

불로소득과 부채가 총 생활비에 미치는 영향

확증적 데이터분석 Semi Project

INDEX

01	연구 개요
02	데이터 전처리
03	상관분석
04	주성분분석(PCA)
05	회귀분석/결과보고
06	잔차분석

01 연구개요

연일뉴스 구독

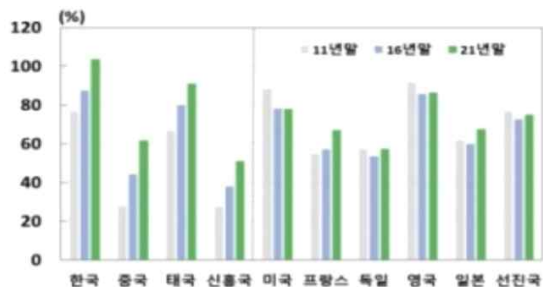
PICK ①

한은 "한국, 가계부채 세계 3위...성장 막고 자산불평등 키워"

입력 2023.07.17, 오후 12:00 | 수정 2023.07.17, 오후 1:42 | 기사원문

신호경 기자

주요국 GDP 대비 가계부채 비율¹⁾



주: 1) 선진국 - 미국, 프랑스, 독일, 영국, 일본, EU, 덴마크 등 11개국, 신흥국 - 한국, 중국, 태국, 아르헨티나, 브라질, 헝가리 등 20개국

현대사회에서 금융 부채 문제와 소득 양극화는 심각한 이슈로 대두되고 있음

연구를 통해 가구별 총생활비와 부채, 불로소득과의 상관관계를 조사한 후 도출된 결론을 바탕으로 다양한 분야에 도움이 될 수 있는 참고 자료를 제공하고자 함

저소득층 근로소득 줄고 고소득층 불로소득 증가...빈부 격차 심화

신호경 기자 | 입력 2022.09.19 09:37 | 댓글 0

코로나19가 확산됐던 지난 2020년 저소득층의 근로소득은 오히려 감소했지만 걸었지만, 고소득층의 불로소득은 빠른 속도로 증가해 소득양극화가 심화된 것으로 나타났다.

18일 진선미 더불어민주당 의원(국회 기획재정위원회)이 국세청으로부터 제출받은 '근로·배당·양도소득 천분위 현황' 자료에 따르면, 2020년 근로소득이 전년 대비 4.5% 증가하는 동안, 배당소득은 27%, 양도소득은 49.7% 증가했다.

02 데이터 전처리

데이터 준비 한국 복지 패널 데이터 활용

	총생활비	이자_배당금_연간소득	임대료_연간소득	기타_소득	부모가보낸_생활비	자녀가보낸_생활비	민간보조금	세금환급금	금융기관대출	주택관련_부채액	주거관련_부채의_이자	기타이자	기초연금	공적연금	재난지원금
0	275	0.0	0	7.0	0.0	500	360.0	0.0	10000	0.0	0	350.0	0	0	25
1	239	0.0	0	7.0	0.0	380	10.0	0.0	0	0.0	0	0.0	360	0	20
2	156	0.0	80	7.0	0.0	0	285.0	0.0	14000	20500.0	720	0.0	360	540	20
3	541	140.0	0	9.0	0.0	0	20.0	0.0	0	0.0	0	0.0	576	0	125
4	85	50.0	0	7.0	0.0	420	3.0	0.0	0	0.0	0	0.0	360	0	35

7865 rows × 15 columns

02 데이터 전처리

데이터형식 확인

```

총생활비          int64
이자_배당금_연간소득  float64
임대료_연간소득      int64
기타_소득          float64
부모가보낸생활비    float64
자녀가보낸생활비    int64
민간보조금          float64
세금환급금          float64
금융기관대출        int64
주택관련_부채액      float64
주거관련_부채의_이자  int64
기타이자            float64
기초연금             int64
공적연금             int64
재난지원금           int64
dtype: object

```

결측치 확인

```

총생활비          0
이자_배당금_연간소득    0
임대료_연간소득      0
기타_소득          0
부모가보낸생활비    0
자녀가보낸생활비    0
민간보조금          0
세금환급금          0
금융기관대출        0
주택관련_부채액      1079
주거관련_부채의_이자    0
기타이자            0
기초연금             0
공적연금             0
재난지원금           0
dtype: int64

```

02 데이터 전처리

결측치 대체 re_df1=df1.fillna(value=0)

```
총생활비          0
이자_배당금_연간소득    0
임대료_연간소득        0
기타_소득          0
부모가보낸생활비      0
자녀가보낸생활비      0
민간보조금          0
세금환급금          0
금융기관대출          0
주택관련_부채액        0
주거관련_부채의_이자    0
기타이자            0
기초연금            0
공적연금            0
재난지원금          0
dtype: int64
```

