

제목: 금리 변동과 주택담보대출이 결혼율에 미치는 영향

조원:2024002939 조주성
2024005107 김민혁

1. 서론

가. 분석 목적

최근 대한민국 청년층의 주거 문제는 매우 심각한 수준으로 보인다.

서울 특별시 부동산 정보광장에서 제공하는 통계에 따르면, 2024년 4월부터 2025 4월까지 1년간 서울 아파트 평균 매매가격은 약 12억 4095만 원이다.

반면 KOSIS의 '성별 연령대별 소득'자료에 따르면, 2023년 기준 20~30대의 평균 연 소득은 3895만 원에 불과하다. 이 수치를 바탕으로 청년층이 소득 전부를 저축한다고 가정하더라도 서울의 아파트를 구매하려면 31년 이상이 걸리는 것을 알 수 있다.

이러한 현실 속 청년층이 자력으로 주거를 마련하는 것은 불가능에 가깝고, 자연스럽게 주택담보대출(주담대)에 의존하는 것이 필수적 선택이 되고 있다. 하지만 기준금리가 인상될 경우 주담대의 금리도 같이 상승하므로 대출 자체에도 큰 부담이 되어 청년층의 내 집 마련이 어려워지고 있는 현실이다.

따라서 주거의 불안정은 결혼이나 출산과 같은 결정에도 부정적 영향을 미치게 되며, 사회적 흐름도 결혼을 미루거나 출산을 감소가 크게 나타나고 있다. 이러한 구조적 문제를 배경으로 금리 변동과 주택담보대출 금리 변동의 청년층의 혼인율 및 초혼 연령에 미치는 영향을 객관적 수치로 분석하고자 한다.

이를 위해 5년간의 기준 금리와 주택담보대출 금리, 주택담보대출의 현황, 연령별 혼인율 및 초혼 연령 데이터를 분석하여 금리와 주택담보대출의 부담이 청년층의 결혼 결정에 어떠한 영향을 미쳤는지 구체적으로 분석할 예정이다

나. 데이터 개요

데이터: 한국은행 경제통계시스템(ECOS)에서 기준금리 데이터와 KOSIS 국가통계포털에서 예금은행 신규취급액 기준 주택담보대출 금리 데이터를 수집한다. 또한 서울 특별시 부동산 정보광장에서 서울 아파트 평균 매매가격 데이터를 확보한다. KOSIS에서 연령별 혼인율과 초혼연령, 성별·연령대별 소득 데이터를 수집하여 금리 및 주택담보대출 흐름과 결혼율 및 초혼 연령 변화 간 상관관계를 분석한다.

형식: CSV(정형 데이터)

주요 변수: 기준금리 변동, 주택담보대출 평균 금리 변동, 서울 아파트 평균 매매가격 변동, 연령별 혼인율, 초혼 연령, 성별·연령별 소득

데이터 건수: 모든 데이터는 2019년 4월~2024년 4월까지 5년치의 양을 가진다.

데이터 건수:

- 혼인율 데이터: 253건 (연도·지역·연령대별)
- 초혼 연령 데이터: 1296건 (성별·지역·연령 기준)
- 주담대 금리 데이터: 19건 (월별, 항목별 금리)
- 아파트 매매가격 데이터: 18건 (월별 평균)
- 기준금리 데이터: 1건 (연도별 기준금리 항목별)

유의사항 및 한계: 결혼율 및 초혼 연령 변화에 경제적 요소만 영향을 미친다고 단정할 수 없고, 문화·사회적 요인 등 다른 변수들도 함께 작용할 가능성이 있다.

1	행정구역별	2019.01	2019.02	2019.03	2019.04	2019.05	2019.06	2019.07	2019.08	2019.09
2	전국	90.5	90.4	90.3	90.1	89.9	89.8	89.7	89.7	89.7
3	서울특별시	93.4	93.2	93.0	92.9	92.8	92.7	92.8	92.9	93.1
4	부산광역시	90.1	90.0	89.8	89.6	89.4	89.3	89.1	88.9	88.8
5	대구광역시	88.1	88.1	88.2	88.2	88.2	88.3	88.3	88.3	88.4
6	인천광역시	85.8	85.7	85.7	85.6	85.6	85.6	85.6	85.5	85.6
7	광주광역시	94.7	94.8	94.9	94.9	94.9	94.8	94.8	94.8	94.8
8	대전광역시	77.0	77.1	77.2	77.2	77.3	77.6	77.9	78.5	79.2
9	울산광역시	91.2	90.8	90.4	89.9	89.6	89.2	88.9	88.7	88.5
10	세종특별자치시	71.5	71.5	71.3	70.9	70.7	70.4	70.2	70.1	70.1
11	경기도	84.5	84.4	84.3	84.0	83.8	83.6	83.6	83.6	83.7
12	강원특별자치도	97.8	97.7	97.4	97.2	96.9	96.6	96.3	95.9	95.5
13	충청북도	95.8	95.6	95.3	95.0	94.6	94.3	94.0	93.6	93.4
14	충청남도	94.7	94.5	94.3	94.1	94.0	93.8	93.7	93.6	93.5

사진 1_집값 평균 데이터

<사진 1>과 같이 집값 평균 데이터는 지역(행정 구역)별로 월 단위로 세분화되어 있으며, 데이터의 범위가 넓고 복잡하다.

따라서 특정 행정구역을 지정하여 변수를 통일하고 비교 가능하도록 정제하는 작업이 필요하다. 이 과정은 분석 정확도 향상 및 해석의 일관성을 유지하기 위해 특정 행정구역을 선택해야 한다.

	연도	기준금리	주담대금리		평균소득	평균주택가격	
0	2019	1.25	3.4475	58640000	90.016667	239159	5.4
1	2020	0.50	2.8	59660000	92.575000	213502	5.5
2	2021	1.00	2.8825	61800000	100.450000	192507	6.7
3	2022	3.25	4.295	64700000	103.866667	191690	6.3
4	2023	3.50	5.19	67620000	96.250000	193657	6.3
5	2024	3.00	4.725	71850000	96.025000	222412	6.3

사진 2_데이터 통합

<사진 2>처럼 여러 csv 파일을 보기 쉽게 테이블 형식으로 만드는 과정이 필요하다. 이처럼 여러 출처에서 수집한 데이터는 구조적으로 상이하기 때문에, 단순 수집만으로는 인사이트를 도출하기 어렵다.

따라서 다양한 csv 파일을 하나의 분석

프레임으로 묶기 위해 전처리, 병합, 변수 통일화 등의 과정이 필요하다.

