2025.05.28

비모수통계학

출생아 사망에 영향을 미치는 생물학적/부모의 사회경제적 요인 분석

2020110467 김민지
2021110206 이선재

목차

01 **추진배경 및 현황** 3

02 **변수정의서** 4

03 **분석계획서** 5

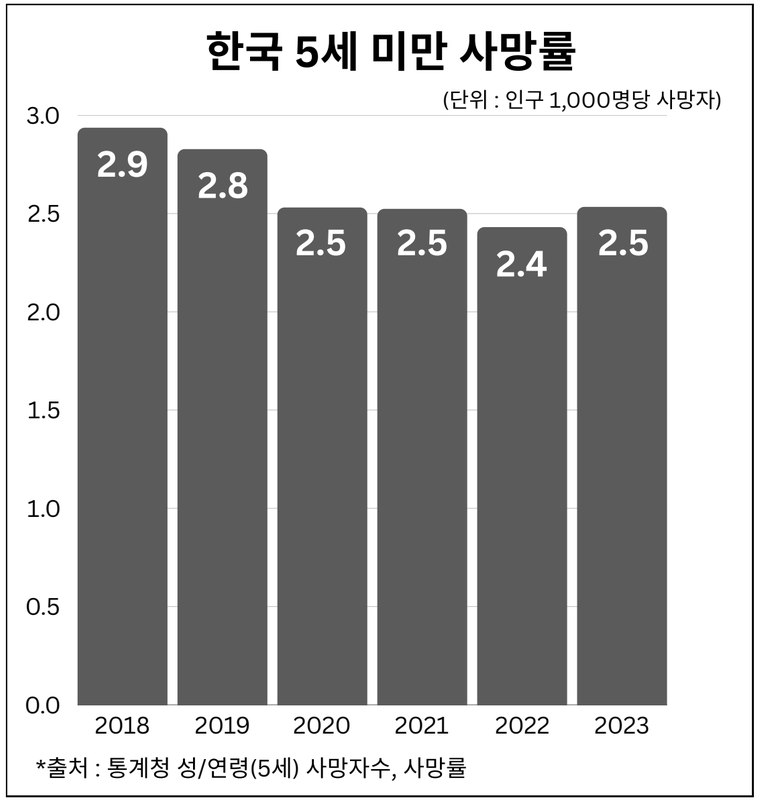
04 **데이터 분석 결과** 6

05 **결론** 10

추진배경 및 현황

5세 미만 사망률(Under-five mortality rate, U5MR)

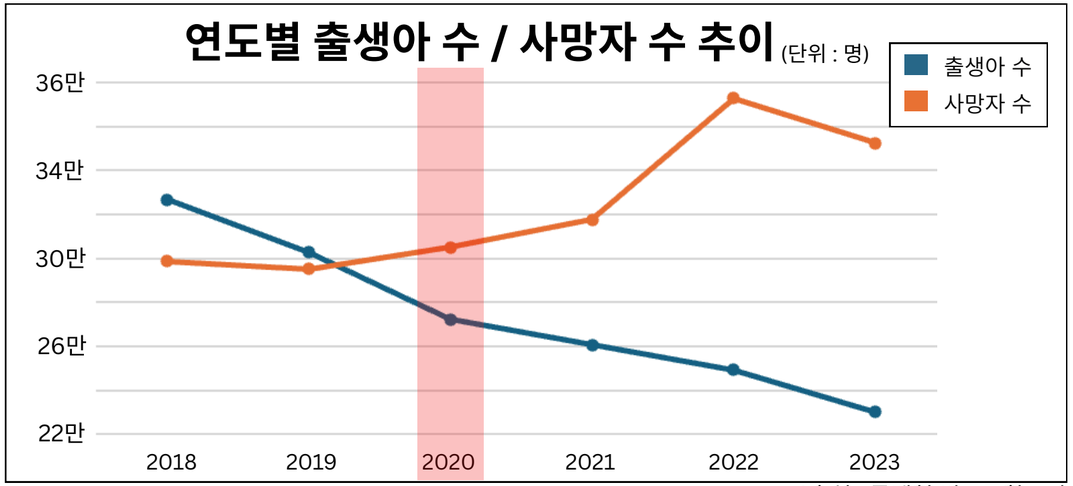
WHO, UN 등 세계 주요기관이 매년 발표하는 주요 보건지표
한국의 경우 **정체된 양상**을 보이고 있음



