

**출생아 사망에 영향을 미치는
생물학적/부모의 사회경제적 요인 분석**

2020110467 김민지

2021110206 이선재

목차

1. 추진배경 및 현황	1
2. 사용한 데이터	2
3. 분석 계획서	3
4. 데이터 분석 결과	4
4.1 사망요인 분석	4
4.2 생존여부에 따른 체중의 대표값 검정	5
4.3 체중 외 사망요인 파악	7
4.4 출생아 생존여부 예측	8
5. 결론	10
6. 코드 및 참고문헌	11

1. 추진배경 및 현황

5세 미만 사망은 출생아가 만 5세가 되기 전인 영유아 시기에 사망하는 것을 의미하며, 이는 나라 간 보건 및 사회 발전 상태를 비교하기 위한 지표로 사용된다.¹ 세계의 각 국가들은 5세 미만 아동 사망률(Under-five mortality rate, U5MR)을 개선시키기 위한 다양한 노력을 시도하고, 요구 받는다. 한국에서도 국가 차원의 영유아 보건 및 복지정책을 수행하고 있지만, 5세 미만 사망률은 정체된 양상을 보이고 있다(그림 1 참고).

한편, 영유아 시기의 건강은 성장한 후인 성인과 밀접한 관련이 있다. 저체중아는 출생 후 성장장애가 발생할 수 있으며, 정상체중아와 비교했을 때 아동기 장애가 정상아에 비해 증가하는 것으로 나타났다. 따라서 영유아 시기 사망요인을 선제적으로 파악하는 것은 보건학적으로 중요한 문제이다. 또한, 한국은 2020년부터 사망자 수가 출생자 수보다 많아 인구가 자연감소하는 데드 크로스(Dead Cross)현상이 나타나고 있다(그림 2 참고). 처음 인구가 감소했던 2020년의 경우에는 코로나19의 영향이 컸던 반면, 최근 들어 발생하고 있는 데드크로스의 원인은 저출산 때문이다. 통계청의 '장래인구추계 : 2022~2072년'에 따르면 앞으로 50년간 한국 인구는 1,550만 명 가량 급감해 1977년 수준인 3,600만 명대까지 쪼그라들고 그중 절반은 63세 이상으로 채워지는 역삼각형 형태의 초고령 화사회가 될 것으로 예측했다. 저출산 시대 건강한 미래인구를 확보하기 위해 5세 미만 영유아의 사망에 영향을 미치는 요인을 파악하는 것이 중요하다.

¹ 5세 미만 사망률(Under-five mortality rate, U5MR)은 WHO, UN 등 세계 주요 기관이 매년 발표하는 주요 보건 지표이다. UN이 정한 지속가능발전목표에는 2030년까지 5세 미만 사망률을 1,000명 당 25명 이하로 낮추는 것이 명시되어 있다.

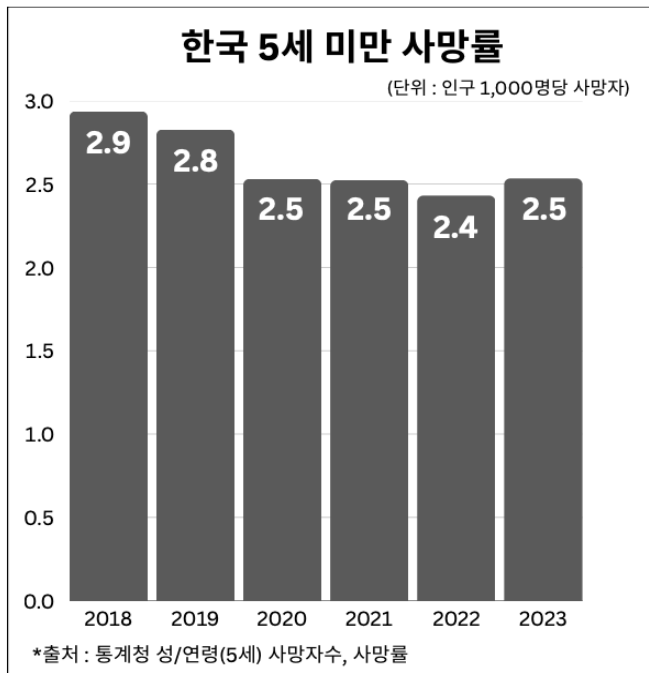


그림 1 한국 5세 미만 사망률(2018~2023)