2. 데이터 전처리

가. 결측치 처리

1) 이 분석에 사용된 데이터는 모두 공공기관에서 제공한 연도별 집계 수치로 되어 있어, 원본 csv파일의 결측치는 존재하지 않으며 별도의 삭제 및 대체 처리할 데이터는 없었다. 하지만 여러 데이터셋을 월 단위 기준으로 병합하는 과정에서 일부 기간에 따라 결측치가 발생하였다. 특히, 연 단위로 제공되는 PIR, 평균소득, 혼인건수 등의 변수는 월 단위 분석을 위해 변환할 때 중간값이 누락되어 결측이 생성되었다.

2) 처리방법

```
final_df = final_df.sort_values('기준월')
final_df = final_df.set_index('기준월').interpolate().ffill().bfill().reset_index()
```

<코드-1>: 결측치 보간 및 채움 처리 코드

<코드-1>과 같이 결측치는 시계열 상 시간 흐름을 고려하여 결측값을 직선적으로 추정하는 선형 보간법과 앞의 값을 이용해 결측을 채우는 선방향 채움, 뒤의 값을 이용해 결측을 보완하는 후방향 채움을 통해 결측으로 인한 왜곡을 방지하고 월별 시계열 데이터 프레임 구성하였다.

	기준월	기준금리	주담대금리	평균소득	평균주택가격	혼인건수	PIR
0	2019-01	1.463889	3.384580	3090000.0	330419.365854	237050.205479	5.4
1	2019-02	1.463889	3.341328	3090000.0	329831.097561	235081.997260	5.4
2	2019-03	1.463889	3.293443	3090000.0	329211.902439	232902.909589	5.4
3	2019-04	1.463889	3.247102	3090000.0	328506.926829	230794.115068	5.4
4	2019-05	1.463889	3.199216	3090000.0	328100.658537	228615.027397	5.4
...
67	2024-08	3.266091	4.761316	3630000.0	425456.097561	222412.000000	6.3
68	2024-09	3.254317	4.761316	3630000.0	427611.902439	222412.000000	6.3
69	2024-10	3.145833	4.761316	3630000.0	428793.487805	222412.000000	6.3
70	2024-11	3.000000	4.761316	3630000.0	429425.487805	222412.000000	6.3
71	2024-12	3.000000	4.761316	3630000.0	429609.463415	222412.000000	6.3

72 rows × 7 columns

<표-1>: 월별 시계열 데이터 프레임

<표-1>은 전처리 후 생성된 통합 데이터 프레임의 일부이다. 이 데이터는 총 72개 (2019년 1월~2024년 12월)의 월 단위를 기준으로 각 열에 기준금리, 주담대금리, 평균소득, 평균주택가격, 혼인건수, PIR 지표가 선형보간(interpolate), 선방향 채움(ffill), 후방향 채움(bfill)을 통해 결측값이 없는 시계열 데이터 셋임을 알 수 있다.

나. 이상치 처리

사용한 데이터는 KOSIS, 한국은행과 같은 공신력있는 공식 기관으로부터 제공된 정제된 수치형 통계자료로 구성되어 있다. 데이터는 모두 연도별 또는 월별 수치로 집계되어 있고, 개별 표본 단위가 아닌 요약 통계 형식으로 제공되어 통계적으로 크게 벗어나는 이상치는 없었다.

다. 스케일링: 본 분석은 다양한 경제 지표 및 사회 지표 데이터를 통합하였다.

변수명	수치 범위(예)	단위
기준금리	0.5~3.5	%
주택담보대출금리	2.43~4.73	%
평균소득	178~488	만 원
평균주택가격	155,831~1,495,939	천 원
혼인건수	약 19만~24만 9천	건
PIR	5.4~6.7	%

<표-2>: 각 변수의 범위 및 단위

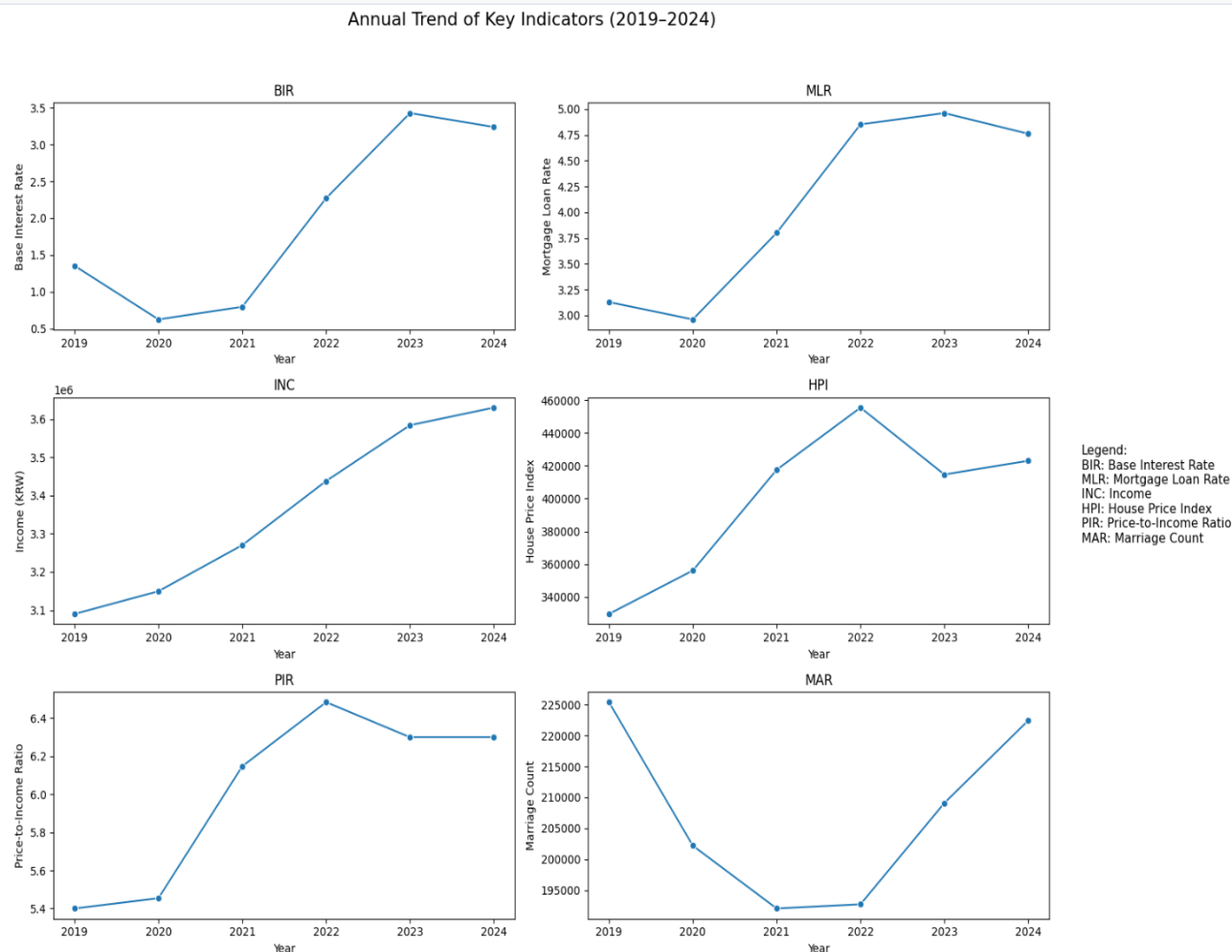
<표-2>와 같이 각 변수들은 단위와 수치의 스케일이 상이하고, 변수 간의 값 범위가 큰 차이를 보인다.

