

2조

심부전

생존분석

20200697 권소연

20201169 김소은

20201235 김하나

20210707 박혜인

목차

01 서론

02 데이터

03 모델링

03-1. 와이블 분포

03-2 비모수적 방법 검정

03-3. 로그순위검정

03-4. Cox비례모형

04 모형 평가

05 결론

서론



심부전

심부전이란 각종 심장질환으로 인해 심장에 구조적 혹은 기능적으로 이상이 생겨 심실의 혈액 충만 또는 박출에 이상이 발생한 상태



연구 이유

- 심부전 치료는 상당한 경제적 부담 동반
- 주요 위험 요인을 사전에 파악하면 조기 개입으로 질병 진행을 늦추거나 예방



생존분석

- 생존분석을 통해 위험 요인을 파악 -예후를 예측하여 환자 특성에 맞는 치료 방안을 제안

데이터

- 파키스탄 파이살라바드에 위치한 심장병원(Institute of Cardiology)과 알라이드 병원(Allied Hospital)에 입원한 심부전 환자들 데이터 기반의 심부전 데이터
- 여성: 105명, 남성: 194명, 총합: 299명
- 40세 이상
- 좌심실 수축 기능 부전 존재
- NYHA 3급 및 4급에 속함 (***NYHA 3급**: 환자는 안정 시에는 괜찮지만, 일상적인 신체 활동시 증상을 유발
NYHA 4급: 환자는 안정 상태에서도 피로, 숨참, 흉통 등 증상이 나타남. 매우 제한적인 활동만 가능하며, 주로 누워서 생활해야 함.)
- 추적 기간: 4일~ 285일 사이 (평균 130일)

출처: <https://archive.ics.uci.edu/dataset/519/heart+failure+clinical+records>

데이터

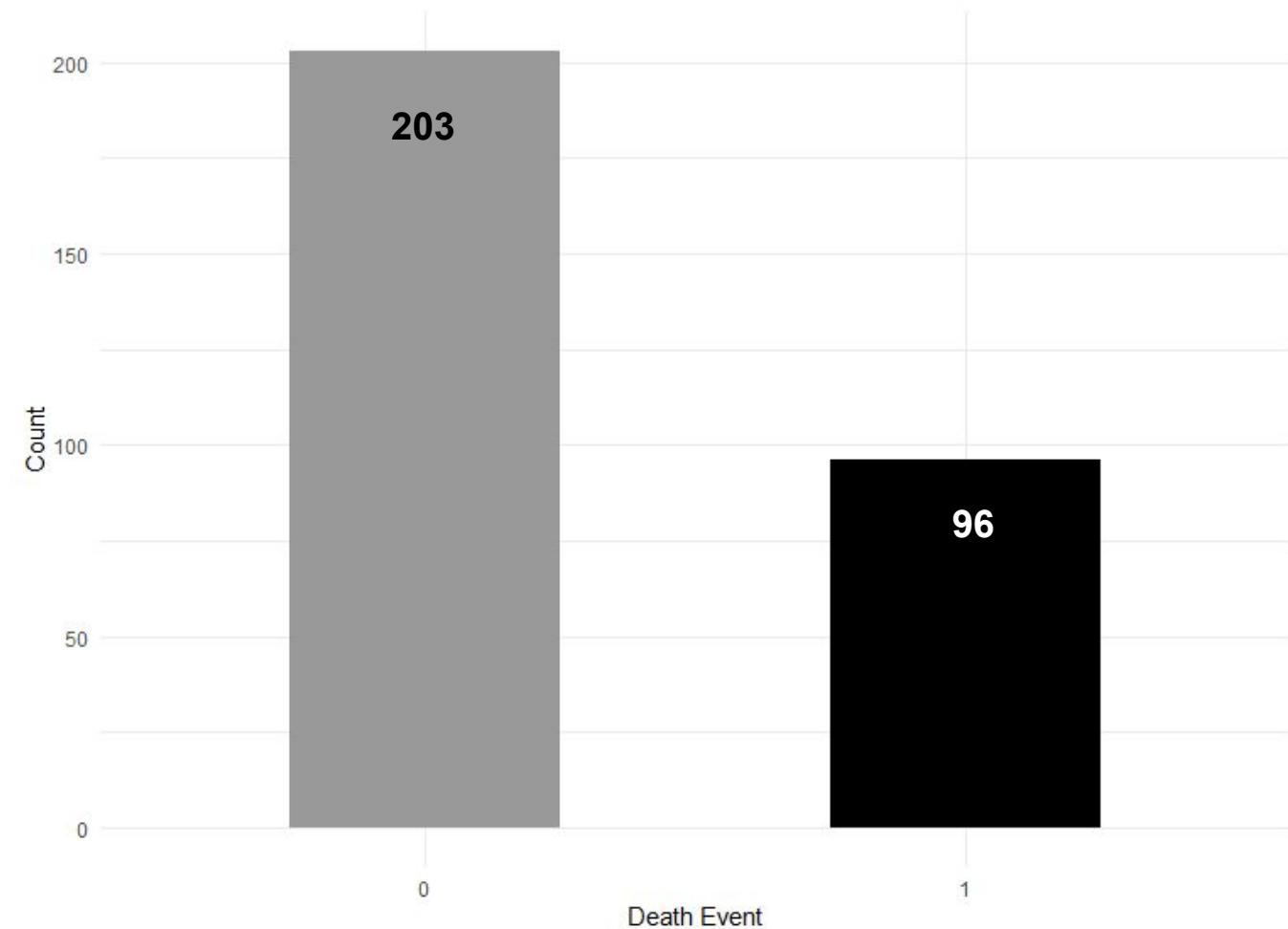
변수	변수 의미	단위
time	치료기간	Days (4-285)
DEATH_EVENT	추적 관찰 기간 동안 사망 여부	사망: 1 생존: 0
age	환자의 나이	Years (40-95)
creatinine_phosphokinase	혈액 내 CPK 효소의 수치	mcg/L (23-7861)
ejection_fraction	박출율: 심장 수축 시 심장에서 나가는 혈액의 비율	Percentage (14-80)
platelets	혈액 내 혈소판의 수	kiloplatelets/mL (25100-850000)
serum_creatinine	혈청크레아디닌: 혈액 내 크레아티닌 수치	mg/dL (0.5-9.4)
serum_sodium	혈청나트륨: 혈액 내 나트륨 수치	mg/dL (113-148)

[표1] 연속형 변수

변수	변수 의미	범주(빈도)
anaemia	빈혈: 적혈구 또는 헤모글로빈의 감소	빈혈: 1 (170) 빈혈X: 0
diabetes	환자의 당뇨병 유무	당뇨병: 1 (125) 당뇨병X: 0 (174)
high_blood_pressure	환자의 고혈압 유무	고혈압: 1 (105) 고혈압X: 0 (194)
sex	성별: 남성 또는 여성	남성: 1 (194) 여성: 0 (105)
smoking	환자의 흡연 여부	흡연: 1 (96) 흡연X: 0 (203)