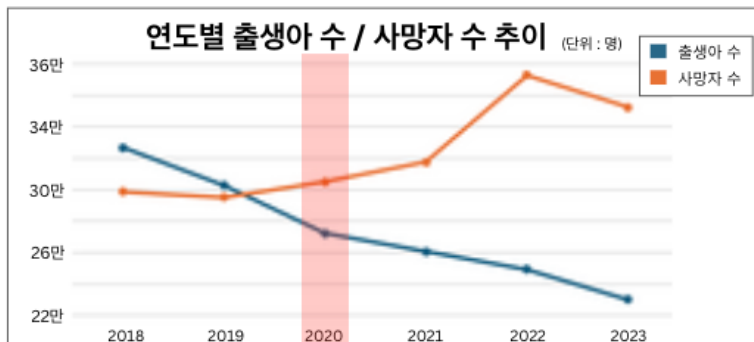


그림 2 연도별 출생아 수 및 사망자 수 추이(2018~2023)

2. 사용한 데이터

본 연구에서 사용한 데이터는 통계청 마이크로데이터 통합서비스(MDIS) '5세 미만 영유아 출생-사망 연계'이다. 이는 2018년에 출생 후 5년 간 사망여부를 추적한 데이터로 약 32만 개의 행과 44개의 변수를 가지고 있다. 분석에 사용된 주요 변수로는 생존여부, 산모 연령, 산모 교육정도코드, 임신주수, 다태아코드, 출생아체중량 등이 있다.

표 1 변수 정의서

변수명	타입	설명
생존여부	범주형	0 : 생존 / 1 : 사망(직접적연계) / 2 : 사망(통계적 연계)
혼인중또는혼인외자녀코드	범주형	산모의 혼인여부
모_연령	연속형	산모의 나이
모_교육정도코드	범주형	산모의 최종학력 수준
임신주수	연속형	임신 유지 기간(주)
다태아분류코드	범주형	단태아 : 태아가 1명인 경우 / 쌍태아 : 2명 / 3명 이상
출생아체중량	연속형	출생 시 체중(kg)
생존여부2	범주형	생존여부 데이터를 재범주화 (0 : 생존, 1 : 사망)
산모연령대	범주형	35세를 기준으로 모_연령 데이터를 재범주화 (35세 미만, 35세 이상)
학력	범주형	고등학교를 기준으로 모_교육정도코드를 재범주화 (학력낮음 : 무학~고등학교, 학력높음 : 대학교 이상)
출생아사망그룹	범주형	사망시점에 따라 사망한 출생아를 재범주화 (신생아사망, 후기신생아사망, 1세이상사망)

3. 분석 계획서

5세 미만 영유아 사망에 미치는 요인을 파악하기 위해 다음과 같이 분석 계획서를 작성하였다(표 2 참고).

표 2 분석 계획서

목적	분석 방법	주요 내용
생존/사망 그룹 간 체중 분포 비교, 출생아 사망 그룹 간 임신주수에 차이가 있는지 분석	KDE Kruskal Wallis Test	생존여부에 따른 출생아 체중 분포 시각화 출생아 사망그룹별(신생아 사망, 후기신생아 사망, 1세 이상 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
부모의 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)과 출생아 생존여부 간의 연관성 검정
출생아와 부모의 정보를 통해 출생아 생존여부를 예측	Tree model Logistic Spline	출생아 체중, 산모 연령, 부모 직업, 다태아출산순위코드 등을 입력하면 출생아의 생존여부를 예측하는 모델 개발

4. 데이터 분석 결과

4.1 사망요인 분석

Kernel Density Estimation(KDE)을 활용해 생존그룹/사망그룹별 체중량 분포를 확인한 결과, 사망그룹의 체중량 분포는 봉우리가 2개인 형태를 보였다(그림 3 참고). 첫 번째 봉우리는 0~1kg 대인 저체중 그룹이라고 해석할 수 있다. 두 번째 봉우리인 3kg대는 생존 그룹에서도 가장 많은 분포를 차지하고 있기 때문에, 체중이 아닌 다른 요인이 사망에 영향을 끼친다고 해석할 수 있다.