본 분석에서는 혼인건수를 예측하기 위한 머신러닝 회귀모델(Ridge, Lasso, ElasticNet)의 학습 안정성과 성능 향상을 위해 데이터 정규화를 적용하였다. 이는 변수 간 단위와 수치 범위가 크게 차이나는 문제를 해소하고, 학습 시 특정 변수가 과도한 영향을 주는 것을 방지한다.

예를 들어, 기준금리와 주택담보대출금리는 소수점 수준(%)으로 제공되며, 평균주택가격은 수백만 원, 평균소득은 수천만 원, 혼인건수는 수십만 건의 범위를 갖는다. 이러한 차이를 고려하여, 모든 입력 변수(기준금리, 주담대금리, 평균소득, 평균주택가격, 혼인건수, PIR)를 사용해 평균 0, 표준편차 1로 정규화(z-score scaling) 하였으며, 종속 변수인 혼인건수 역시 동일한 방식으로 정규화하여 모델 학습에 사용하였다.

학습 이후에는 예측 결과를 다시 원래 단위로 복원하기 위해 역정규화과정을 수행하였다. 이러한 스케일링 과정은 모델의 수렴 속도를 높이고, 정규화 회귀모델의 효과를 극대화하며, 전체적인 예측 성능 향상에 기여한다.

3. 데이터 분석
가. 상관관계 분석



<그래프-1>: 주요 지표의 연간 추세(2019~2024)

<그래프-1>은 월별 시계열 데이터를 연간 시계열 데이터로 변환 후 그래프 형식으로 나타낸 것이다. 각 그래프의 설명은 다음과 같다.

1) BIR(Base Interest Rate)-기준금리

설명: 기준금리는 2020년 코로나 19의 영향으로 0.5%까지 낮아진 후, 이후 글로벌 인플레이션의 영향으로 2022년부터 급격히 상승하여 2023년 3.5%까지 도달하였다. 2024년에는 인플레이션 완화에 따른 소폭 인하된 모습이다.

의미: 기준 금리는 한 나라의 중앙은행(한국은행)이 설정하는 정책 금리로, 금융기관들이 한국은행과 거래할 때 기준이 된다. 통화정책의 핵심 수단이므로 물가 안정과 경기 조절에 직접적 영향을 미친다. 금리가 낮으면 기업 및 개인의 대출 여력 증가 및 소비 촉진이 활성화되며, 금리가 높으면 자금 조달 비용이 상승하여 소비·투자가 위축되는 경향을 가진다. 본 데이터에서는 기준금리의 변화가 주담대 금리, PIR, 혼일율 등 여러 변수에 연쇄적으로 영향을 미친다는 점에서 매우 중요한 변수로 작용한다.

2) MLR(Mortgage Loan Rate)-주택담보대출 금리

설명: 주담대 금리도 2020년 약 3.0%까지 하락 후, 2021년부터 상승을 시작하여 2023년에는 약 5%에 도달하였다. 이는 기준금리의 영향과 은행권의 대출 리스크를 반영한 결과이다.

의미: 주담대 금리는 실소요자의 주택 구입 결정에 직결되는 요인 중 하나이다. 금리 상승은 대출 상환 부담을 높이고 자가 마련 계획을 지연시키는 원인이다. 청년층이 주담대에 의존하여 내 집 마련을 시도하는 구조에서 주담대 금리 상승은 결혼율과 출산에 부정적 영향을 미칠 수 있다.

3) INC(income)-평균소득

설명: 분석 기간 동안 평균소득은 꾸준히 증가하는 모습을 보인다. 2019년 약 3,090만 원에서 2024년에는 약 3,630만 원까지 상승했으며, 이는 약 17.5% 증가율이다. 이는 최저임금 인상, 노동시장 회복, 서비스업 회복세 등이 복합적으로 작용한 결과로 볼 수 있다.

의미: 소득의 증가는 소비 여력을 늘려 경기 회복에 긍정적으로 작용하지만, 동일 기간 동안 주택가격 및 대출금리의 상승 속도가 더 빨랐기 때문에 체감 소득 여력은 오히려 감소했을 가능성이 있다. 특히 PIR의 동반 상승은 실질 주거 구매력이 약화되었음을 시사한다.

4) HPI(House Price Index)-평균주택가격

설명: 2019년 약 340,000 수준이던 평균 아파트 가격은 2022년 약 450,000 이상으로 급등한 후, 2023년에 다소 조정되었다가 2024년에는 소폭 상승하였다. 수도권을 중심으로 한 공급 부족, 저금리 기조, 부동산에 대한 투자수요 증가가 복합적으로 작용하였다.

의미: 주택가격 상승은 PIR을 끌어올려 주거비 부담을 가중시키며, 특히 신혼부부나 청년층의 자가 구입의 장벽을 높이는 주요 요인으로 작용한다. 이는 혼인율에도 부정적인 영향을 줄 수 있으며, 전세난·매매 불균형과 같은 사회적 문제로 확대될 수 있다.

5) PIR(Price to Income Ratio)-소득 대비 주택가격 비율

설명: PIR은 2019년 5.4에서 2022년 6.5로 급등하였다가, 이후 약간 하락하여 2024년에는 6.3으로 유지되고 있다. 이는 집값이 소득보다 더 빠르게 상승했음을 나타낸다.