[표2] 범주형 변수

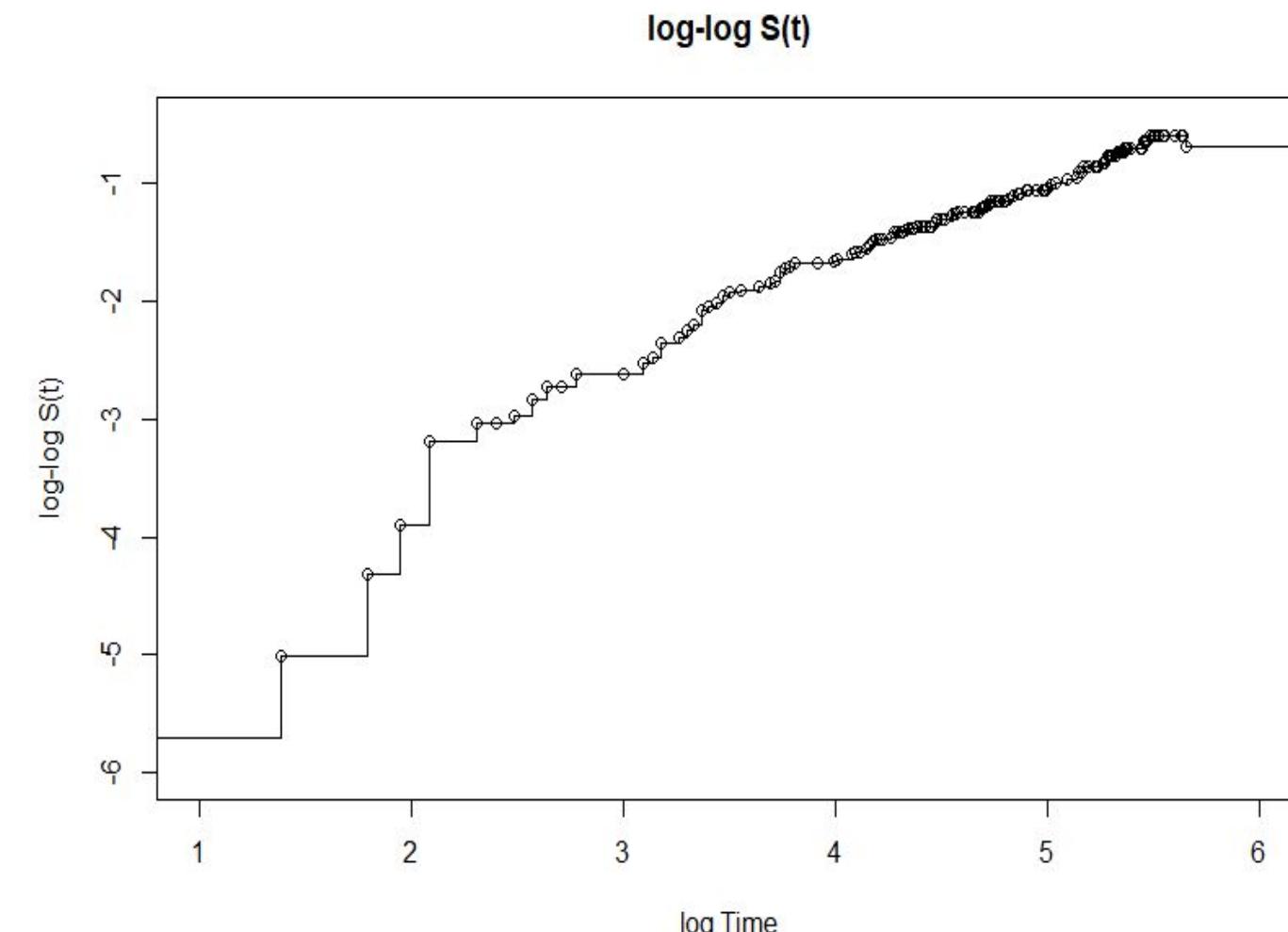
데이터



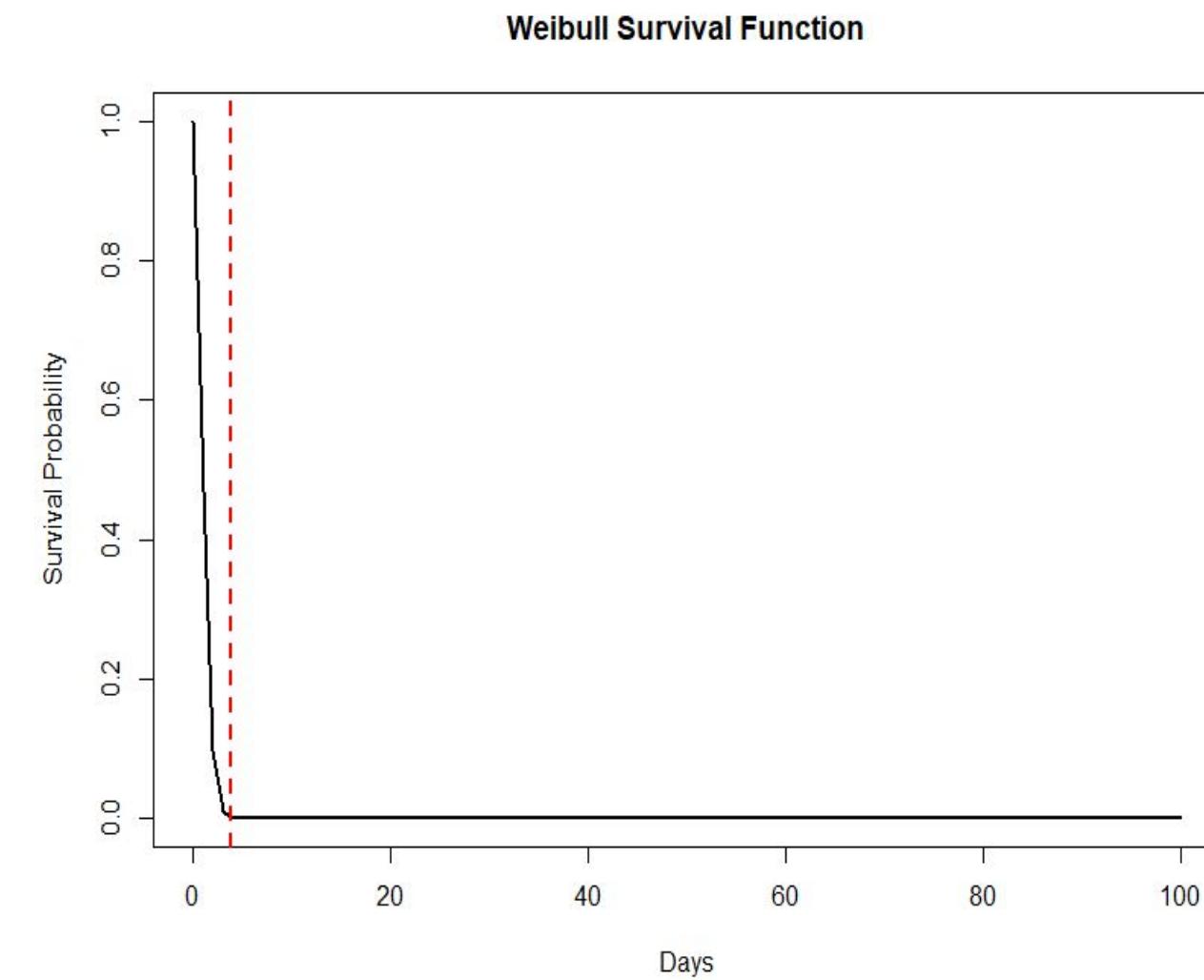
연속형 변수 변환

- 혈청크레아티닌
 > 1.5 신장 기능 장애 | ≤ 1.5 정상
- 혈청나트륨
 ≥ 135 정상 | < 135 저나트륨혈증
- 심박출율
 ≤ 30 심장 기능 매우 저하(Level 1)
30~45 경미한 저하(Level 2)
 > 45 정상(Level 3)

와이블 분포

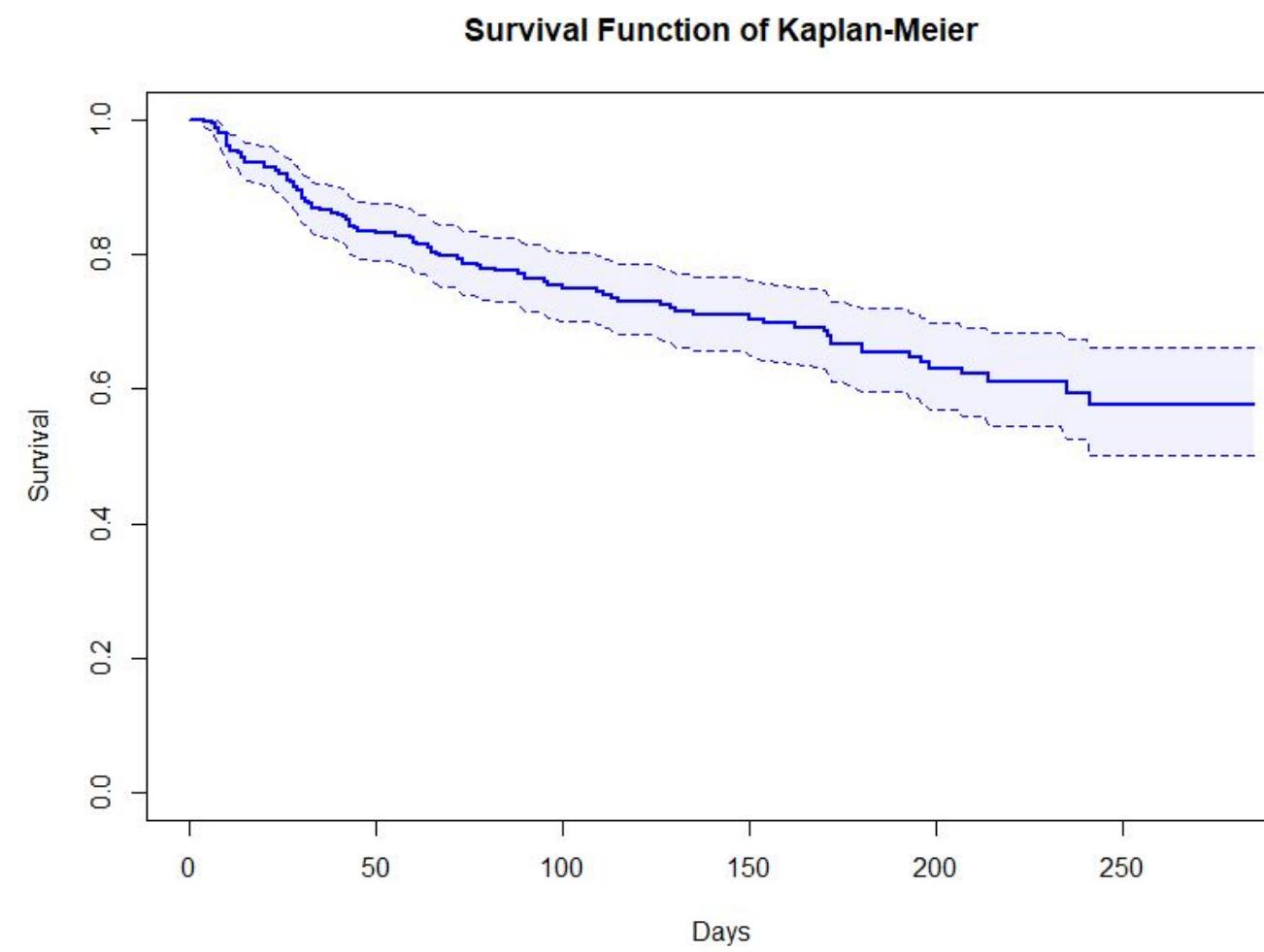


[log 분포 적합성 그래프]

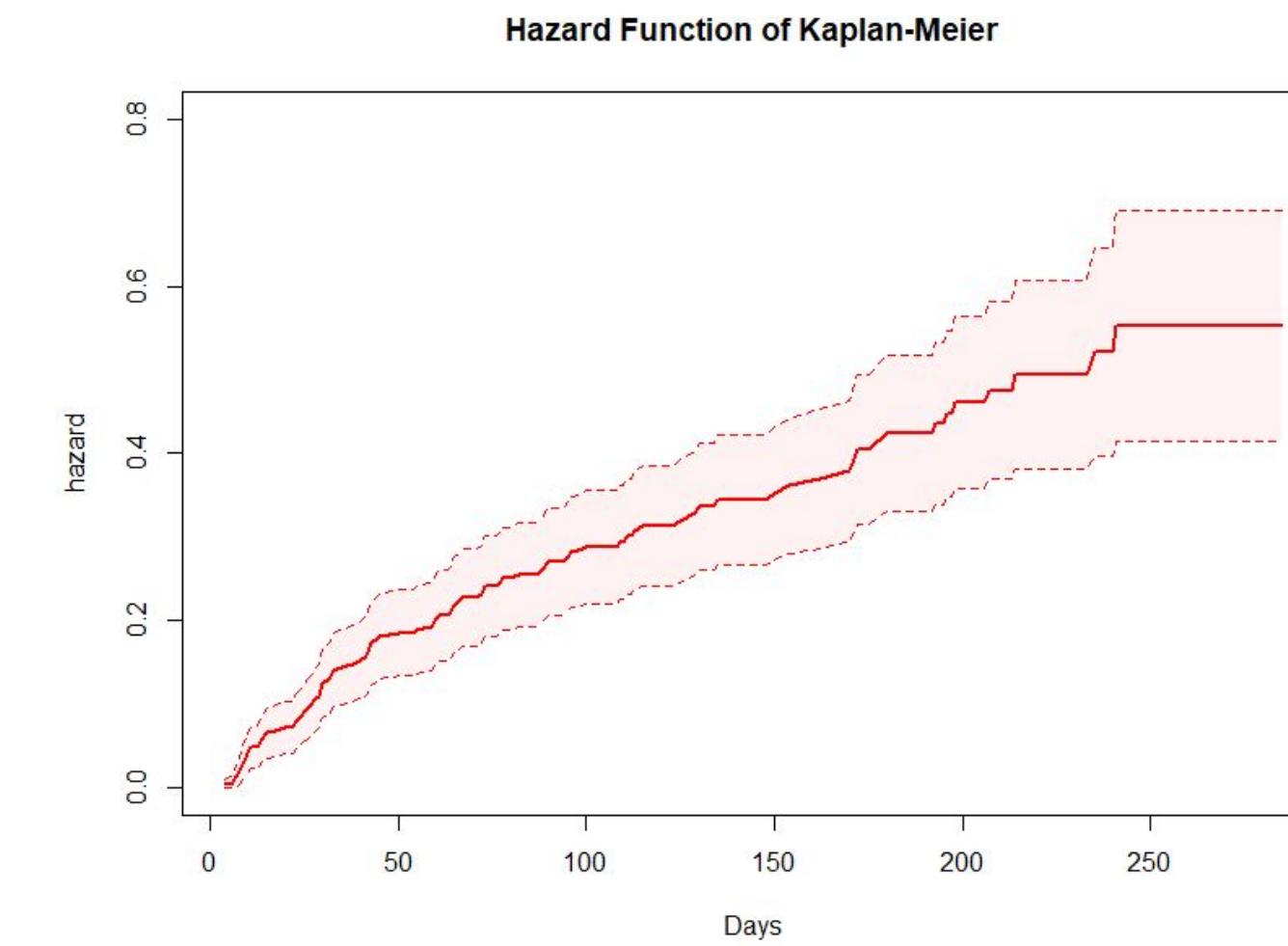


[와이블 분포 생존함수]