한편, 생존일수에 따라 사망한 영유아를 세 그룹으로 범주화하였다. 생존일수가 0~27일인 경우 '신생아사망', 28~364일인 경우 '후기신생아사망', 365일 이상인 경우 '1세이상사망'으로 구분했다. 사망그룹에 따른 임신주수에 차이가 있는지 검정하기 위해 먼저 Box

Plot으로 시각화하였다(그림 4 참고). 집단이 세 그룹 이상일 때 중위수 검정을 수행하는 Kruskal-Wallis Test을 진행하였다. 이때 귀무가설은 '세 그룹 간 임신주수 중위수에는 차이가 없다'이고, 대립가설은 '적어도 한 그룹의 임신주수 중위수가 다른 그룹과 유의하게 다르다'이다. 가설검정 결과, p-value값은 0.0001로 0.05보다 작기 때문에 귀무가설을 기각하였다. 따라서 세 그룹의 출생 당시 임신주수의 분포는 유의하게 다르다고 볼 수 있다. 특히, 신생아사망의 경우 임신주수 중위수는 27주로 조산에 의한 사망 가능성이 큰 반면, 1세이상사망의 경우 만삭에 태어난 경우가 많기 때문에 출생 이후의 다른 요인에 의한 사망 가능성이 크다고 볼 수 있다.

사망요인 분석을 종합해보면, 5세 미만 영유아 사망은 체중과 체중 외의 요인 2가지로 인해 발생한다고 볼 수 있다.

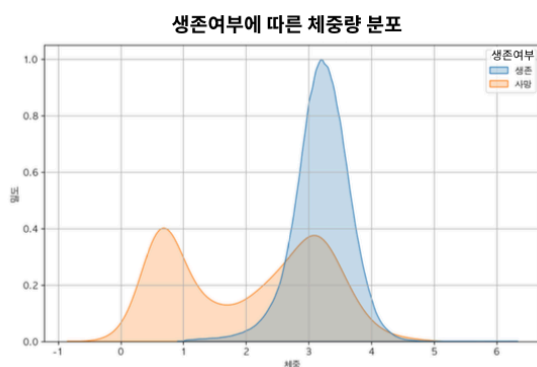


그림 3 생존여부에 따른 체중량 분포

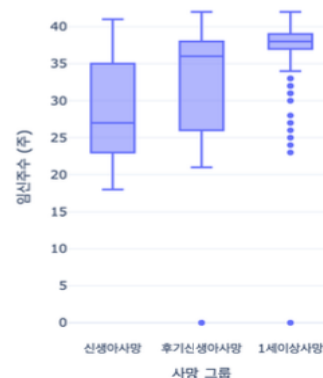


그림 4 사망 그룹에 따른 임신주수 분포

4.2 생존여부에 따른 체중의 대표값 검정

체중과 관련된 검정을 통해 생존여부에 따라 대표값에 차이가 있는지 확인해보고자 한다. 체중량에 대한 정규성 검정 결과, 귀무가설(체중량은 정규분포를 따른다.)을 기각하여 비

모수적으로 접근하였다.

먼저 Wilcoxon Ranksum Test을 통해 두 그룹 간 체중량의 대표값에 차이가 있는지 알아보았다. 귀무가설은 '두 그룹(생존, 사망)간 체중량의 중앙값에 차이가 없다.'이고, 대립가설은 '두 그룹(생존, 사망)간 체중량의 중앙값에 차이가 있다.'로 설정했다. 그 결과, p-value가 $2.590e-241$ 로 0.05보다 작았기 때문에 귀무가설을 기각했다. 따라서 두 그룹 간 체중량의 중앙값에는 통계적으로 유의미한 차이가 있다고 볼 수 있다.