*출처 : 통계청 성/연령(5세) 사망자수, 사망률

데드 크로스(Dead Cross) 현상

사망자 수가 출생자 수보다 많아 **인구가 자연감소**하는 현상
한국은 **2020년**부터 시작되었음



*출처 : 통계청 인구동향조사

서울신문 | 2023-12-15

2025년 인구 '데드크로스' 가속… 50년 뒤엔 둘 중 한 명은 고령층

"2025년에 찾아올 데드크로스 현상의 원인은 **저출산**"
"앞으로 50년간 한국인구 중 절반은 **63세 이상의 역삼각형 초고령화 형태**"

저출산 시대 건강한 미래인구 확보를 위해선 **5세미만 영유아의 사망에 영향을 미치는 요인을 파악**하는 것이 중요함
따라서 사전에 예방할 수 있도록 사망 관련 요인을 파악하는 것이 필요함

변수정의서

- 통계청 마이크로데이터 5세 미만 영유아 출생-사망 연계자료(2018) - Row : 326822, Columns : 44
- 2018년에 출생 후 5년 간 사망여부를 추적한 데이터

파생변수

변수명	타입	설명
생존여부	범주형	0 : 생존 / 1: 사망(직접적 연계) / 2 : 사망(통계적연계)
모_연령	연속형	산모의 나이
모_직업코드	범주형	산모의 직업을 나타내는 코드
모_교육정도코드	범주형	산모의 최종 학력 수준
임신주수	연속형	임신 유지 기간(주)
다태아코드	범주형	단태아/다태아 코드
출생아체중량	연속형	출생 시 체중(kg)
…	…	…
생존여부2	범주형	생존여부 데이터를 재범주화 0 : 생존 / 1: 사망
산모연령대	범주형	35세를 기준으로 모_연령 데이터를 범주화
학력	범주형	고등학교를 기준으로 모_교육정도코드 데이터를 범주화 학력낮음 : 무학~고등학교 / 학력높음 : 대학교 이상
출생아사망그룹	범주형	사망 시점에 따라 사망한 출생아를 범주화 (신생아사망/후기신생아사망/1세이상사망)