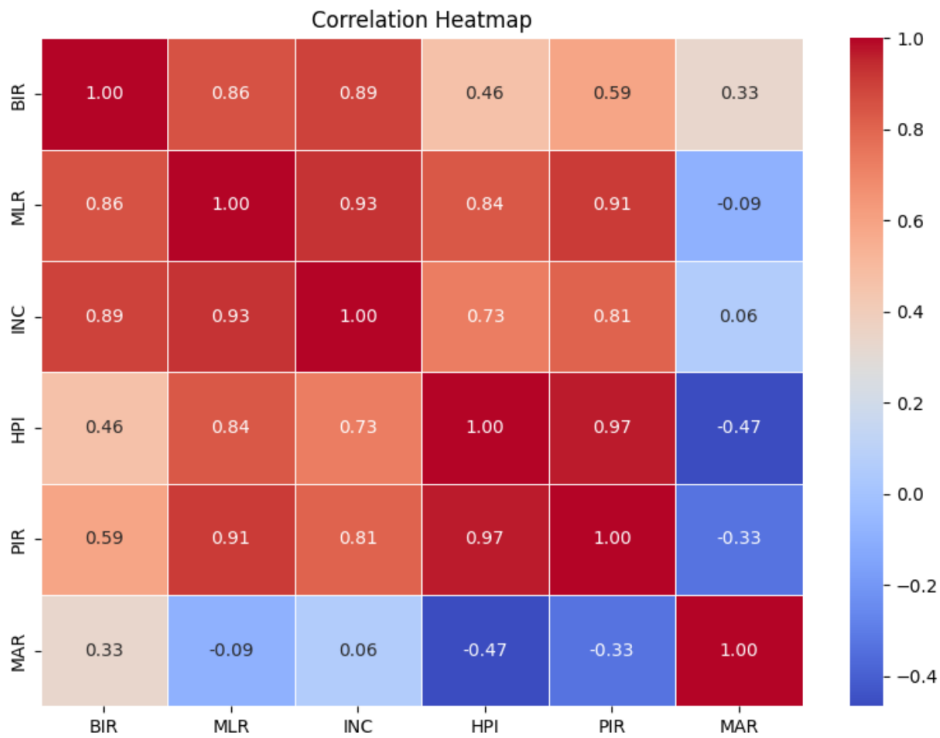
의미: PIR은 주거 구입력의 지표로 사용되며, 수치가 높을수록 '내 집 마련'이 어려워지는 구조를 의미한다. PIR의 상승은 청년층이 결혼을 미루거나 주택 구입 대신 전·월세에 머무는 상황으로 이어질 수 있으며, 장기적으로 혼인율·출산율 저하와 같은 사회구조적 문제와도 연결될 수 있다.

6) MAR(Marriage Count)-혼인건수

설명: 2019년 약 22만 5천 건이던 혼인건수는 2021년 약 19만 3천 건까지 감소하였고, 이후 완만한 회복세를 보이며 2024년에는 약 22만 건 수준으로 회복되었다.

의미: 혼인율 감소는 주거 불안정, 고용 불안, 경기침체, 코로나19와 같은 외생 변수들이 복합적으로 작용한 결과로 해석된다. 주거비 부담(PIR 상승)과 혼인건수 감소 간의 음의 상관관계는 이러한 구조를 뒷받침하며, 정책적 개입 없이는 혼인율 회복이 쉽지 않다는 점을 시사한다.

나. 히트맵



<그림-1>: 주요변수 상관관계 히트맵

<그림-1>은 본 분석에 사용된 주요 변수 간의 상관관계 히트맵(Correlation Heatmap)을 나타낸 것이다.

이 히트맵은 변수 간의 선형 상관관계를 시각적으로 표현한 것으로, 값이 +1에 가까울수록 강한 양의 상관관계, -1에 가까울수록 강한 음의 상관관계, 그리고 0에 가까울수록 상관관계가 거의 없음을 의미한다.

각 변수별 상관관계는 다음 <표-2>와 같다.