Kaplan-Meier

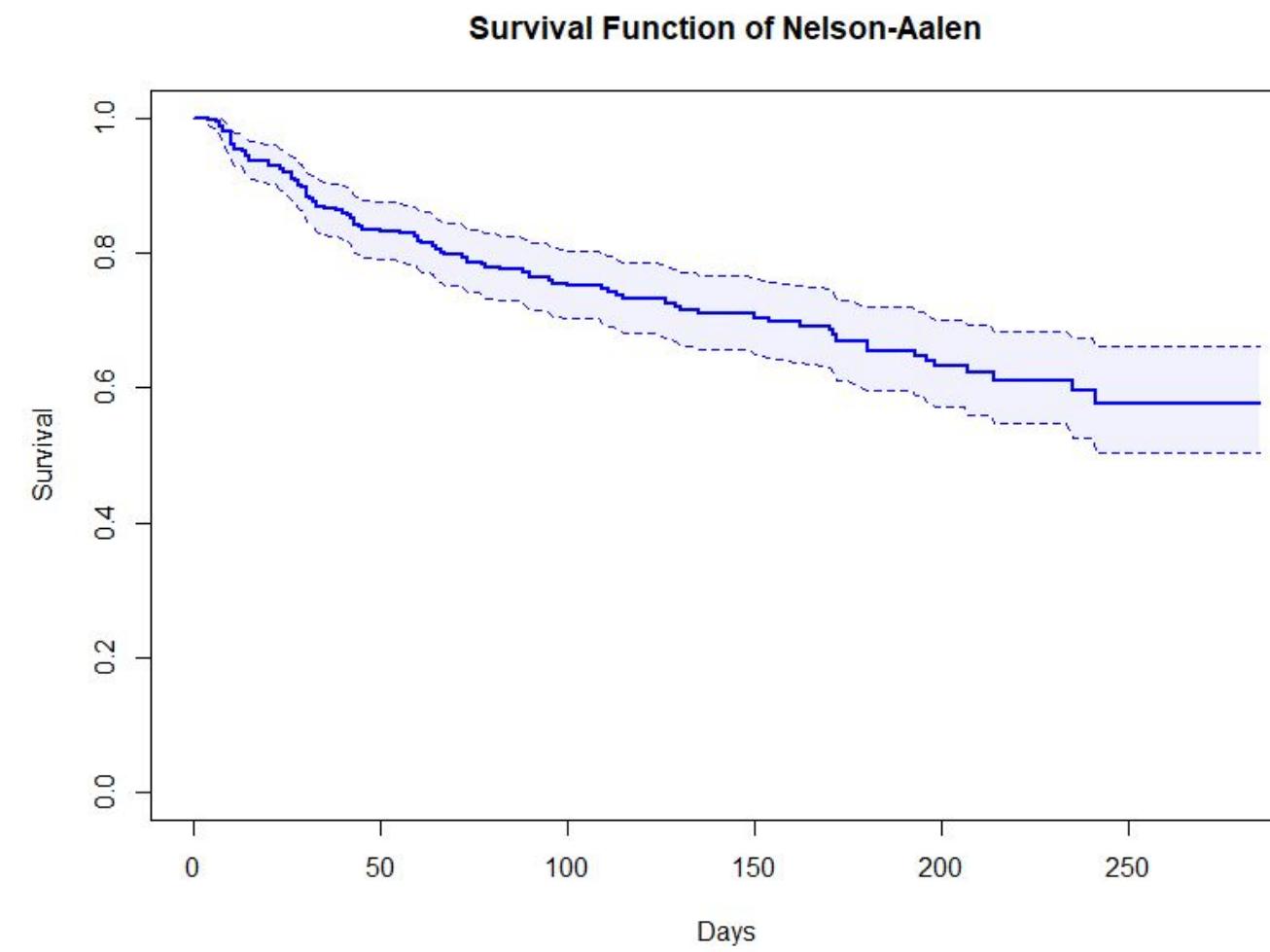


[Kaplan-Meier 생존함수]

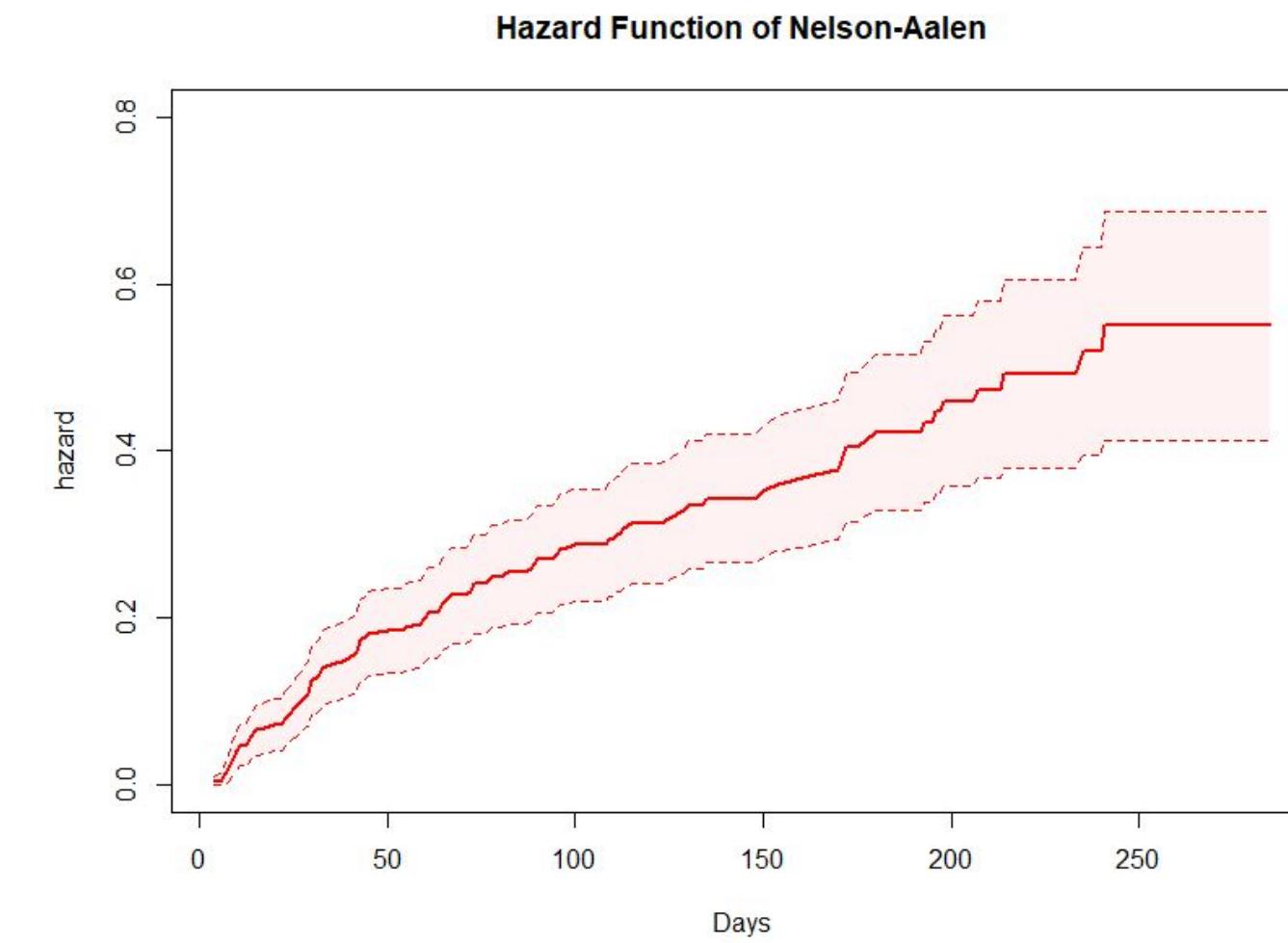


[Kaplan-Meier 누적위험함수]