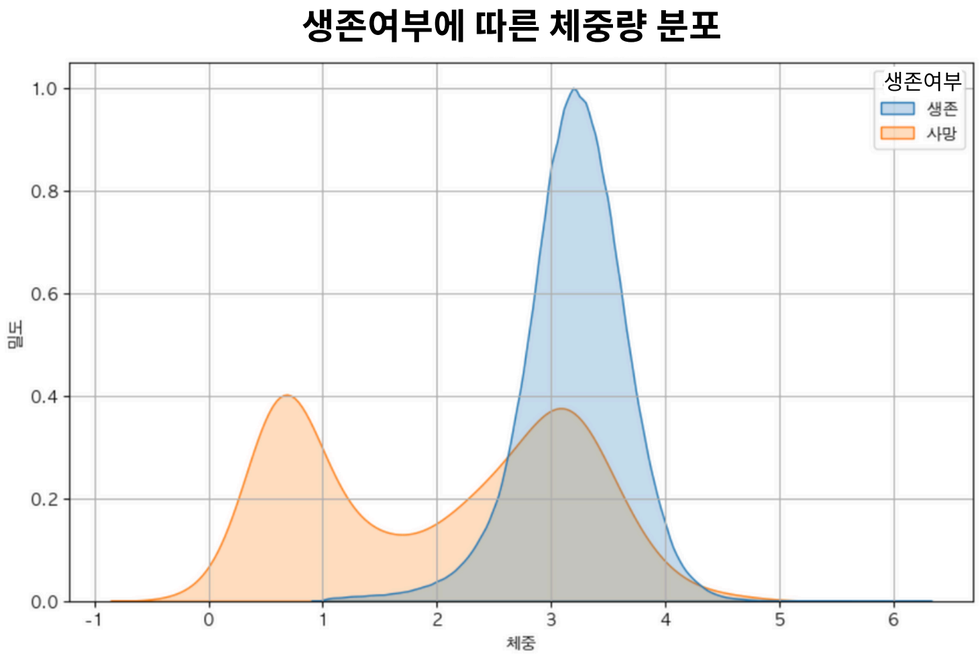
분석계획서

목적	분석방법	주요내용
생존/사망 그룹 간 체중 분포 비교, 출생아 사망 그룹 간 임신주수에 차이가 있는지 분석	Kerenel Density Estimation Krsukal-Wallis test	• 생존 여부에 따른 출생아 체중 분포 시각화 • 출생아 사망그룹별(신생아 사망, 후기신생아 사망, 영아 사망) 임신주기 중위수 검정
출생아 생존여부에 따라 체중량에 유의미한 차이가 있는지 분석	Wilcoxon Ranksum Test Permutation Test Binomial Test	• 출생아 생존여부에 따른 체중의 평균과 중위수 검정
산모의 연령과 사회경제적 특성이 출생아 생존여부와 관련이 있는지 파악	Fisher’s exact Test	• 산모 연령대(35세 미만, 35세 이상) , 부모 학력(낮음, 높음), 혼인여부(기혼, 미혼)와 출생아 생존여부간의 연관성 분석
출생아와 부모의 정보를 통해 출생아 생존여부를 예측	Decision Tree Logistic Spline	• 출생아 체중, 산모 연령, 부모 직업, 다태아출산순위코드 등을 입력하면 출생아의 생존여부를 예측하는 모델 개발

데이터 분석 결과 | 사망요인 분석

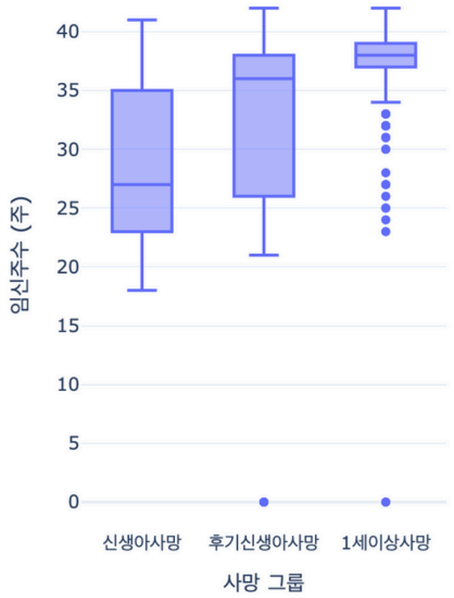
생존여부에 따른 체중량 분포

KDE(Kernel Density Estimation)을 활용해 생존그룹/사망그룹별 체중량 분포를 확인



사망그룹의 체중량 분포는 **0~1kg대, 3kg대**가 많은 것을 알 수 있음
생존그룹 역시 **3kg대**가 많기 때문에, **체중과 체중이 아닌 다른 요인**이
5세 미만 영유아 사망에 영향을 준다고 볼 수 있음

사망 그룹에 따른 임신주수



	생존일수
신생아사망	0-27일
후기신생아사망	28-364일
1세이상사망	365일 이상

Kruskal-Wallis Test	
귀무가설	세 그룹 간 임신주수 중위수에는 차이가 없다.
대립가설	적어도 한 그룹의 임신주수 중위수가 다른 그룹과 유의하게 다르다.
P.value(=0.0001) < 0.05 (귀무가설 기각)	
세 그룹의 출생 당시 임신주수의 분포는 **유의하게 다르다.**	