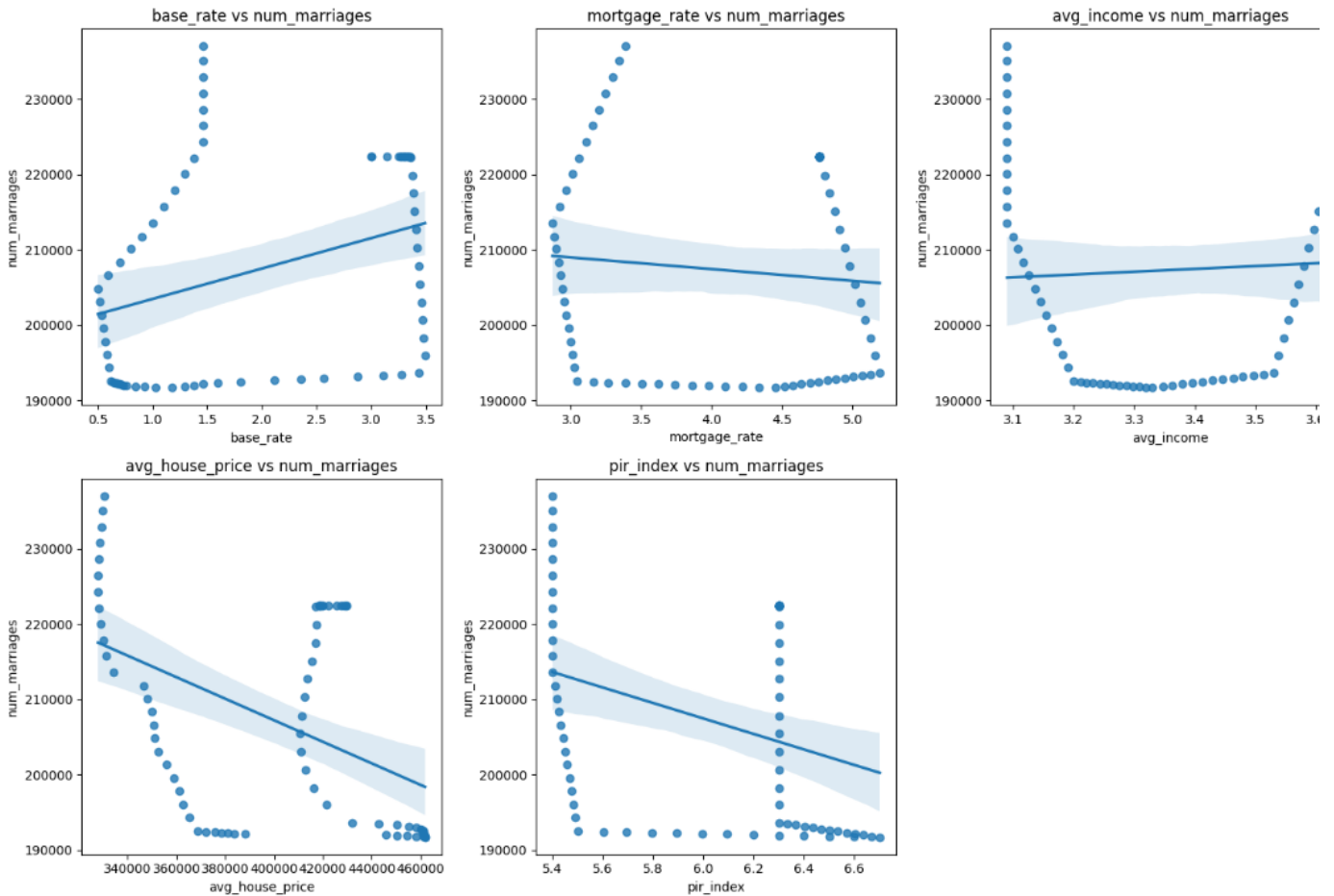
변수 쌍	상관 계수	관계 해석	원인 분석
BIR- MLR	0.86	매우 강한 양의 상관	기준금리(BIR)가 오르면, 시중은행의 주담대금리(MLR)도 거의 동시에 인상됨. 이는 통화정책 전이 효과 때문이다
BIR- INC	0.89	매우 강한 양의 상관	기준금리(BIR)가 경기 과열 시기에 상승하므로, 소득(INC)이 증가하는 경기확장기와 동반 상승하는 경향을 보인다.
BIR- HPI	0.46	중간 양의 상관	기준금리 인상 시점이 주택 가격이 이미 많이 오른 이후 정책 지연 효과로 양의 상관관계를 가질 수 있다
BIR- PIR	0.59	중간 양의 상관	PIR = 집값/소득이며, 기준금리 인상은 소득 증가 시기와 겹치고, 집값이 단기적으로 덜 떨어지기 때문에 긍정적 상관 가능하다
BIR- MAR	0.33	약한 양의 상관	기준금리와 혼인건수는 직접적 관계는 적고, 소득 향상기와 일치한 상승으로 인해 약한 양의 상관 가능성. 실질적 인과는 희박하다
MLR- INC	0.93	매우 강한 양의 상관	주담대금리는 기준금리 영향을 받으며, 소득이 증가하는 시기에 대출수요가 늘어나며 동반 상승하기도 한다
MLR- HPI	0.84	매우 강한 양의 상관	집값이 오르면 대출 수요도 증가하므로 주담대금리 상승에 압력을 주고, 은행의 위험 프리미엄 반영한다.
MLR- PIR	0.91	매우 강한 양의 상관	PIR 상승은 집값 급등 혹은 소득 감소 의미한다. 둘 다 MLR 증가와 동반될 수 있고, 집값 폭등 시기와 강하게 연결한다.
MLR- MAR	-0.09	거의 무관 (약한 음의 상관)	대출금리 높을수록 결혼 비용 부담이 크지만, 직접적 영향 보다는 간접 효과임을 알 수 있고, 사실상 상관이 없다.
INC- HPI	0.73	강한 양의 상관	소득 증가 시점에 자산 투자 수요가 늘고, 주택 수요의 증가로 집값이 상승한다. 주택 구매 여력과 연결된다.
INC- PIR	0.81	강한 양의 상관	이론적으로 소득 증가는 PIR의 하락이지만, 실제로는 집값이 더 가파르게 오르기 때문에 동시에 상승할 수 있다.
INC- MAR	0.06	상관없음	소득과 결혼은 직접적 관계보다 주거 및 일자리의 안정성과 연결된다. 소득이 높다고 무조건 결혼 증가는 아니다.
HPI- PIR	0.97	거의 완전 한 양의 상 관	PIR=HPI/INC이며, 구조상 매우 높은 상관 관계를 보인다. 특히 주택가격이 소득보다 급격히 오를 경우 더욱 도드라진다.
HPI- MAR	-0.47	중간 음의 상관	주택가격 상승은 결혼에 필요한 주거 마련 비용을 증가시키므로 결혼 포기 또는 연기를 유도한다. 강한 부정적 영향을 알 수 있다.
PIR- MAR	-0.3	약한 음의 상관	PIR의 상승은 주거비 부담을 의미하고 혼인율에 부정적 영향을 미친다.

<표-2>: 주요 변수 쌍별 상관관계 분석

위 <표-2>를 통해 확인할 수 있는 점은 주택 가격(HPI)과 주거부담지수(PIR)는 혼인건수(MAR)와 유의미한 음의 상관관계를 보이며, 이는 주거 비용이 결혼 결정에 부정적인 영향을 미칠 수 있음을 시사한다.

반면, 소득(INC), 기준금리(BIR), 주택담보대출금리(MLR) 등은 혼인건수(MAR)과 뚜렷한 상관관계를 보이지 않으며, 이는 혼인건수에 직접적 영향보다는 간접적·구조적 영향을 미치는 것임을 알 수 있다.

다. 혼인건수(Num marriages)와 주요 변수 간 관계 시각화



<그림-2>: 혼인건수(num_marriages)와 주요 변수들 간의 관계를 시각화한 회귀 산점도

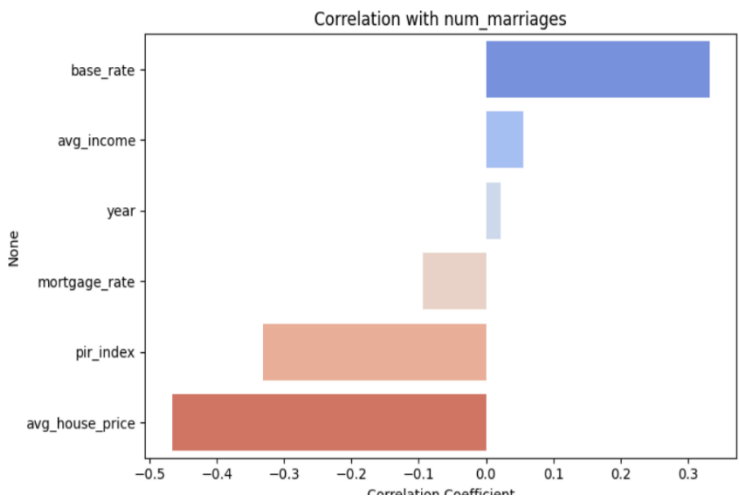
<그림-2>는 혼인건수(num_marriages)와 주요 경제 지표 간의 관계를 시각적으로 분석한 결과이다. 각각의 그래프는 해당 변수와 혼인건수 간의 산점도와 함께 선형 회귀선을 포함하고 있다. 변수의 변화에 따라 혼인건수가 어떤 추세를 보이는지를 직관적으로 확인할 수 있다.