Nelson-Aalen

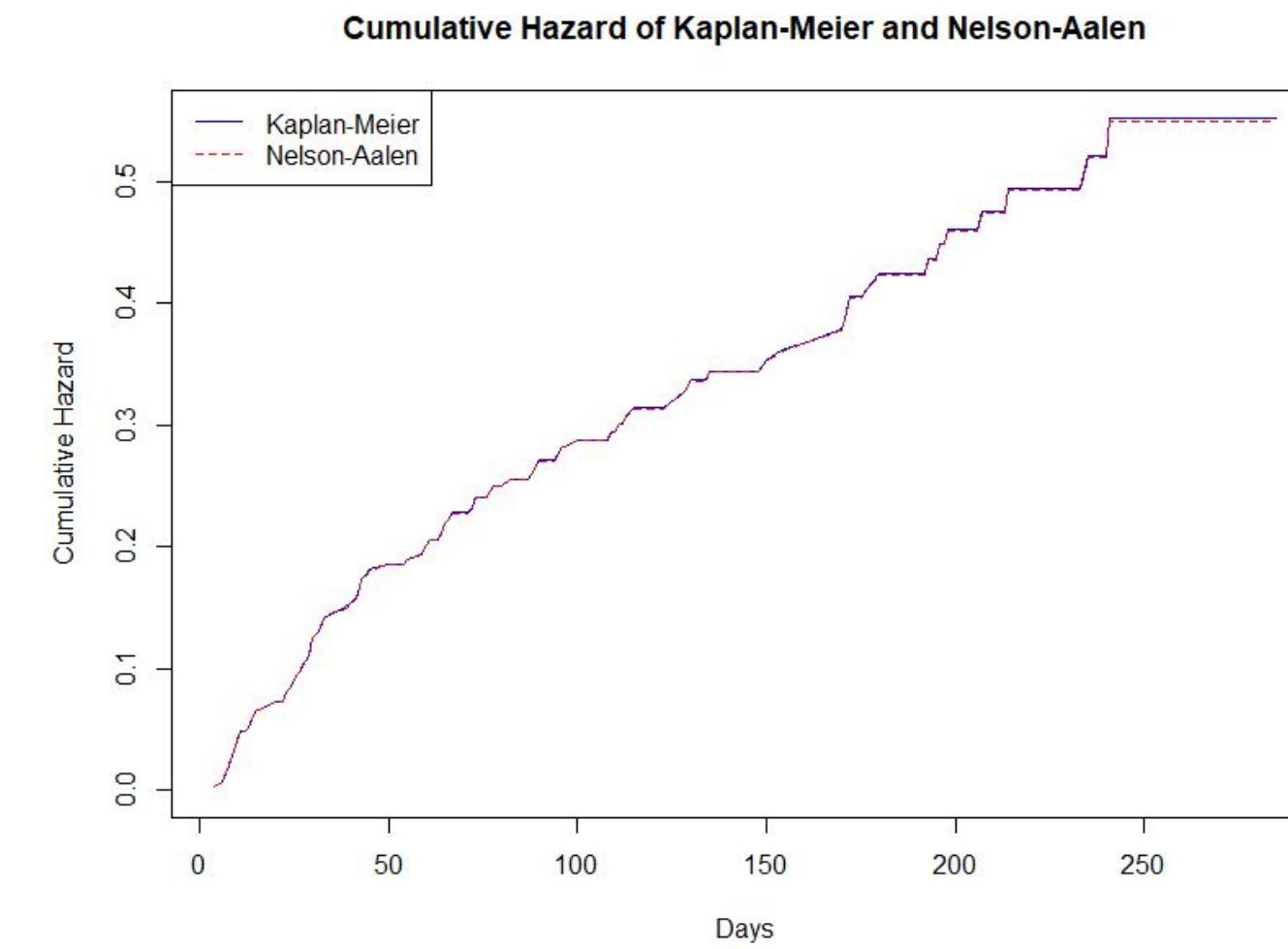
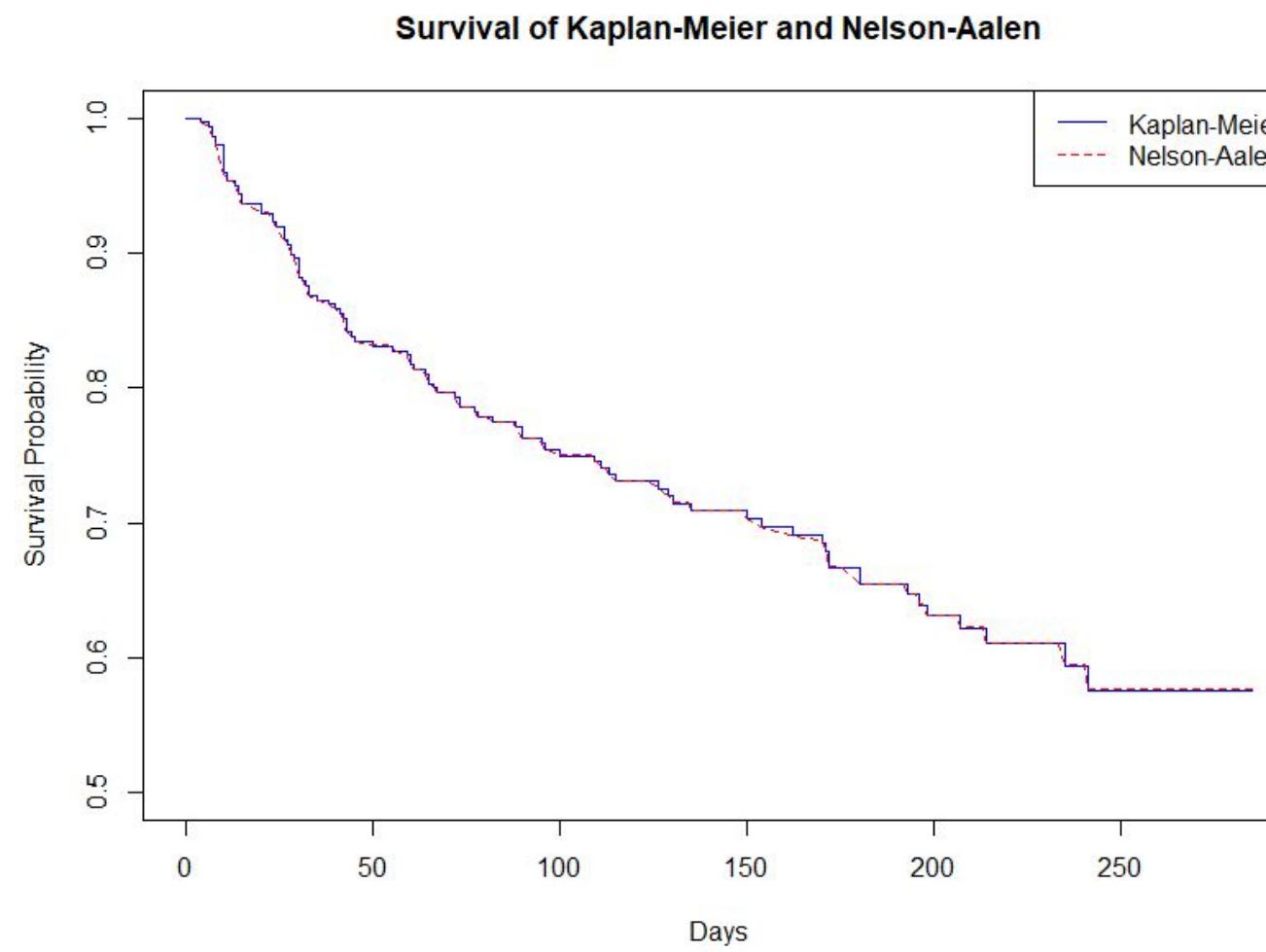


[Nelson-Aalen 생존함수]



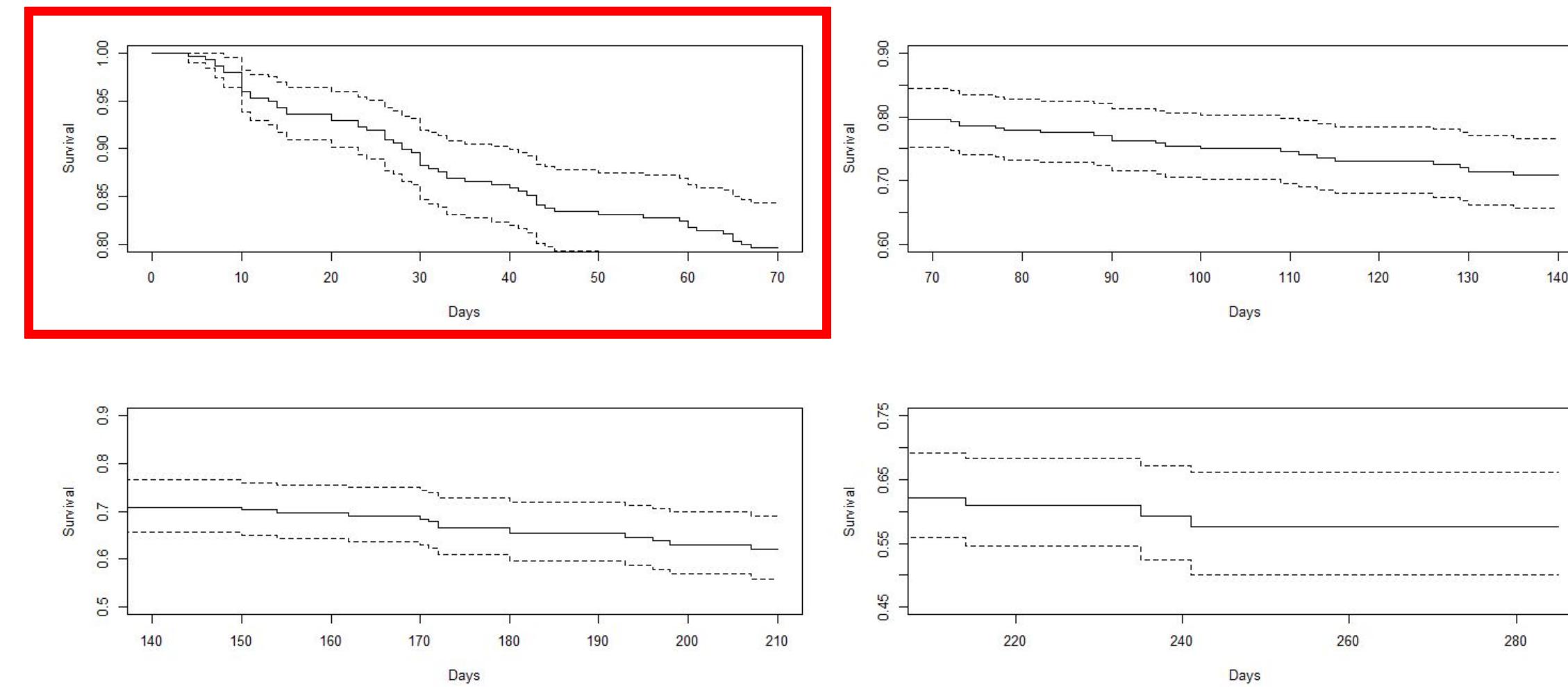
[Nelson-Aalen 누적위험함수]

Kaplan-Meier vs Nelson Aalen



[Kaplan-Meier, Nelson-Aalen 생존함수와 누적위험함수 비교]

Kaplan-Meier vs Nelson Aalen



[시간에 따른 Kaplan-Meier 생존함수]

로그순위검정

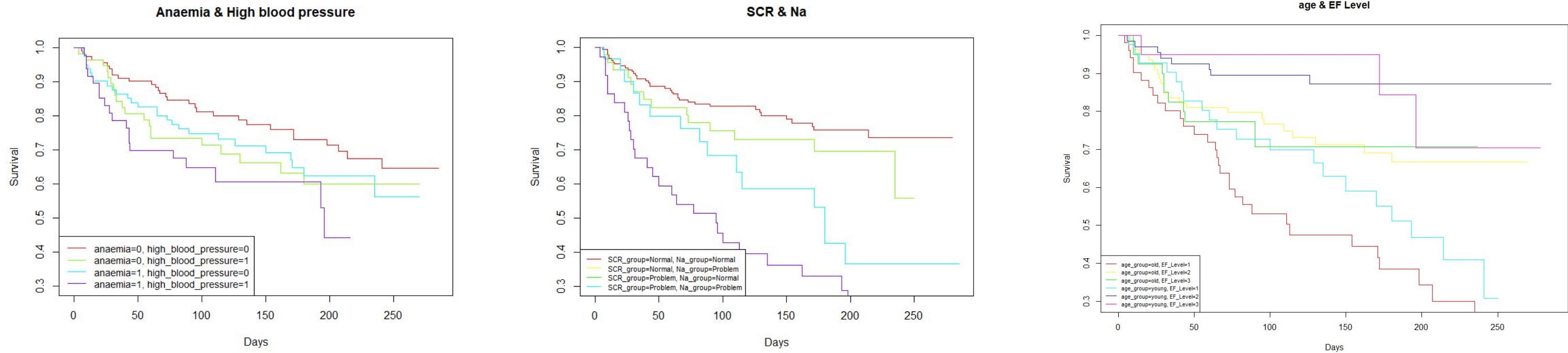
- 연속형 변수 변환

- 나이
< 60 적음 | >=60 많음
- CPK 효소
<250 정상 | 250~1000 스트레스 | >1000 심각
- 혈소판
<=Q1 적음 | Q1~Q3 적당 | > Q3 많음

