록 사망확률이 높은지에 대한 추세검정 (trend test)
- 단계 효과와 생존률의 관계
- 비례 위험률 회귀모형을 통한 설명변수와 위험률의 관계

생존 데이터 사례와 분석 문제

단계 효과와 생존율의 관계, 비례 위험을 회귀모형을 통한 여러 설명변수들과 위험률의 관계 등에 대해 분석하고자 한다.

[표 1.3(a)] 남성 후두암 데이터

| I | | II | | III | | IV | |
|-------|----------|-------|----------|-------|----------|-------|----------|
| Death | Censored | Death | Censored | Death | Censored | Death | Censored |
| 0.6 | 2.5 | 0.2 | 2.2 | 0.3 | 3.7 | 0.1 | 2.9 |
| 1.3 | 3.2 | 1.8 | 2.6 | 0.3 | 4.5 | 0.3 | 4.3 |
| 2.4 | 3.3 | 2 | 3.3 | 0.5 | 4.8 | 0.4 | |
| 3.2 | 4.5 | 3.6 | 3.6 | 0.7 | 4.8 | 0.8 | |
| 3.3 | 4.5 | 4 | 4.3 | 0.8 | 5 | 0.8 | |
| 3.5 | 5.5 | 6.2 | 4.3 | 1 | 5.1 | 1 | |
| 3.5 | 5.9 | 7 | 5 | 1.3 | 6.5 | 1.5 | |
| 4 | 5.9 | | 7.5 | 1.6 | 8 | 2 | |
| 4 | 6.1 | | 7.6 | 1.8 | 9.3 | 2.3 | |
| 4.3 | 6.2 | | 9.3 | 1.9 | 10.1 | 3.6 | |
| 5.3 | 6.5 | | | 1.9 | | 3.8 | |
| 6 | 6.7 | | | 3.2 | | | |
| 6.4 | 7 | | | 3.5 | | | |
| 6.5 | 7.4 | | | 5 | | | |
| 7.4 | 8.1 | | | 6.3 | | | |
| | 8.1 | | | 6.4 | | | |
| | 9.6 | | | 7.8 | | | |
| | 10.7 | | | | | | |

[표 1.3(b)] 남성 후두암 액셀 데이터셋

| time | status | stage | age | year |
|------|--------|-------|-----|------|
| 0.6 | 1 | 1 | 77 | 76 |
| 1.3 | 1 | 1 | 53 | 71 |
| 2.4 | 1 | 1 | 45 | 71 |
| 3.2 | 1 | 1 | 58 | 74 |
| 3.3 | 1 | 1 | 76 | 74 |
| 3.5 | 1 | 1 | 43 | 71 |
| 3.5 | 1 | 1 | 60 | 73 |
| 4 | 1 | 1 | 52 | 71 |
| 4 | 1 | 1 | 63 | 76 |
| 4.3 | 1 | 1 | 86 | 74 |
| 5.3 | 1 | 1 | 81 | 72 |
| 6 | 1 | 1 | 75 | 73 |
| 6.4 | 1 | 1 | 77 | 72 |
| 6.5 | 1 | 1 | 67 | 70 |
| 7.4 | 1 | 1 | 68 | 71 |
| 2.5 | 0 | 1 | 57 | 78 |
| 3.2 | 0 | 1 | 51 | 77 |
| 3.3 | 0 | 1 | 63 | 77 |
| 4.5 | 0 | 1 | 48 | 76 |
| 4.5 | 0 | 1 | 68 | 76 |
| 5.5 | 0 | 1 | 70 | 75 |
| 5.9 | 0 | 1 | 47 | 75 |
| 5.9 | 0 | 1 | 58 | 75 |
| 6.1 | 0 | 1 | 77 | 75 |
| 6.2 | 0 | 1 | 64 | 75 |
| 6.5 | 0 | 1 | 79 | 74 |
| 6.7 | 0 | 1 | 61 | 74 |
| 7 | 0 | 1 | 66 | 74 |
| 7.4 | 0 | 1 | 73 | 73 |
| 8.1 | 0 | 1 | 56 | 73 |
| 8.1 | 0 | 1 | 73 | 73 |
| 9.6 | 0 | 1 | 58 | 71 |
| 10.7 | 0 | 1 | 68 | 70 |
| 0.2 | 1 | 2 | 86 | 74 |
| 1.8 | 1 | 2 | 64 | 77 |
| 2 | 1 | 2 | 63 | 75 |
| 3.6 | 1 | 2 | 70 | 77 |
| 4 | 1 | 2 | 81 | 71 |

골수이식 (bone marrow transplant) : transplant.csv

골수이식을 받은 백혈병(leukemia) 환자들에 대해 골수이식 방법에 따른 생존시간 연구

1. **연구대상자** 국제골수이식협회에 보고된 최신 골수성 백혈병을 가진 환자 101명
 - 자가(autologous) 골수이식 환자 51명
 - 동종(allogeneic) 골수이식 환자 50명
2. **데이터** 골수이식 방법에 따라 골수이식 후 백혈병 없는 생존기간(월)
3. **분석문제** 생존시간과 이식 후 무 백혈병 생존 (leukemia free survival) 시간을 이용하여 두가지 이식 방법의 효과 비교

생존 데이터 사례와 분석 문제

- 골수종 (myeloma) : myeloma.csv
- 신장염 (kidney infection) : kidney.csv
- 유방암 (breast cancer) : breastcancer.csv

R패키지에서 사용할 수 있는 데이터

survival과 KMsurv 패키지 등에 내장된 데이터

```
library(survival); try(data(package="survival"))
```

```
library(KMsurv); try(data(package="KMsurv"))
```

- 일차성 담즙성 간경화증 (primary biliary cirrhosis:PBC) :
survival 패키지 내 data(pbc)
- 심장이식수술 (heart transplant) : survival 패키지 내 data(heart)

과제

과제 제출: LMS에 9월 13일까지 올려주세요.

R과 SAS를 이용한 기초 분석

설암 (김은진), 일차성 담즙성 간경화증 (박선), 백혈병 골수 이식 (전바울), 골수종 (정다영), 신장염 (조진실), 심장이식수술 (지현준), 당뇨병 (황성윤) 데이터를 이용하여 다음을 구하시오

1. 그룹별로 관측값과 중도절단 개수와 비율
2. 그룹별로 관측값 데이터만 고려한 생존시간 평균
3. 그룹별로 관측값과 중도절단 데이터를 모두 포함한 생존시간 평균

사례 조사

신문, 인터넷, 논문 등에서 생존분석 사례를 찾고 다음을 기술하시오.

1. 데이터
2. 분석문제와 결과