기준금리는 혼인건수와 약한 양의 상관관계를 보인다. 금리가 상승한 시점에서 혼인건수는 일부 회복세를 보였으나, 이는 간접적인 관계일 가능성이 높다. 반면, 주택담보대출 금리는 약한 음의 상관관계를 나타내며, 대출금리 인상이 결혼 결정에 부담 요인으로 작용할 수 있음을 시사한다.

평균소득은 혼인건수와 거의 무관한 경향을 보인다. 소득 수준의 상승은 전체적인 경제 안

정성을 높이지만, 혼인율에는 직접적인 영향을 주지 않는 것으로 분석된다.

특히 평균주택가격과 PIR은 혼인건수와 뚜렷한 음의 상관관계를 보이며, 주택 구매 부담이 커질수록 결혼을 미루거나 포기하는 경향이 강화됨을 보여준다. 이는 실제 생활 속에서 주거 안정이 결혼의 전제 조건으로 작용하고 있음을 의미하며, 주거 부담이 혼인율 하락의 구조적 요인 중 하나로 작용하고 있음을 뒷받침하는 결과이다.



<그림-3>: 혼인건수와 주요 변수 간 상관계수 막대그래프

<그림-3>은 혼인건수(num_marriages)와 기준금리, 주담대금리, 평균소득, 주택가격, PIR 간의 피어슨 상관계수를 막대 형태로 나타낸 그래프. 주택 관련 변수들이 혼인건수와 음의 상관관계를 보이며, 경제적 주거 부담이 결혼 결정에 부정적 영향을 줄 수 있음을 시사한다.

4. 머신러닝 모델 적용

가. 사용된 모델

다중 회귀 분석 모델(선형): Ridge, Lasso, ElasticNet

앙상블 회귀 모델(비선형): Random Forest

```
# 1. 데이터 구성
X = final_df[['base_rate', 'mortgage_rate', 'pir_index', 'avg_house_price', 'avg_income']].astype(float)
y = final_df['num_marriages'].astype(float).to_numpy().reshape(-1, 1)

# 2. 정규화
scaler_X = StandardScaler()
X_scaled = scaler_X.fit_transform(X)

scaler_y = StandardScaler()
y_scaled = scaler_y.fit_transform(y)

# 3. 예측용 2025년 입력 데이터
X_input_raw = np.array([[2.75, 4.17, 7.12, 408934, 2910000]])
X_input_scaled = scaler_X.transform(X_input_raw)

# 4. 모델들 정의
models = {
    "Ridge": Ridge(alpha=1.0),
    "Lasso": Lasso(alpha=0.1),
    "ElasticNet": ElasticNet(alpha=0.1, l1_ratio=0.5)
}

results = []

for name, model in models.items():
    model.fit(X_scaled, y_scaled)

    # 학습 데이터에 대한 예측 및 평가
    y_pred_scaled = model.predict(X_scaled)
    y_pred_original = scaler_y.inverse_transform(y_pred_scaled.reshape(-1, 1))

    mse = mean_squared_error(y, y_pred_original)
    r2 = r2_score(y, y_pred_original)

    # 2025년 예측
    y_2025_scaled = model.predict(X_input_scaled)
    y_2025_original = scaler_y.inverse_transform(y_2025_scaled.reshape(-1, 1))[0][0]
    y_2025_final = round(max(0, y_2025_original))

    results.append({
        "모델": name,
        "R²": round(r2, 4),
        "MSE": round(mse, 2),
        "2025년 혼인건수": y_2025_final
    })
```

<코드-2>: 다중 회귀 모델 별 2025 혼인건수 예측

나. 모델 성능

1) 다중회귀분석모델(선형)

<코드-2>와 같이 혼인건수 예측을 위해 총 세 가지 선형 회귀 기반 머신러닝 모델인 Ridge, Lasso, ElasticNet 회귀를 적용하였다. 이들 모델은 모두 정규화를 통해 회귀계수의 크기를 제약하는 방식이며, 과적합을 방지하거나 변수 선택 효과를 제공하는 특성이 있다.

변수로는 기준금리, 주담대금리, PIR, 평균주택가격 소득을 이용해 2025년 혼인건수를 예측하였다. 그리고 각 모델의 정확도는 MSE(평균오차제곱)와 R^2 결정계수를 기준으로 비교 분석하였다.

2025년 혼인건수 예측에는 실제 경제지표 기반의 입력값(2.75, 4.17, 7.12, 408934, 29100000)을 사용하였고, 예측 결과는 역정규화 과정을 통해 실제 혼인건수 단위로 변환되었다.

	모델	R ²	MSE	2025년 예측 혼인건수
0	Ridge	0.6671	65961715.19	235696
1	Lasso	0.5628	86622324.24	210527
2	ElasticNet	0.5764	83937368.57	210802

결과는 <그림-4>와 같다. 정확도는 Ridge>ElasticNet>Lasso이다.

Ridge 회귀 모델이 실제 혼인건수와 비교해도 가장 근접한 수준으로 판단된다.

<그림-4>: 다중 회귀 모델의 예측 정확도

반면, Lasso는 상대적으로 낮은 설명력을 보였고, 예측값도 210,527건으로 다소 보수적인 추정 결과를 나타냈다.

ElasticNet은 Lasso보다는 다소 높은 설명력을 보였지만, 여전히 Ridge에 비해서는 낮은 성능을 보였다. 예측값은 Lasso와 유사한 수준이었다.