다음으로는 Permutation Test을 통해 평균 체중량에 대한 검정을 실시했다. 이 때 귀무가설은 '생존그룹의 평균 체중량과 사망그룹의 평균 체중량은 같다.', 대립가설은 '생존그룹의 평균 체중량이 사망그룹의 평균 체중량보다 높다.'로 단측검정(우측)을 진행했다. 그 결과 p-value가 0.001로 0.05보다 작았기 때문에 귀무가설을 기각했다. 따라서 생존그룹의 평균 체중량이 사망그룹의 평균 체중량보다 높다고 볼 수 있다.

이외에도 체중량 중위수가 특정 값보다 작은지 검정하는 Binomial Test을 진행했다. 먼저 생존그룹의 경우, 귀무가설은 '생존그룹의 체중량 중위수는 3.5kg이다.'. 대립가설은 '생존그룹의 체중량 중위수는 3.5kg보다 작다'로 설정했다. 그 결과 p-value 값이 0으로 0.05보다 작아 귀무가설을 기각했다. 또한 중위수에 대한 95% 신뢰구간은 [3.21kg, 3.217kg]인 것으로 나타났다. 따라서 생존그룹의 체중량 중위수는 3.5kg보다 작다는 주장이 통계적으로 유의미하다고 볼 수 있다. 사망그룹에 대해서도 Binomial Test을 진행했다. 그 결과, 귀무가설을 기각하였으며 사망그룹의 체중량 중위수에 대한 95% 신뢰구간은 [2.08kg, 2.38kg]인 것으로 나타났다.

따라서 생존여부에 따라 체중의 대표값에 유의미한 차이가 있고, 생존그룹에 비해 사망그룹의 체중량이 낮아 체중량이 사망의 요인이라고 해석할만한 충분한 통계적 근거가 있다.

4.3 체중 외 사망요인 파악

신생아의 체중 이외에, 산모의 사회경제적 배경이 신생아의 생존에 미치는 영향을 통계적으로 분석하고자 하였다. 구체적으로는 산모의 연령, 학력, 혼인 여부라는 세 가지 변수를 중심으로, 출생아의 생존 여부와의 관계를 각각 교차표(그림 5 참고) 및 Fisher's Exact Test를 통해 검토하였다.

일부 교차표의 기대빈도가 충분히 크다는 점에서 일반적으로는 카이제곱 검정이 사용될 수 있으나, 본 연구에서는보다 보수적인 해석을 위해 전 항목에 Fisher's Exact Test를 일관되게 적용하였다. 이는 분포나 표본 조건에 덜 민감한 비모수 검정 방식을 통해 분석 전반의 해석 신뢰도를 높이고, 사회경제적 변수 전반에 대한 통계적 일관성을 확보하려는 목적에 부합한다.

먼저 산모 연령을 기준으로 35세 미만과 35세 이상 두 집단으로 나누어 출생아의 생존 여부를 비교하였다. 이는 한국 의료계에서 35세 이상을 '고령 출산'으로 분류하는 기준에 따른 것이다. 검정 결과, p-value는 0.0001로 유의수준 0.05보다 충분히 작아 귀무가설(산모 연령대와 출생아 생존여부 간에 관련이 없다.)은 기각되었으며, 이에 따라 "산모의 연령이 높을수록 출생아 사망률이 높아지는 경향이 있다"는 대립가설을 채택하였다.

다음으로 산모의 학력 수준을 고등학교를 기준으로 '학력 낮음' 집단과 '학력 높음' 집단으로 구분하여 출생아의 생존 여부를 비교하였다. 검정 결과, p-value는 2.028e-27로 유의수준 0.05보다 작아 귀무가설(산모 학력과 출생아 생존여부 간에 관련이 없다.)은 기각되었으며, 이에 따라 "산모의 학력이 낮을수록 출생아 사망률이 높아지는 경향이 있다"는 대립가설이 채택되었다.

