
맞춤형 사망연금보험을 위한 생존 분석 모형

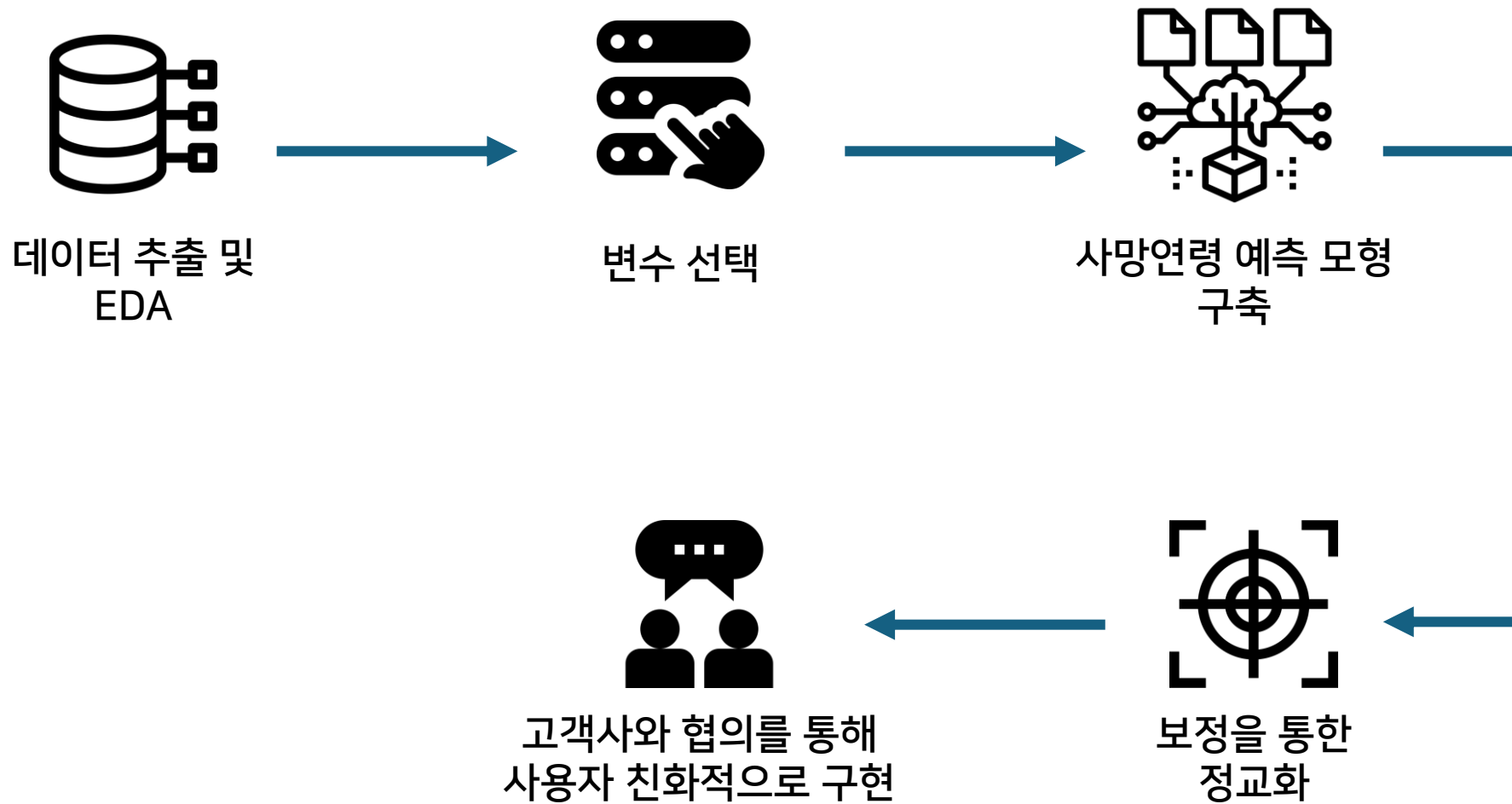
#생존분석 #R #SQL

서상우

프로젝트 필요성

- AS-IS 현재 사망연금 보험은 고객의 개별 건강 상태를 고려하지 않고, 평균적인 통계청 데이터를 활용하여 연금 수령 조건을 제공
- TO-BE 고객의 건강 상태를 기반으로 예상 사망 연령을 산출하여 고객의 개별 건강 상황에 맞는 합리적인 연금 수령 조건을 제안할 수 있음