종합적으로 볼 때, Ridge 회귀 모델이 가장 안정적이고 우수한 예측력을 가진 모델이다

2) 앙상블 회귀 모델(비선형)

```
# 1. 입력 데이터 및 타겟값 구성
X_rf = final_df[['base_rate', 'mortgage_rate', 'pir_index', 'avg_house_price', 'avg_income']].astype(float).to_numpy()
y_rf = final_df['num_marriages'].astype(float).to_numpy()

# 2. 랜덤 포레스트 모델 정의 및 훈련
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_rf, y_rf)

# 3. 2025년 입력값 구성 및 예측
X_2025_rf = np.array([[2.75, 4.17, 7.12, 408934, 29100000]])
y_2025_rf_pred = rf_model.predict(X_2025_rf)[0]

print(f"2025년 예측 혼인건수(랜덤 포레스트): {round(y_2025_rf_pred)}건")
```

2025년 예측 혼인건수(랜덤 포레스트): 215638건

<코드-3>과 같이 비선형 회귀 모델인 Random Forest Regressor를 적용하여 혼인건수를 예측하였다. 여러 개의 결정 트리를 앙상블 방식으로 구성한 모델인 Random Forest는 각 트리의 예측 결과를 평균 내어 최종 결과를 산출함으로써 과적합을 방지하고 예측의 안정성을 확보하는 특징을 가진다.

<코드-3>: RandomForest의 예측 결과

입력 변수로는 기준금리, 주담대금리, PIR, 평균주택가격 소득을 이용해 2025년 혼인 건수를 예측하였다 Random Forest는 입력 변수 간의 비선형적 상호작용과 변수 중요도에 따른 분기 구조를 학습할 수 있어, 단순한 선형 회귀로는 포착하기 어려운 복잡한 관계를 모델링하는 데 유리하다.

실제 모델 적용 결과, 2025년의 혼인건수를 예측한 값은 약 215,638건으로 나타났다. 이는 선형 기반 모델의 예측값과 비교하여 상대적으로 보수적인 추정이며, 변수 간의 비선형 관계와 상호작용을 고려한 결과로 해석할 수 있다.

3) 모델 일반화 성능 검증

릿지 회귀와 랜덤 포레스트 회귀 모델을 MAPE 검증을 통해 예측값을 비교 검증함으로써, 모델의 일반화 가능성과 정책 적용 타당성을 확보하고자 하였다.

```
from sklearn.metrics import mean_absolute_percentage_error
import numpy as np

# === 검증 입력값 원본
val_x_17_raw = np.array([[1.75, 3.27, 5.6, 255966, 27330000]])
val_x_18_raw = np.array([[1.5, 3.39, 5.7, 285680, 28100000]])

# === 정규화된 입력값 (Ridge 모델 학습 시 사용된 scaler_X 활용)
val_x_17_scaled = scaler_X.transform(val_x_17_raw)
val_x_18_scaled = scaler_X.transform(val_x_18_raw)

# === Ridge 예측 (정규화된 X → 정규화된 y → 역정규화)
ridge_pred_17_scaled = ridge_model.predict(val_x_17_scaled)
ridge_pred_18_scaled = ridge_model.predict(val_x_18_scaled)

ridge_pred_17 = scaler_y.inverse_transform(ridge_pred_17_scaled.reshape(-1, 1))[0][0]
ridge_pred_18 = scaler_y.inverse_transform(ridge_pred_18_scaled.reshape(-1, 1))[0][0]

# === Random Forest은 정규화 불필요
rf_pred_17 = rf_model.predict(val_x_17_raw)[0]
rf_pred_18 = rf_model.predict(val_x_18_raw)[0]

# === 실제 혼인건수
true_17 = 264455
true_18 = 257622

# === MAPE 계산
ridge_mape_17 = mean_absolute_percentage_error([true_17], [ridge_pred_17])
ridge_mape_18 = mean_absolute_percentage_error([true_18], [ridge_pred_18])
rf_mape_17 = mean_absolute_percentage_error([true_17], [rf_pred_17])
rf_mape_18 = mean_absolute_percentage_error([true_18], [rf_pred_18])

# === 결과 출력
print(f"★ Ridge MAPE (2017): {ridge_mape_17:.2%}")
print(f"★ Ridge MAPE (2018): {ridge_mape_18:.2%}")
print(f"--- ★ 30")
print(f"▲ Random Forest MAPE (2017): {rf_mape_17:.2%}")
print(f"▲ Random Forest MAPE (2018): {rf_mape_18:.2%}")

★ Ridge MAPE (2017): 6.15%
★ Ridge MAPE (2018): 4.53%

▲ Random Forest MAPE (2017): 14.31%
▲ Random Forest MAPE (2018): 11.51%
```

<코드-4>: MAPE 검증

변수 간 상호작용을 포착하는 데 강점이 있으며, 복잡한 패턴을 학습할 수 있다는 장점이 있다.

다만, 데이터 특성이나 노이즈에 민감해 오차가 클 수 있으며, 해석력이 낮고 예측의 일관성 측면에서는 다소 불안정할 수 있다.

이처럼 선형 모델과 비선형 모델은 각각의 장점이 다르기 때문에 선형 모델의 정확도가 높다고 선형 모델만 사용할 것이 아니다.

모델의 일반화 성능을 확인하기 위해, 실제 혼인건수가 이미 존재하는 과거 시점 (2017년, 2018년)의 경제지표를 입력값으로 사용하여 릿지 회귀 모델과 랜덤 포레스트 모델의 예측 정확도(MAPE)를 비교 분석하였다.

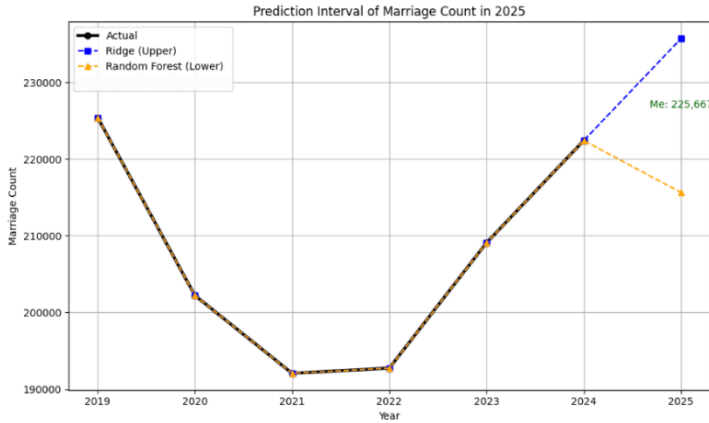
각 연도별 실제 혼인건수를 기준값으로 설정하고, 모델의 예측값과의 평균 절대 백분율 오차(MAPE)를 산출하였다.

분석 결과, 릿지 회귀 모델은 2017년 6.15%, 2018년 4.53%의 낮은 MAPE를 기록하여, 과거 데이터에 대해서도 비교적 안정적인 예측 성능을 보였다.