변수	검정통계량	p-value
크레아티닌	39.612	0.000
심박출량	31.053	0.000
혈청나트륨	15.65	0.000
나이	8.565	0.003
고혈압	4.406	0.036
빈혈	2.727	0.099
혈소판	3.453	0.178
CPK 효소	0.885	0.643
당뇨병	0.041	0.841
성별	0.004	0.95
흡연	0.002	0.964

[표] 로그순위검정 결과

로그순위검정



[유의한 변수들 조합별 Kaplan-Meier 생존함수]

변수간의 상호작용도 생존에 영향

Cox비례위험모형

변수 선택으로 • 모형 1 = p값 유의 기준

모형 비교 • 모형 2= 단계적 선택 방법 (AIC 기준)

- 모형1 p값 유의

나이, CPK효소, 고혈압, 크레아티닌, 심박출량

- 모형2 단계적 선택방법

나이, CPK효소, 고혈압, 크레아티닌, 심박출량 + **빈혈, 혈청나트륨**

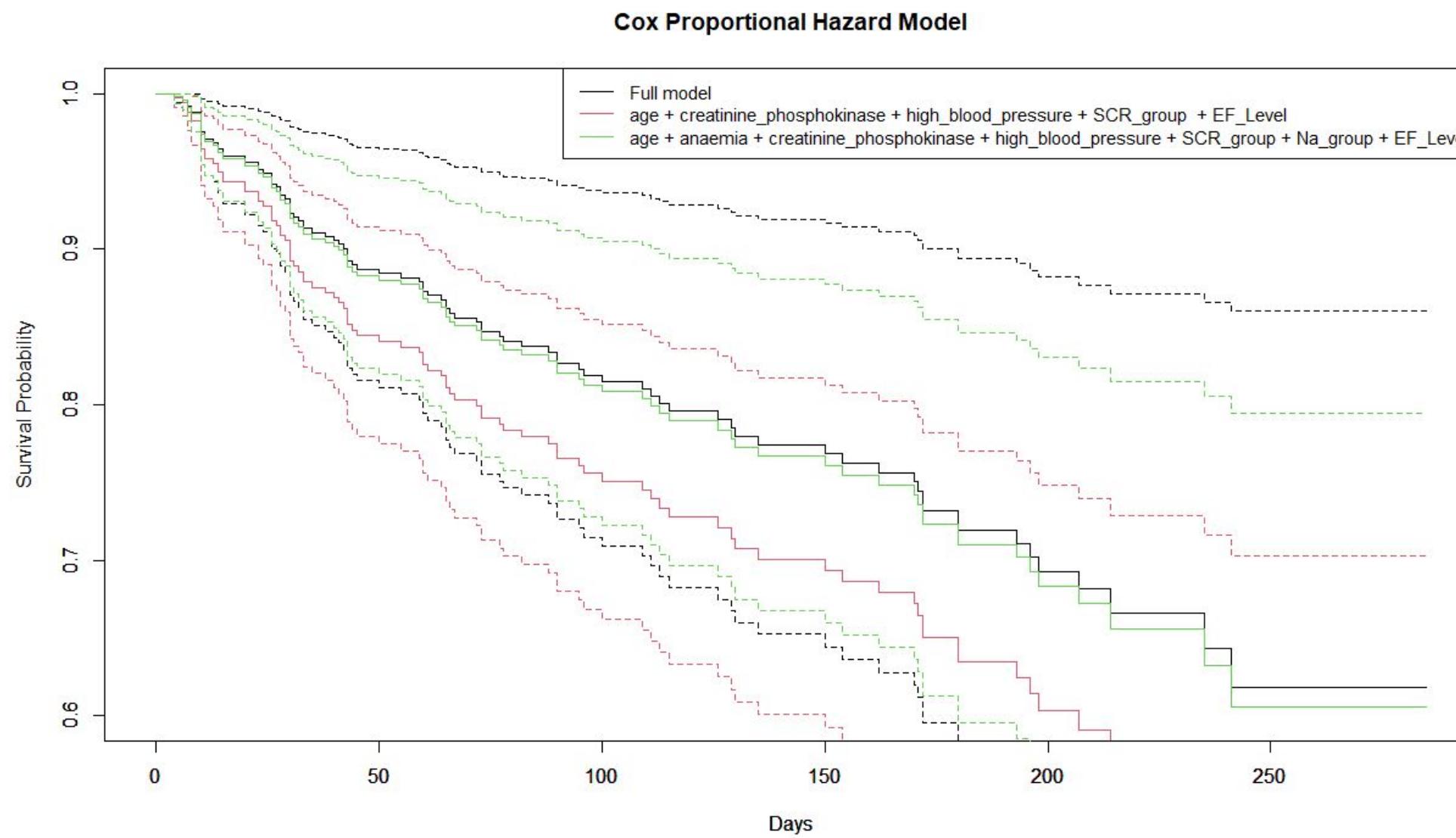
Cox비례위험모형

	coef	exp(coef)	lower .95	upper .95	z	p-value
나이	0.0412	1.0421	1.0237	1.0608	4.537	0
심박출량_Level2	-1.01	0.366	0.2267	0.5908	-4.114	0
크레아티닌_Problem	0.823	2.2768	1.4399	3.6	3.519	0.0004
CPK효소	0.00026	1.0003	1.0001	1.0005	2.61	0.0091
고혈압	0.519	1.6797	1.1039	2.556	2.422	0.0155
심박출량_Level3	-0.772	0.4621	0.2452	0.8708	-2.388	0.0169
빈혈	0.37	1.4476	0.9438	2.2202	1.695	0.0901
혈청나트륨_Problem	0.345	1.4123	0.9063	2.2007	1.525	0.1271

[표] 단계적 선택방법 적용한 위험률 모형

p값이 유의하지 않은 변수 2개 존재

Cox비례위험모형

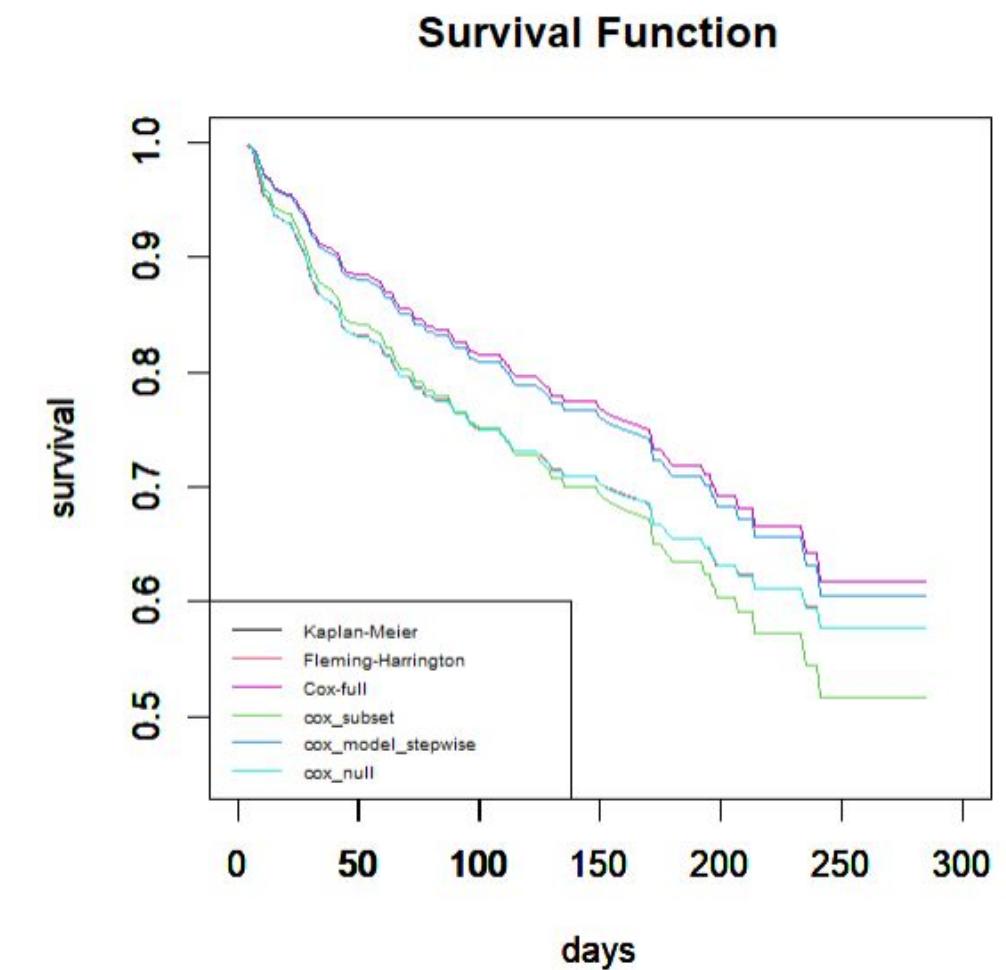


[Cox비례위험 모형(Full, p값 유의, 단계적 방법 비교)]