프로세스



데이터 및 방법론 I

① 데이터 추출



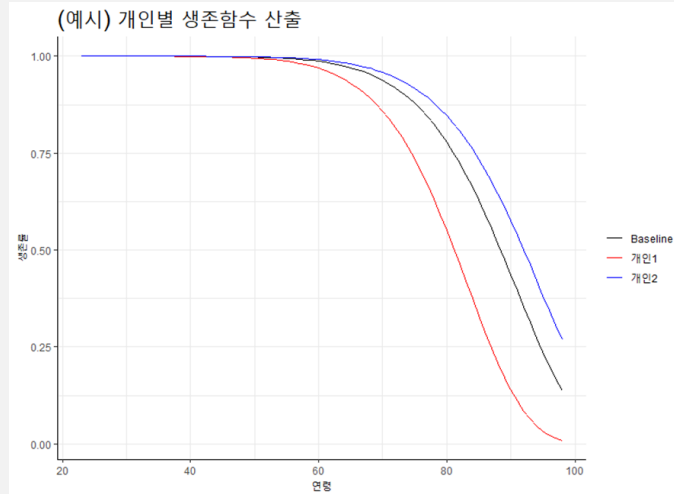
인구통계학적
정보

건강검진정보

주요 만성질환
이력

변수명	개인1	개인2
연령	46	51
성별	M	M
BMI 지수	28.1	23.2
수축기혈압	150	95
...
(주요 기질병) 당뇨 여부	1	0
(주요 기질병) 고혈압 여부	0	0
...

② 사망 연령 예측 모형 구축



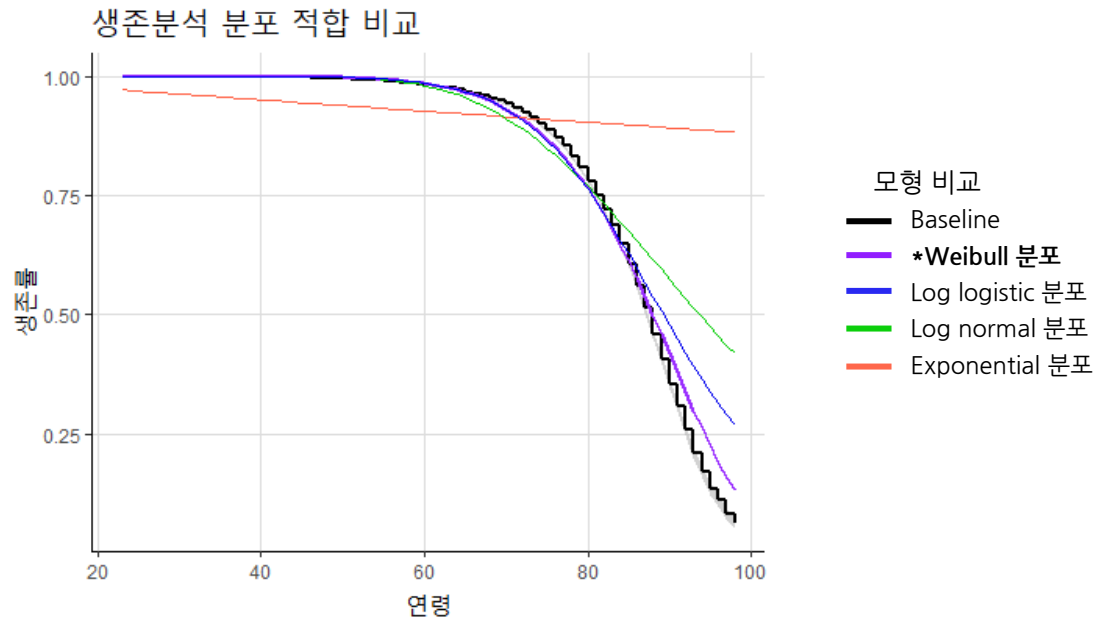
- ① 최적 모형 탐색: Weibull AFT 생존 함수
- ② 모델에 변수를 하나씩 가감해가며
최적 변수 선택
- ③ 개인별 예측 사망 연령 산출