Ridge 회귀는 선형 기반 모델로 해석 가능성과 학습 안정성이 높으며, 비교적 적은 데이터에서도 과적합에 강하고 일반화 성능이 뛰어난 경향을 보인다. 실제로 2017년과 2018년의 과거 데이터에 대한 MAPE 검증에서도 가장 낮은 오차율을 기록하였다.

반면, Random Forest는 비선형 관계나

4) 릿지 회귀모델과 랜덤포레스트 회귀 모델의 비교



<그림-4>: 선형 및 비선형 회귀모델의 2025년 결혼 건수 예측

Forest 모델에서는 약 215,638건으로 각각 산출되었으며, 두 값의 평균인 225,667건을 중앙 예측값으로 설정하여 시각화하였다. 그래프에서는 Ridge 예측을 상한선(파란 점선)으로, Random Forest 예측을 하한선(주황색 점선)으로 하였고 중앙값을 초록색으로 표현하였다.

이러한 시각적 비교는 단일 모델의 수치를 절대 기준으로 제시하는 것보다, 모델 간 편차와 예측의 불확실성까지 함께 고려할 수 있는 직관적인 의사결정 자료로 활용될 수 있다.

특히 혼인건수와 같이 사회적 함의가 크고 정책적 민감성이 높은 변수의 경우, 복수 모델 기반의 예측 구간 설정은 정책 수립 과정에서 리스크를 줄이는 데 매우 효과적인 기준이 될 수 있다.

더불어, 중앙값 활용은 두 모델의 구조적 특성과 예측 경향의 차이를 보완하고, 극단적인 편향 없이 현실성 있는 수치를 제시하는 접근으로서 의미를 지닌다.

즉, 선형 모델의 안정성과 비선형 모델의 유연성을 균형 있게 반영한 결합적 예측 방식이라 할 수 있다.

혼인건수 예측 결과에 대한 해석력과 설득력을 높이기 위해, Ridge 회귀 모델과 Random Forest 모델의 2025년 예측값을 시각적으로 비교 분석하였다.

<그림-4>는 2019년부터 2024년까지의 실제 혼인건수 추이와, 2025년의 모델 기반 예측값을 함께 보여준다.

2025년 예측값은 Ridge 회귀 모델에서 약 235,696건, Random

5. 결론 및 제안

가. 결론

본 분석에서는 혼인건수에 영향을 미치는 주요 경제 지표로서 기준금리(BIR), 주택담보대출 금리(MLR), PIR(소득 대비 주택가격 비율), 평균소득 및 평균주택가격을 중심으로 분석을 수행하였다. 다양한 회귀 모델과 과거 데이터를 활용한 예측 검증을 통해, 혼인율 변화는 단순한 소득 수준보다는 주거 관련 경제 요인에 더 밀접하게 연관되어 있음을 확인할 수 있다.

분석 결과, 기준금리(BIR)와 주담대금리(MLR)가 오를수록 PIR이 상승하고, 이는 혼인건수에 부정적인 영향을 미치는 경향이 강하게 나타났다. 다시 말해, **BIR → MLR → PIR → 혼인건수 감소**라는 일련의 경제적 연결고리가 확인된 것이다. 이는 단순한 상관분석을 넘어, 시계열적 흐름과 예측 모델에서 일관되게 나타난 구조적 인과 경향으로 해석된다.

따라서, 혼인율 저하 문제를 단기적인 사회적 분위기나 가치관의 변화만으로 설명하기보다는, 금리 및 대출금리의 상승이 주거 부담을 높이고, 결과적으로 혼인 결정을 지연 또는 포기하게 만드는 근본 원인 중 하나로 작용하고 있음을 주목해야 한다.

이러한 인과 구조에 기반할 때, 향후 혼인율 정책을 수립하거나 인구정책을 보완할 때에는, **당시 또는 수개월 후에 나타날 금리 및 주담대금리 변화의 흐름을 고려한 선제적 대응이 필요하다.** 금리와 대출금리 변동이 혼인건수에 미치는 파급력과 시차적 영향을 정량적으로 파악하고, 이에 따라 주거 지원 정책이나 혼인 장려 인센티브의 시점을 정교하게 조정할 필요가 있다.

다만 본 분석은 경제적 요인을 중심으로 혼인율 변화의 구조를 파악하고자 하였기 때문에, **문화적 가치관의 변화, 결혼에 대한 사회적 인식, 비경제적 요인에 대한 고려는 포함되지 않았다.** 향후에는 이러한 비정량적·사회문화적 요인까지 포괄하는 다차원적 분석을 통해 보다 입체적인 혼인율 변화 원인 분석이 필요할 것이다.

나. 제안

1) 금리 및 주담대금리 연동 혼인율 예측 체계 구축

기준금리와 주택담보대출금리의 변동이 일정 수준에 도달할 경우, 향후 수개월 내 혼인건수에 미치는 영향을 사전에 예측할 수 있는 혼인율 위험 예측 모델을 도입할 필요가 있다. 이를 통해 혼인 감소가 예견되는 시점을 선제적으로 파악하고, 정책 대응 시기를 조절할 수 있다.

2) 금리 상승기 혼인 부담 완화형 주거지원 정책 설계

금리 인상 구간에서 PIR이 상승하면 혼인결정이 지연되는 경향이 나타나므로, 청년층 및 신혼부부를 대상으로 한 한시적 주거지원 정책(예: 전세자금 이자 감면, 대출 금리 상한 보장 등)을 만들어 청년층 및 신혼부부의 부담을 한시적으로 줄인다.

3) 시계열 기반 정책 타이밍 최적화 시스템 도입

경제지표 변화와 혼인율 사이의 시차적 관계를 반영하여, 주택금융지원, 혼인 인센티브 정

책 등의 개입 시점을 데이터 기반으로 조정하는 시스템을 구축해야 한다. 금리 인상 후 일정 시점부터 PIR이 급격히 오를 경우 유연하게 대응 정책이 가동되도록 설계되어야 한다.

4) 부처 간 혼인율 대응 협업 체계 강화

본 사안은 단일 부처의 과제가 아닌 만큼, 금융정책을 담당하는 한국은행 및 금융위원회, 주거 정책을 담당하는 국토교통부, 인구정책을 주관하는 보건복지부 등과의 연계 협력 체계 강화가 필요하다. 중장기적 대응을 위해서는 각 기관이 데이터를 공유하고 예측 기반으로 정책을 동기화하는 체계가 필요하다.