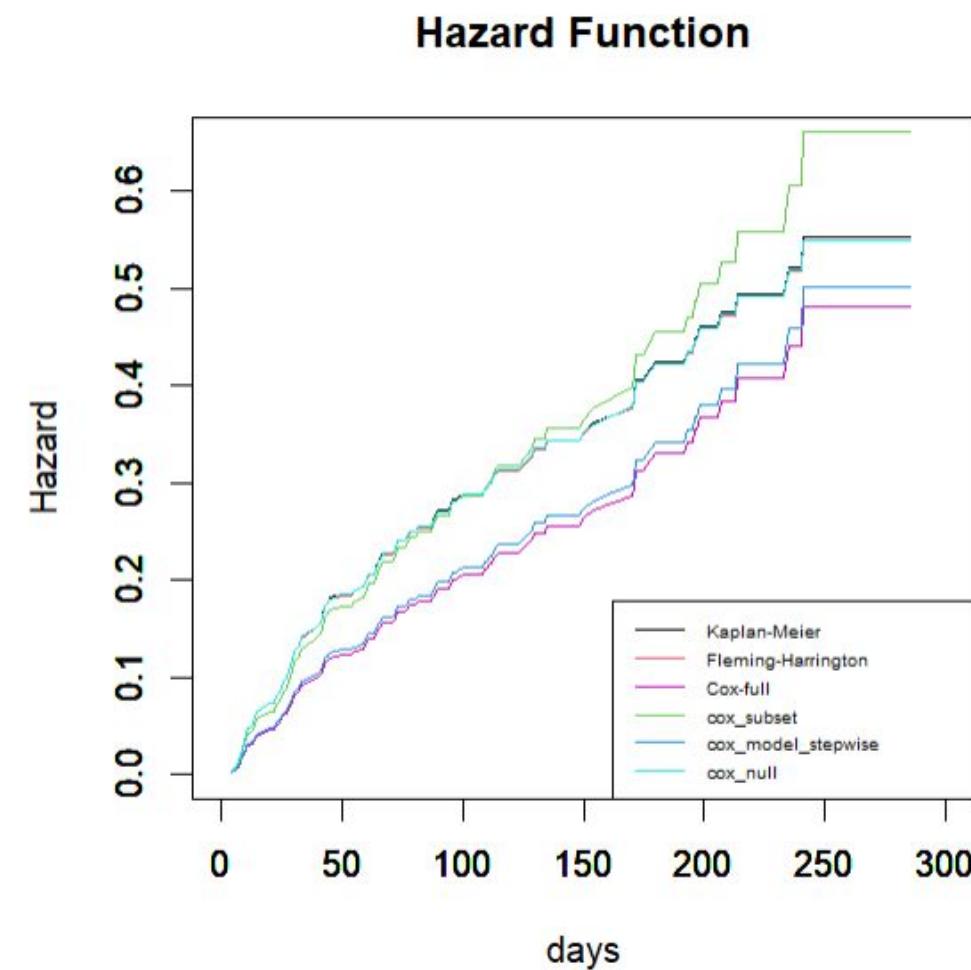


단계적 방법
축소 모형 선정

Cox비례위험 모형



[모형 생존함수]



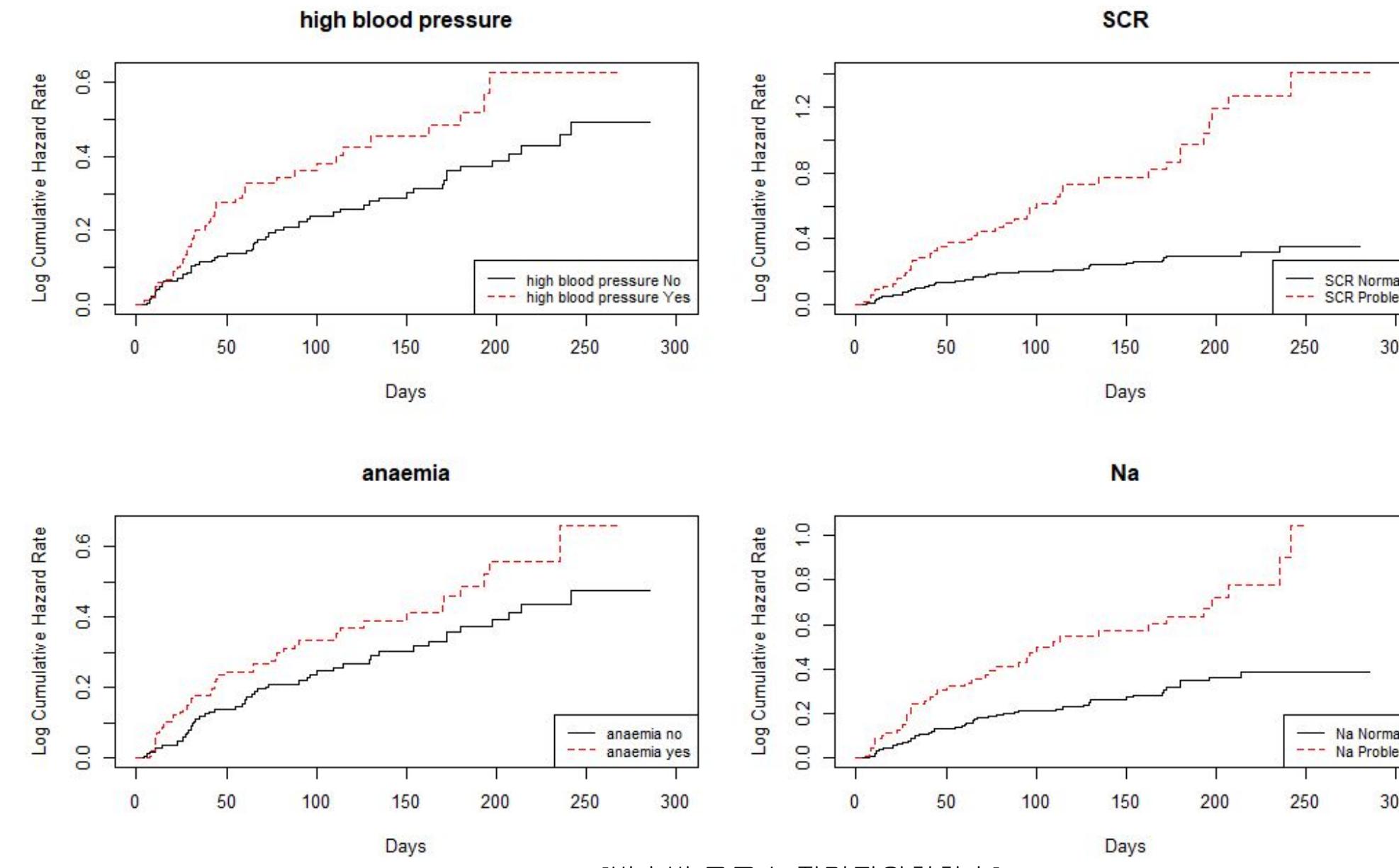
[모형 위험함수]

모형평가

적합도 검정	검정통계량	자유도	p-value
Likelihood ratio test	84.33	8	6.555325e-15
Wald test	87.1	8	1.797901e-15
Score (logrank) test	93.34	8	=<2e-16

[표] 적합도 검정 결과

모형 평가

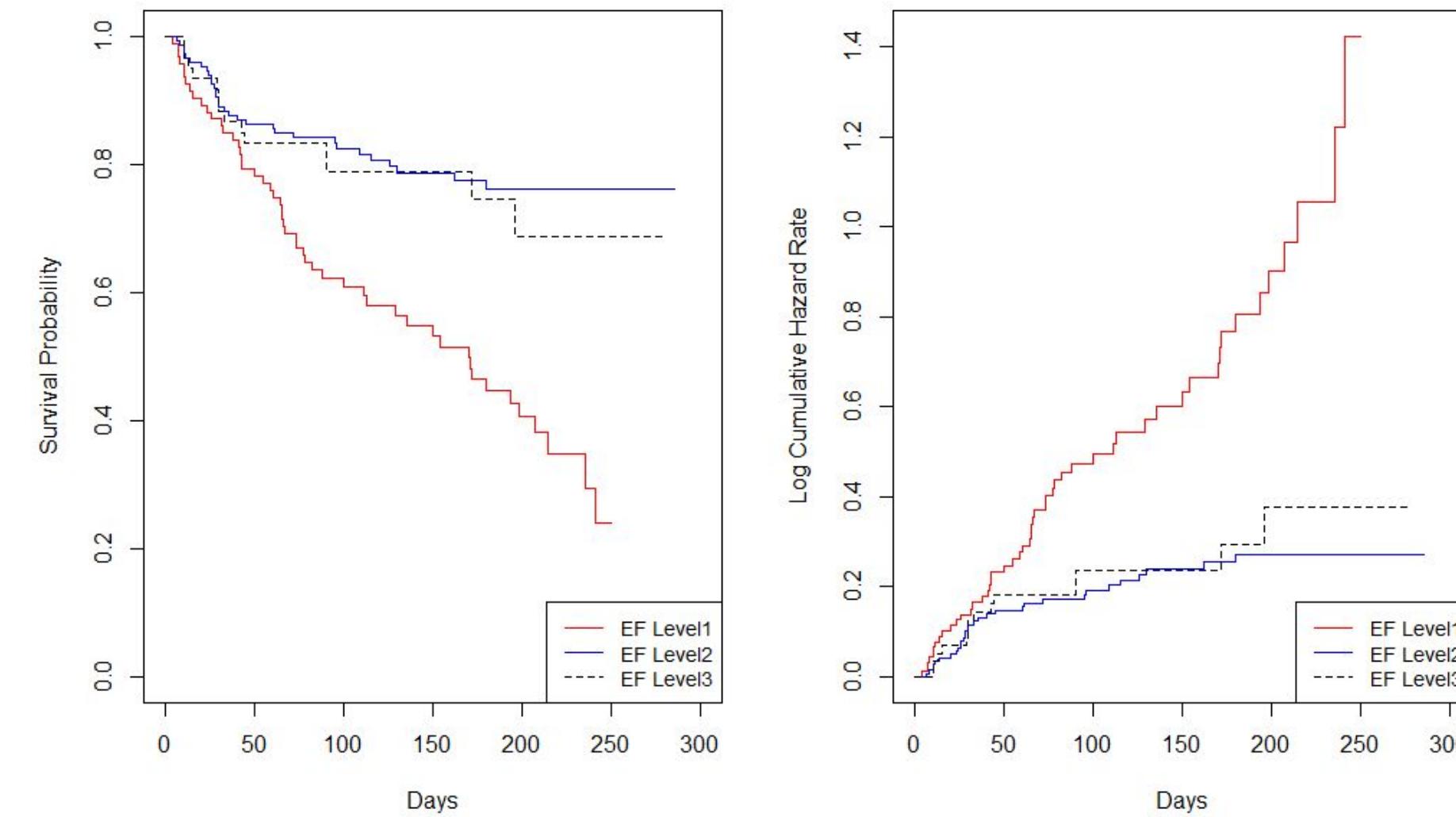


[변수별 로그 누적기저위험함수]