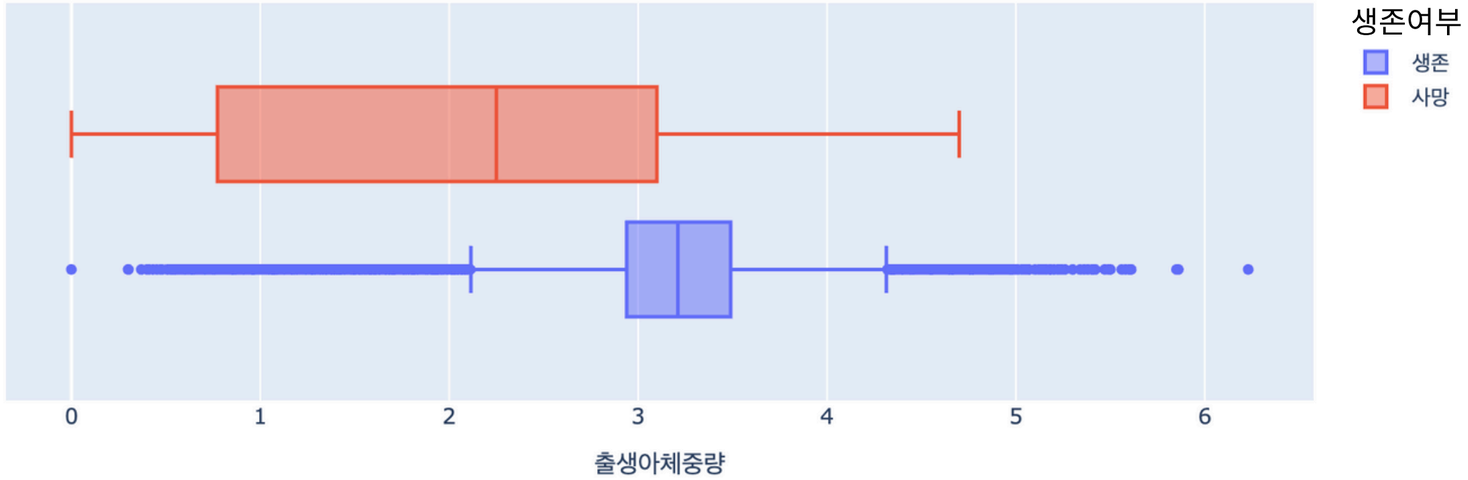
*조산 : 20-37주 출산

신생아사망 : 임신주수 중위수는 27주로 **조산*에 의한 사망 가능성**이 큼
1세이상사망 : 거의 만삭에 태어났기 때문에, **출생 이후 요인에 의한 사망 가능성**이 큼

5세 미만 영유아 사망은 **체중과 체중 외의 요인 2가지**로 인해 발생하기 때문에, 그 외의 요인을 파악해보고자 함

데이터 분석 결과 | 생존여부에 따른 체중의 대표값 검정

생존여부에 따른 체중량 분포



생존여부에 따른 체중의 대표값 검정

체중량에 대한 정규성 검정 결과, 귀무가설(체중량은 정규분포를 따른다.)을 기각하여 **비모수적으로 접근**하였음

Wilcoxon Ranksum Test	
귀무가설	두 그룹(생존, 사망)간 체중량의 대표값(중앙값)에 차이가 없다.
대립가설	두 그룹(생존, 사망)간 체중량의 대표값(중앙값)에 차이가 있다.
P.value(=2.590e-241) < 0.05 (귀무가설 기각)	
두 그룹(생존, 사망)간 체중량의 대표값(중앙값)에 **차이가 있다.**	