마지막으로 산모의 혼인 여부를 기준으로 '기혼'과 '미혼' 두 집단으로 구분하여 출생아의

생존 여부를 비교하였다. 검정 결과, p-value는 3.87e-17로 유의수준 0.05보다 충분히 작아 귀무가설(혼인여부와 출생아 생존여부 간에 관련이 없다)은 기각되었으며, 이에 따라 “기혼인 경우 출생아 사망률이 낮아진다”는 대립가설이 채택되었다.

체중량 외에도 산모의 연령, 학력, 혼인 여부가 출생아 사망과 유의미한 관련이 있음을 통계적으로 확인하였다. 따라서 출산 및 영유아 사망 예방을 위해서는 이러한 배경 요인을 고려한 맞춤형 지원이 병행되어야 한다.

산모 연령대 - 생존여부 교차표 (단위: 명)			산모 학력 - 생존여부 교차표 (단위: 명)			산모 혼인여부 - 생존여부 교차표 (단위: 명)		
	사망	생존		사망	생존		사망	생존
35세 이상*	415	103698	학력낮음	349	66071	기혼	988	318459
35세 미만	706	222003	학력높음	631	254439	미혼	73	7091

*한국 의료계에서는 35세 이상을 고령 출산이라고 규정

*산모의 학력이 미상인 경우 제외하였음

그림 5 교차표

4.4 출생아 생존여부 예측

출생 직후 사망은 발생률이 낮은 희귀 사건이지만, 발생 시 치명적인 결과를 초래한다는 점에서 조기 선별이 무엇보다 중요하다. 이에 본 연구는 사망 고위험군을 사전에 식별하는 모델을 구축하고자 하였다.

사망 여부를 이진 분류하기 위한 예측 모델로는 CART와 Spline 회귀모형을 적용하였다. 두 모델 모두 분포 가정 없이 자료의 구조적 패턴을 탐색할 수 있는 비모수적 접근 방식으로, 변수 간 관계가 비선형적이거나 복잡한 경우에도 유연하게 적용 가능하다.

CART 모델의 성능은 Recall 0.66, Accuracy 0.8955로 나타났다. 사용된 변수는 출생아 체중, 임신주수, 모 연령, 부 연령, 혼인 여부, 모·부의 직업 코드, 교육 정도 코드, 출생자 성별 코드, 출생장소 코드, 다태아 여부, 출생자 유형 코드 등으로 구성되었다. 특히 출생아 체중을 기준으로 성능을 구간별로 분석한 결과, 2.5kg 이하 저체중군에서 Recall은

0.87로 매우 높았던 반면, 2.5kg 초과군에서는 Recall이 0.34로 낮아졌다. 즉, 체중이 정상인데 사망한 사례는 제대로 예측하지 못했고, 이는 모델이 반영하지 못한 다른 위험 요인이 존재함을 시사한다.

Spline 회귀모형은 출생아 체중이라는 단일 연속형 변수와 사망 확률 간의 비선형 관계를 추정하기 위해 적용되었다. 해당 모델의 예측 성능은 Recall 0.56, Accuracy 0.913으로 나타났으며, 단일 변수만을 활용한 모형임에도 일정 수준 이상의 분류 성능을 확인할 수 있었다. 또한 임신주수, 출생아 체중, 모 연령, 부 연령 등 네 가지 연속형 변수를 함께 입력한 경우, Recall은 0.62, Accuracy는 0.9329로 향상되었다. 이는 단일 변수보다 다변수 기반의 예측이 성능 면에서 더 우수함을 보여준다. 단일 변수 모델 결과를 시각화한 사망 확률 곡선(그림 6 참고)을 통해, 출생아 체중이 낮을수록 사망 확률이 급격히 상승하고 일정 수준 이상에서는 완만하게 감소하는 비선형적 경향이 확인되었다. 이러한 결과는 출생아 체중만으로도 사망 위험의 전반적인 경향성을 파악할 수 있음을 보여준다.