고혈압, 빈혈,
액청나트륨,
크레아티닌

비례성 만족

모형 평가



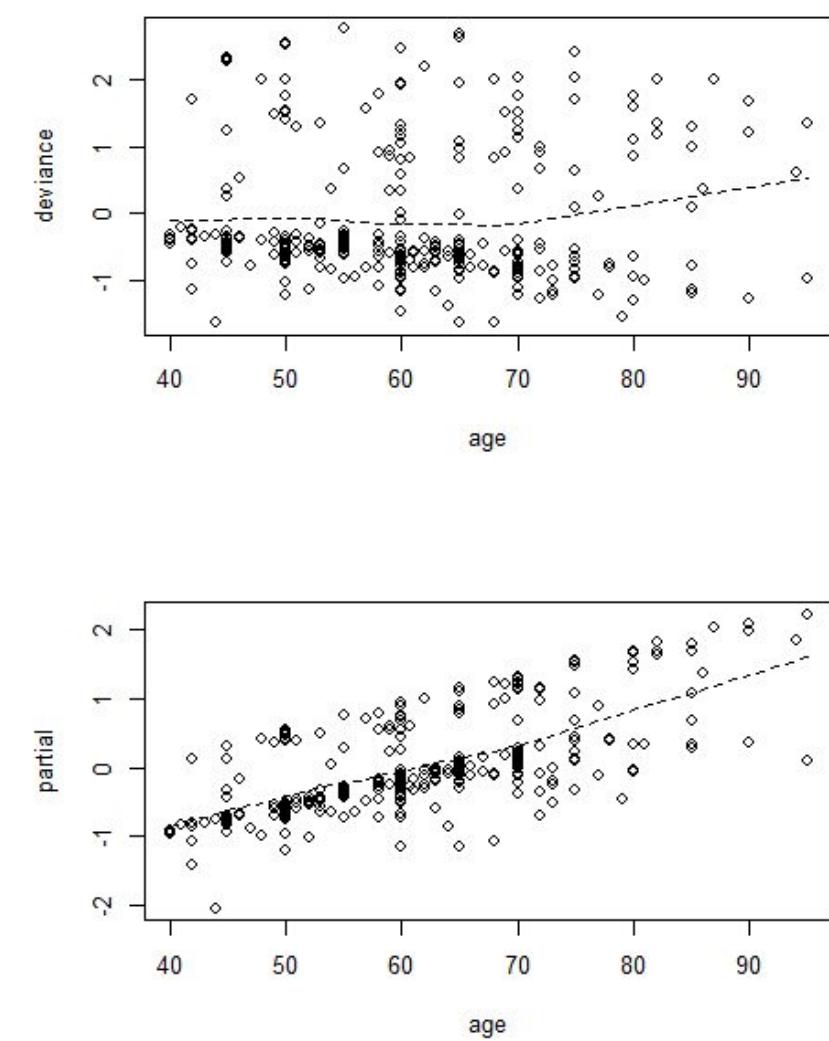
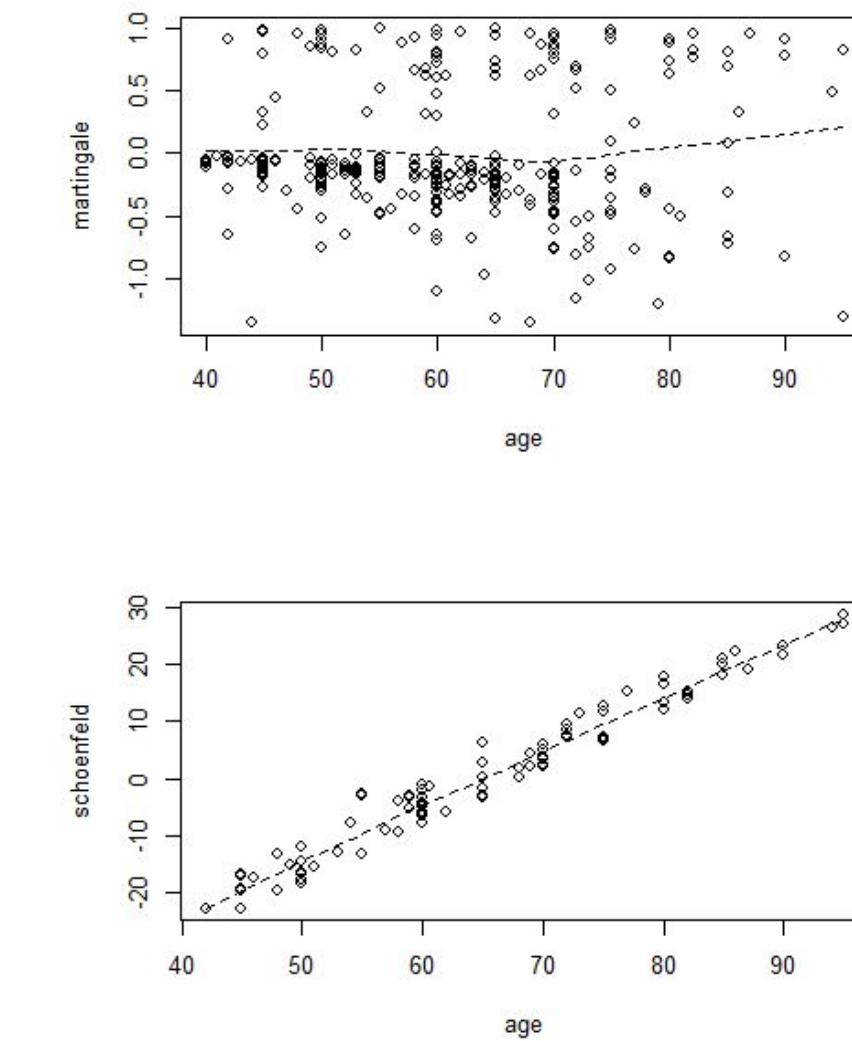
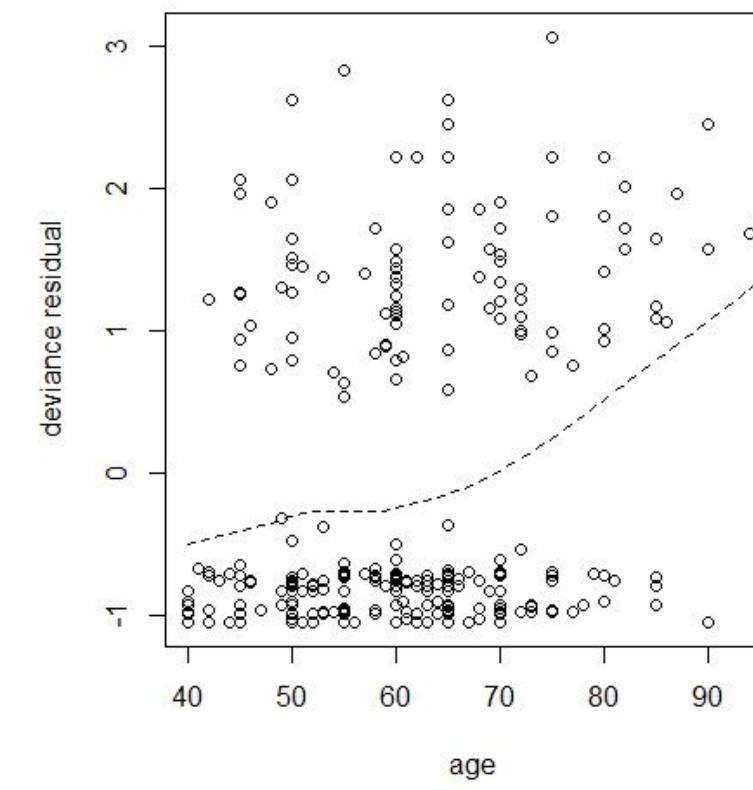
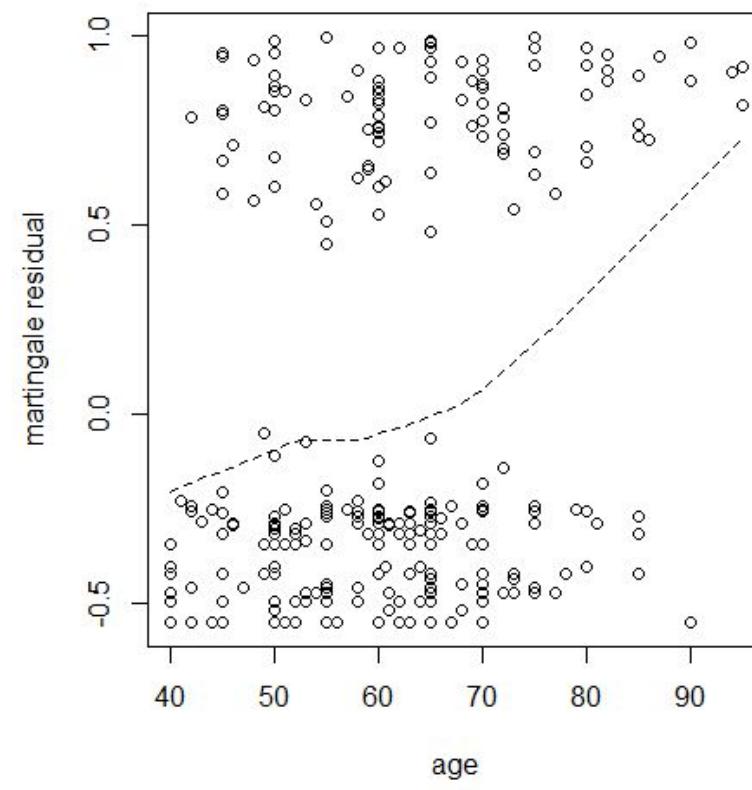
[심박출량의 생존함수 및 로그 누적위험함수]