Permutation Test	
귀무가설	생존그룹의 평균 체중량과 사망그룹의 평균 체중량은 같다.
대립가설	생존그룹의 평균 체중량이 사망그룹의 평균 체중량보다 높다.
P.value(=0.001) < 0.05 (귀무가설 기각)	
생존그룹의 평균 체중량이 사망그룹의 평균 체중량보다 **높다.**	

Binomial Test*	
귀무가설	생존그룹의 체중량 중위수는 3.5kg이다.
대립가설	생존그룹의 체중량 중위수는 3.5kg보다 작다.
P.value(=0.0) < 0.05 (귀무가설 기각)	
생존그룹의 체중량 중위수는 3.5kg보다 **작다.(95% 신뢰구간 : [3.21, 3.217])**	

*사망그룹 역시 귀무가설 기각하였음(귀무가설 : 사망그룹의 체중량 중위수는 2.5kg이다/대립가설 : 사망그룹의 체중량 중위수는 2.5kg보다 작다.)
*사망그룹 체중량 95% 중위수 신뢰구간 : [2.09., 2.38]

생존여부에 따라 체중의 대표값에 차이가 있고, 생존그룹에 비해 사망그룹의 체중량이 낮아 **체중량이 사망의 요인**이라고 볼 수 있음

데이터 분석 결과 | 체중 외 사망요인 파악

산모 연령대(35세 미만 vs 35세 이상)

산모 연령대 - 생존여부 교차표 (단위 : 명)

	사망	생존
35세 이상*	415	103698
35세 미만	706	222003

*한국 의료계에서는 35세 이상을 고령 출산이라고 규정

Fisher's exact Test (산모 연령대 - 생존여부)	
귀무가설	산모 연령대(35세 미만, 35세 이상)와 출생아 생존여부 간에 관련이 없다.
대립가설	연령대가 높을수록 출생아 사망률이 높아진다.(양의 상관관계가 있다.)
P.value(=0.0001) < 0.05 (귀무가설 기각)	
연령대가 높을수록 출생아 사망률이 높아진다.**(양의 상관관계가 있다.)**	

산모 학력(학력낮음 vs 학력높음)

산모 학력 - 생존여부 교차표 (단위 : 명)

	사망	생존
학력낮음	349	66071
학력높음	631	254439

*산모의 학력이 미상인 경우 제외하였음

Fisher's exact Test (산모 학력 - 생존여부)	
귀무가설	산모 학력(낮음, 높음)과 출생아 생존여부 간에 관련이 없다.
대립가설	학력이 낮을수록 출생아 사망률이 높아진다.(양의 상관관계가 있다.)
P.value(=2.028e-27) < 0.05 (귀무가설 기각)	
학력이 낮을수록 출생아 사망률이 높아진다.**(양의 상관관계가 있다.)**	

산모 혼인여부(기혼 vs 미혼)

산모 혼인여부 - 생존여부 교차표 (단위 : 명)

	사망	생존
기혼	988	318459
미혼	73	7091

Fisher's exact Test (혼인여부 - 생존여부)	
귀무가설	혼인여부(기혼, 미혼)와 출생아 생존여부 간에 관련이 없다.
대립가설	기혼인 경우 출생아 사망률이 낮아진다.(음의 상관관계가 있다.)
P.value(=3.87e-17) < 0.05 (귀무가설 기각)	
기혼인 경우 출생아 사망률이 낮아진다.**(음의 상관관계가 있다.)**	

체중량 외에도 **산모 연령대, 학력, 혼인여부** 모두 출생아 사망과 연관이 있음
따라서, **산모의 사회경제적 배경**을 고려한 맞춤형 지원이 중요함