③ 보정을 통한 정교화

: 데이터세트에서 반영하지 못하는
의료 기술의 발달, 생활 습관의 변화
등으로 인한 정보를 반영하기 위해
통계청에서 발표하는 최신 기대여명
자료를 이용한 보정 처리

데이터 및 방법론 II: 사망 연령 예측 모형

1) 최적 분포 선정



> 비교 대상 4개의 분포 중 Weibull 분포가 실제 생존 커브(Baseline 커브)와 가장 유사함

분포	AIC 스코어
Weibull 분포	201,982.4
Log logistic 분포	203,792.3
Log normal 분포	209,988.8
Exponential 분포	290,475.9

> 비교 대상 4개의 분포 중 Weibull 분포가 AIC 점수가 가장 낮아 적합도가 가장 높은 분포임을 알 수 있음

* AIC 스코어가 낮을수록 모델이 데이터를 잘 설명하면서 모형의 복잡성을 낮추었음을 의미

최종 선택 모형: **Weibull AFT** 모형

- 사망에 관한 생존 커브가 Weibull 분포를 따른다고 가정한 AFT 모형
- 참고문헌: *Using the Weibull accelerated failure time regression model to predict time to health events* (2018, Enwu Liu, Karen Lim)

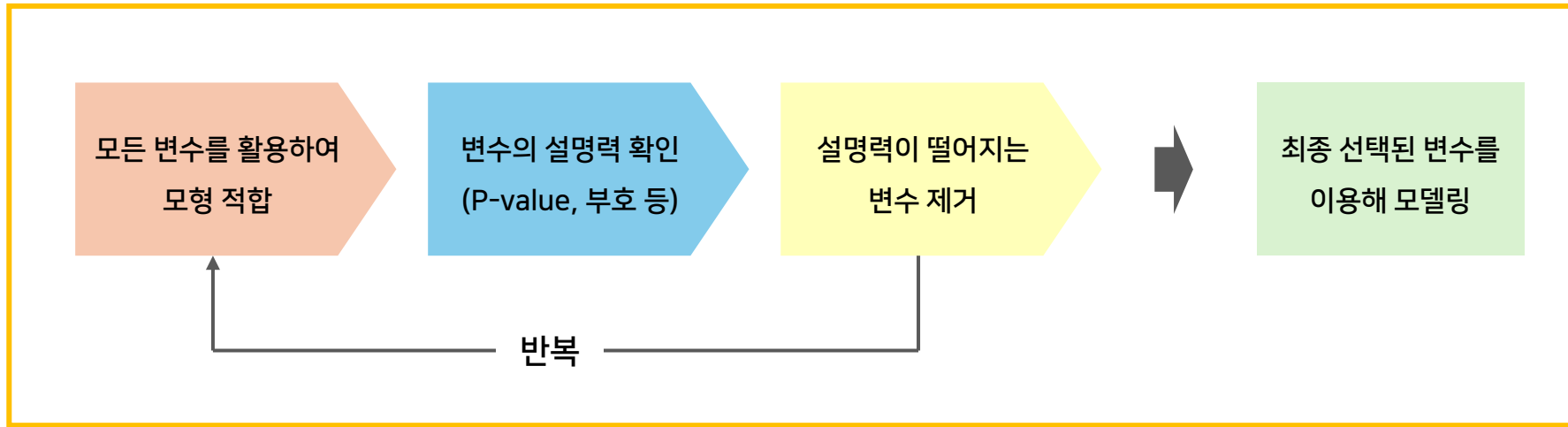
데이터 및 방법론 II: 사망 연령 예측 모형

2) 활용 가능한 변수

- 기준년도 당시 건강 검진 당시 활용 가능한 건강 검진 지표 및 의료 기록

3) 최종 변수 선택: 후진 제거법(Backward Elimination)

: 먼저 모든 독립 변수를 포함한 전체 모델을 구축한 후, 설명력이 떨어지는 변수들을 하나씩 제거한 후 남은 변수들로 다시 모델을 재적합 하는 과정을 반복하여 최종적으로 모형에 투입될 변수를 선택하는 변수 선택 프로세스



데이터 및 방법론 II: 사망 연령 예측 모형

4) 최종 선택된 변수 및 모델링 결과

(모델링 결과)