Tree 모델과 Spline 모델은 모두 최소한의 정보만으로 일정 수준 이상의 예측 성능을 보였으며, 두 모델을 활용해 5세 미만 영유아 사망 고위험군을 조기에 선별할 수 있다.

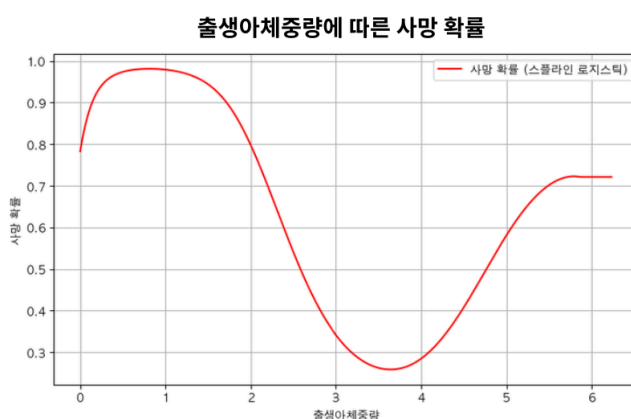


그림 6 출생아체중량에 따른 사망 확률

5. 결론

본 연구는 출생아 사망에 영향을 미치는 생물학적 요인과 부모의 사회경제적 요인을 분석하고, 5세 미만 영유아 사망의 주요 원인을 분석하고 예측 모델을 구축하고자 하였다.

분석 결과, 5세 미만 영유아 사망은 단순히 한 가지 요인만으로 설명되기 어려우며, 체중과 체중 외의 요인 2가지로 인해 발생하였다.

생존 여부에 따라 체중의 대표값에 차이가 있었고, 사망 그룹의 체중이 생존 그룹보다 유의하게 낮게 나타나 체중 자체가 사망의 한 요인으로 해석될 수 있었다. 또한 산모의 연령, 학력, 혼인 여부 등 다양한 사회경제적 요인과 유의미한 관련이 있는 것으로 나타났다.

CART 및 Spline 회귀모형 예측에서는 최소한의 기초 정보만으로도 일정 수준 이상의 예측 정확도가 확인되었으며, 이를 통해 출생아 생존 여부를 바탕으로 사망 고위험군을 조기에 선별하는 것이 가능함을 보여주었다. 다만 체중이 정상 범위임에도 불구하고 사망한 사례가 다수 존재했으며, 이는 모델이 반영하지 못한 외부 요인 또는 사회적 조건이 영향을 미쳤을 가능성을 시사한다.

따라서 출생아 사망 예방을 위한 개입은 저체중 출생의 예방에만 국한되어서는 안 되며, 산모의 연령, 학력, 혼인 상태 등 다양한 배경 요인을 함께 고려한 다각적이고 통합적인 대응이 필요하다.

6. 코드 및 참고문헌

사용한 데이터 및 코드

[1] 통계청 마이크로데이터 통합서비스(MDIS) 5세 미만 영유아 출생-사망 연계

<https://mdis.kostat.go.kr/ofrData/selectOfrDataDetail.do?survId=1000017&itmDiv=1&nPage=3&itmId=2001&itemNm=%EC%9D%B8%EA%B5%AC>

[2] 깃허브

<https://github.com/iseonjae/Nonparametric-Statistics-Project>

논문 및 보고서

[1] 김동섭, "산모 연령 증가가 영아 사망에 미치는 영향", 차의과학대학교(2019).

[2] 박상희, "5세 미만 영유아 세부시기별 사망위험요인과 사망원인 분석: 통계청 5세 미만 영유아 출생사망연계자료를 활용하여(2012~2015년)", 보건사회연구 44(2), 50-66.(2024).

뉴스기사

[1] 이영준. "2025년 인구 '데드크로스' 가속... 50년 뒤엔 둘 중 한 명은 고령층." *서울신문*, 2023.

<https://www.seoul.co.kr/news/economy/2023/12/15/20231215001005>