데이터 분석 결과 | 출생아 생존여부 예측

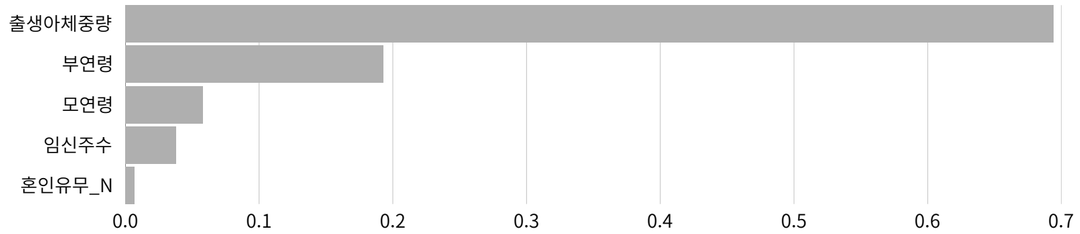
Tree 모델 성능 요약

Recall (실제 사망자 중 사망으로 예측한 비율) 중심의 모델 평가 진행

	Recall	Accuracy
Tree	0.66	0.8955

Tree 모델 변수 중요도

임신주수, 출생아체중량, 모_연령, 부_연령모_직업코드, 부_직업코드, 모_교육정도코드, 부_교육정도코드, 출생자성별코드, 다태아분류코드, 출생장소코드, 혼인중또는혼인외자녀코드



주요 변수인 출생아체중량 기준으로 성능 구간 분석

	<= 2.5kg	> 2.5kg
Recall	0.87	0.34

2.5kg 초과 시 Recall이 급감하여
다른 요인의 영향 가능성 시사

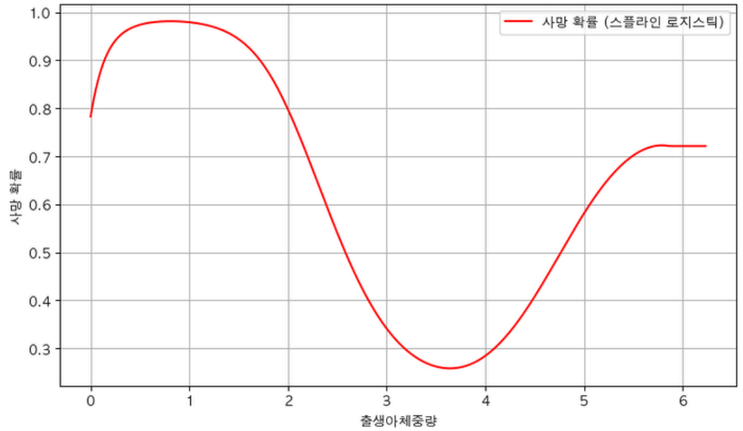
Spline 모델 성능 요약

변수 중요도 결과, 출생아체중량이 주요 요인으로 나타나 **단일 변수 기반 예측** 실험

	Recall	Accuracy
출생아체중량	0.56	0.9130
연속형변수	0.62	0.9329

*연속형 변수 : 임신주수, 출생아체중량, 모_연령, 부_연령

출생아체중량에 따른 사망 확률



저체중일수록 사망 확률이 급증하며, **출생아체중량만으로도 경향성 포착 가능**

Tree 모델과 Spline모델을 통해 출생아 생존여부를 예측하여, **5세 미만 사망 고위험군**을 사전에 파악할 수 있음

결론

프로젝트 목적

출생아 사망에 영향을 미치는 생물학적/부모의 사회경제적 요인 분석

5세 미만 영유아 사망은 **체중과 체중 외의 요인 2가지**로 인해 발생

생존여부에 따라 체중의 대표값에 차이가 있고, 생존그룹에 비해 사망그룹의 체중량이 낮아 **체중량이 사망의 요인**이라고 볼 수 있음

산모 연령대, 학력, 혼인여부 모두 출생아 사망과 연관이 있음

출생아 생존여부 예측을 통해 **사망 고위험군 사전 파악 가능**

저체중 출생 예방뿐만 아니라 산모의 연령, 학력, 혼인 상태 등 다양한 요인에 대한 **다각적인 개입이 병행**되어야 함

감사합니다

2020110467 김민지
2021110206 이선재