변수	계수	표준오차	z-통계량	P-value
(Intercept)	4.5075	0.0074	597.38	< 2e-16
성별 (남)	0.0779	0.0074	10.54	< 2e-16
공복 혈당 (가분)	0.0008	0.0001	12.88	< 2e-16
SGOT (국제)	0.0000	0.0004	0.00	0.99918
SGOT (가분)	0.0000	0.0001	-14.88	< 2e-16
GGT (국제)	0.0000	0.0000	0.00	0.99917
GGT (가분)	0.0000	0.0000	-25.48	< 2e-16
혈색소 (국제)	0.0000	0.0000	1.00	0.31512
혈색소 (가분)	0.0000	0.0000	-14.88	< 2e-16
요단	0.0000	0.0001	0.00	< 2e-16
Log(scale)	-2.5159	0.0047	-488.90	< 2e-10

Scale(σ) 0.0987

- Weibull AFT 모형의 생존함수

$$S_T(t) = \exp \left[- \left(\frac{t}{\exp(x'\beta)} \right)^{\frac{1}{\sigma}} \right]$$

- 평균 사망 연령 산출식

$$E(T) = \exp(x_i'\beta) \Gamma(\sigma + 1)$$

데이터 및 방법론 II: 사망 연령 예측 모형

5) AFT 생존 모형의 장점 및 활용 방안

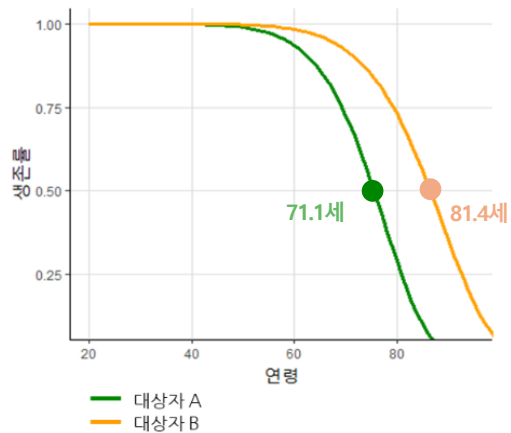
① 모형의 장점

- Cox 모형과 달리, 개별 고객의 특성에 따라 평균 또는 중앙 생존 시간 직접 수치화 가능
- 서비스에 이탈하지 않은 고객의 데이터까지 포함하여 불완전한 데이터도 효율적으로 활용 가능
- 각 고객 특성이 생존 시간에 미치는 영향을 직관적으로 해석 가능

② 활용 방안

- **고객 이탈 예측:** 일반적인 분류 모델(이탈 vs 비이탈)보다 더 정밀하게 '언제' 이탈할지를 예측 가능
→ 고객군별 이탈 위험도 시각화 및 우선 대응 순위 결정
- **고객 생애 가치(CLV) 추정:** 고객의 평균 생존 시간과 월간 평균 수익을 곱해 예상 CLV 산출 가능
- **프로모션 효과 분석:** 특정 캠페인이 고객의 서비스 유지 기간에 실제 얼마나 영향을 주었는지 통계적으로 검정 가능
→ (예시) 할인 프로모션을 받은 고객군의 생존 시각 변화 분석
- **이탈 위험군 조기 경고 시스템:** 특정 고객 특성(공변량) 조합이 이탈을 가속화할 경우 조기 경고하여 고위험 고객 선별 가능

연금 지급 전략 설계



고객의 건강상태를 반영한 기대 사망 연령 산출

기대 사망 연령	
대상자 A	71.1세
대상자 B	81.4세

- 대상자A : 회당 연금 지급액 ↑ & 연금 지급 기간 ↓
- 대상자B : 회당 연금 지급액 ↓ & 연금 지급 기간 ↑

역할 및 성과

주요 역할

- 고객사의 요청 사항 정의
: 사망연금보험 가입자의 건강 상태를 기반으로 합리적인 연금 수령 조건을 제안할 수 있는 모형 구축
- 고객사의 요청사항을 구현할 수 있는 최적 모형 리서치
- 데이터 전처리 및 분석, 모델 설계 및 최적화
- 고객사와의 협의를 통한 결과물 구현

성과

- 고객사의 요청사항을 충족시키는 모델 개발
- 프로젝트의 성공적인 완료 및 피드백 반영