심박출

비례성 불만족

모형평가

연속형 : 나이

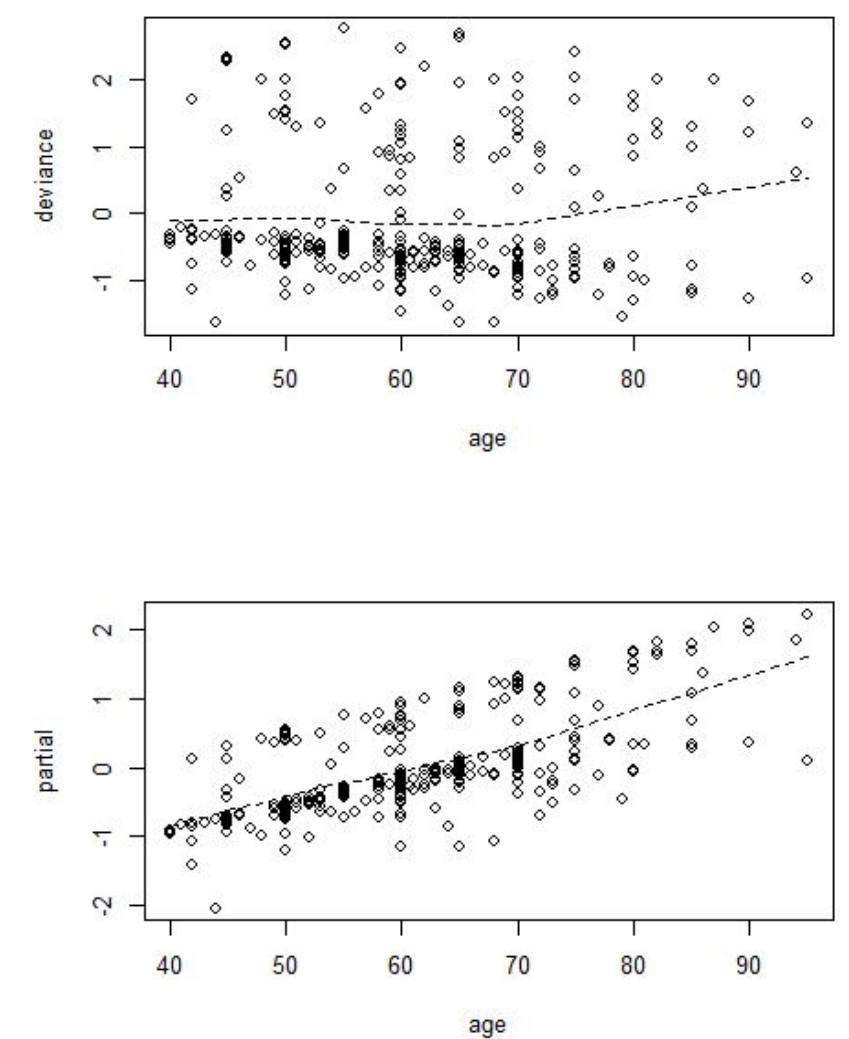
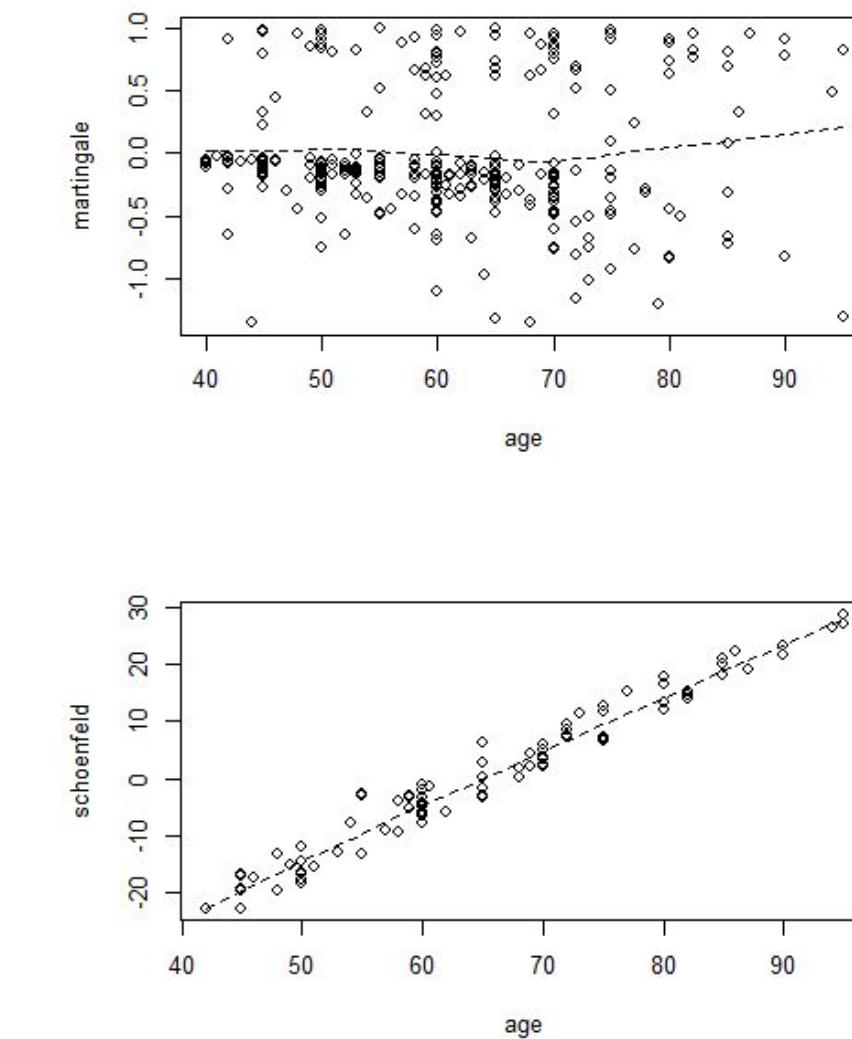
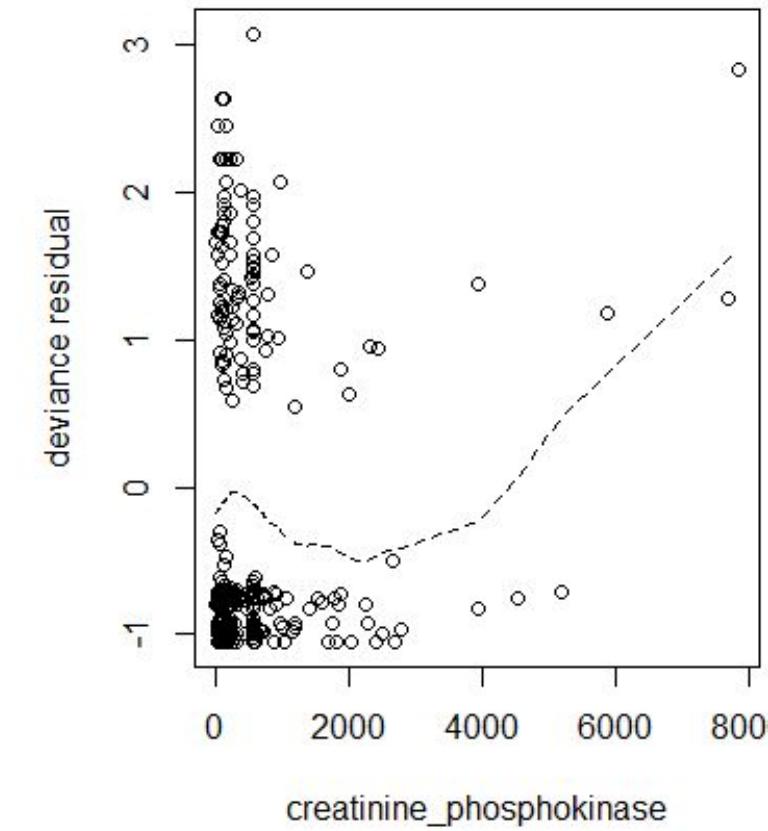
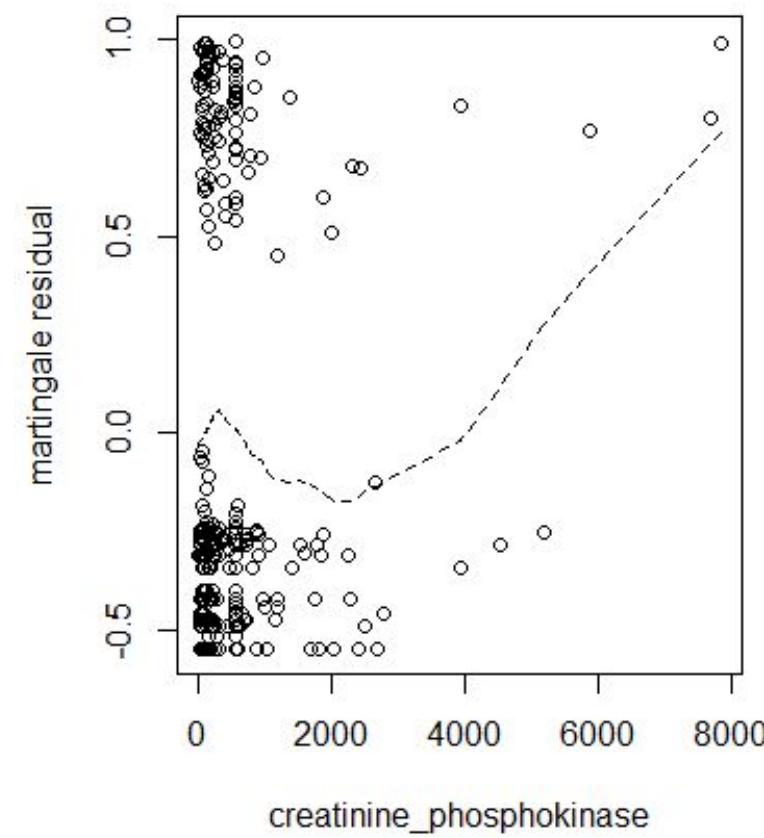


[공변량 없는 경우 age 잔차분석]

[최종 모형의 age 잔차분석]

모형평가

연속형 : CPK 효소



[공변량 없는 경우 CPK효소 잔차분석]

[최종 모형의 CPK효소 잔차분석]

모형평가

	검정통계량	p-value
나이	0.00	0.988
빈혈	0.01	0.915
고혈압	0.19	0.66
혈청나트륨	0.24	0.621
CPK효소	1.15	0.284
크레아티닌	4.35	0.037
심박출량	8.81	0.012
전체모형	13.4	0.097

심박출량 & 크레아티닌
비례성 불만족

But

전체모형은
비례성 만족하는 것으로 보임

[표] Schoenfeld 잔차를 사용한 Cox 비례위험 가정 검정

결론

생존함수

$$\begin{aligned}
 h(t | \text{나이}, \text{심박출량_Level2}, \text{크레아티닌_Problem}, \text{CPK효소}, \text{고혈압}, \text{심박출량_Level3}, \text{빈혈}, \text{혈청나트륨_Problem}) \\
 = h_0(t) \cdot \exp(0.0412 \times \text{나이} + 0.823 \times \text{크레아티닌_Problem} + 0.00026 \times \text{CPK효소} \\
 + 0.519 \times \text{고혈압} + 0.37 \times \text{빈혈} + 0.345 \times \text{혈청나트륨_Problem} \\
 - 1.01 \times \text{심박출량_Level2} - 0.772 \times \text{심박출량_Level3})
 \end{aligned}$$

비례

- 나이 ↑
- 고혈압 존재 ↑
- 빈혈 존재 ↑
- 크레아티닌 ↑ = 신장 기능 장애
- CPK효소 ↑

반비례

- 혈청나트륨 ↓ = 저나트륨혈증
- 심박출량 ↓ = 심장 기능 매우 저하

심부전으로 인한 사망 확률 증가

결론

연구 한계

- 국외 데이터 기반
- 데이터 수집 기간의 제한
- 표본의 특정성
- 환자의 중증과 추가적인 요인 미포함
- 모형 선택과 통계적 한계

추후 연구 방향

- 더 큰 규모의 국내 데이터 수집
- 장기적 추적조사
- 기계 학습 모델을 활용한 생존

분석 비교 연구 가능성

참고논문

- 이시은, 심병용, 김재희 (2015) 직장암 데이터에 대한 위험률 함수 추정 및 위험률 변화점 추정, *한국데이터정보과학회지*, 26(6), 1225-1238.
- Ahmad, T., Munir, A., Bhatti, S. H.*, Aftab, M., Raza, M. A. (2017), Survival Analysis of Heart Failure Patients: A Case Study, Department of Statistics, Government College University, Faisalabad, Pakistan.
- Chicco, D., Jurman, G. (2020), Machine Learning Can Predict Survival of Patients with Heart Failure from Serum Creatinine and Ejection Fraction Alone, *BMC Medical Informatics and Decision Making*, 20(1), 16. DOI: 10.1186/s12911-020-1023-5.

THANK YOU

감사합니다