02 데이터 전처리

0 값 확인

```
for column in re_df1.columns:
    zero_count = re_df1.loc[re_df1[column] == 0].shape[0]
    number_count = re_df1.loc[re_df1[column] != 0].shape[0]
    print('0 개수 :',{column : zero_count})
    print('0 아닌 수 :',{column : number_count})
    print('-'*30)
```

0이 아닌 값들의 수를 확인하여 분석에 용이한 데이터인지 확인



분석에 활용 가능한 데이터로 판단하여 제외하는 변수 없이 진행

	변수	0 값인 수	0 값이 아닌 수
0	총생활비	0	7865
1	이자_배당금_연간소득	5489	2376
2	임대료_연간소득	6707	1158
3	기타_소득	1439	6426
4	부모가보낸생활비	6366	1499
5	자녀가보낸생활비	3296	4569
6	민간보조금	4019	3846
7	세금환급금	6275	1590
8	기초연금	4155	3710
9	공적연금	4777	3088
10	재난지원금	239	7626
11	금융기관대출	5500	2365
12	주택관련_부채액	6421	1444
13	주거관련_부채의_이자	6414	1451
14	기타이자	6598	1267

02 데이터 전처리

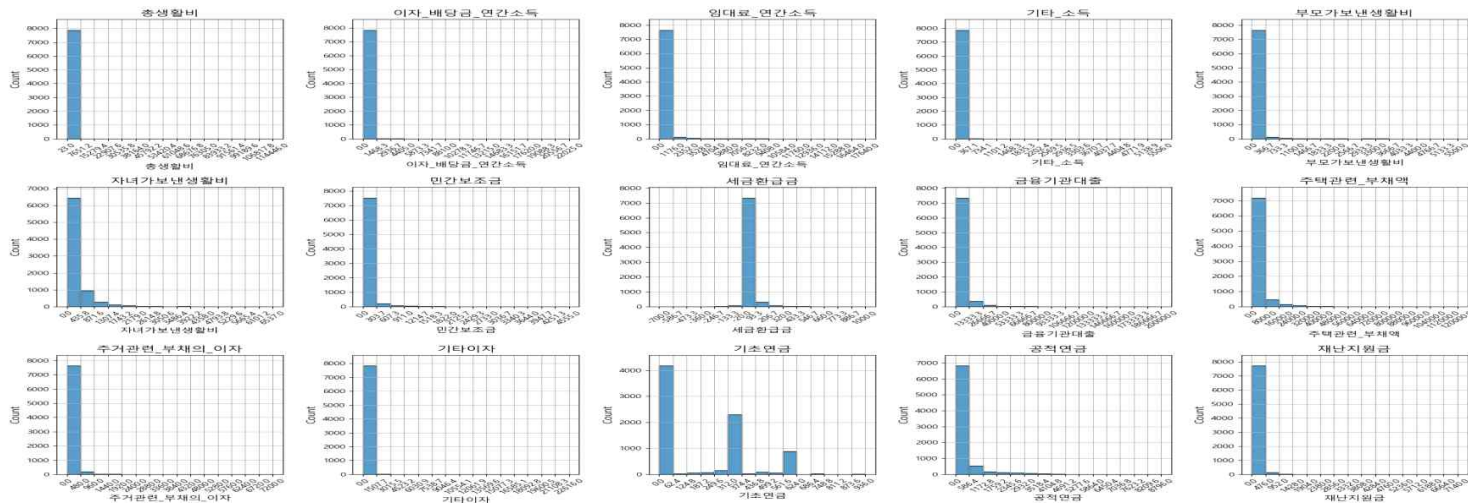
기초통계량 데이터의 통계적 속성 확인

	총생활비	이자_배당 금_연간소득	임대료_연 간소득	기타_소득	부모가 보낸생활비	자녀가 보낸생활비	민간보 조금	세금환 급금	금융기관대 출	주택관련_ 부채액	주거관 련_부채 의_이자	기타이자	기초연금	공적연금	재난지원금
count	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000
→ mean	329.822	55.538	129.336	8.456	45.949	230.697	58.166	14.506	2958.797	1773.448	53.392	47.885	193.396	288.639	94.908
std	1314.994	347.927	644.463	87.830	218.872	411.962	202.689	59.831	8277.347	5527.462	195.466	358.612	219.695	675.031	143.003
min	23.000	0.000	0.000	0.000	0.000	0.000	0.000	-700.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25%	121.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	35.000
→ 50%	228.000	0.000	0.000	7.000	0.000	70.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	60.000
75%	435.000	20.000	0.000	8.000	0.000	300.000	30.000	0.000	1700.000	0.000	0.000	0.000	360.000	288.000	105.000
→ max	114446.000	22025.000	17640.000	5506.000	5500.000	6537.000	4555.000	1000.000	200000.000	120000.000	7200.000	22616.000	936.000	8796.000	7140.000

max 값에 이상치로 판단될 수 있는 비정상적으로 높은 값이 있음을 확인 가능
모든 변수의 평균이 중앙값보다 크기 때문에 오른쪽으로 치우쳐 있을것으로 예상 가능

02 데이터 전처리

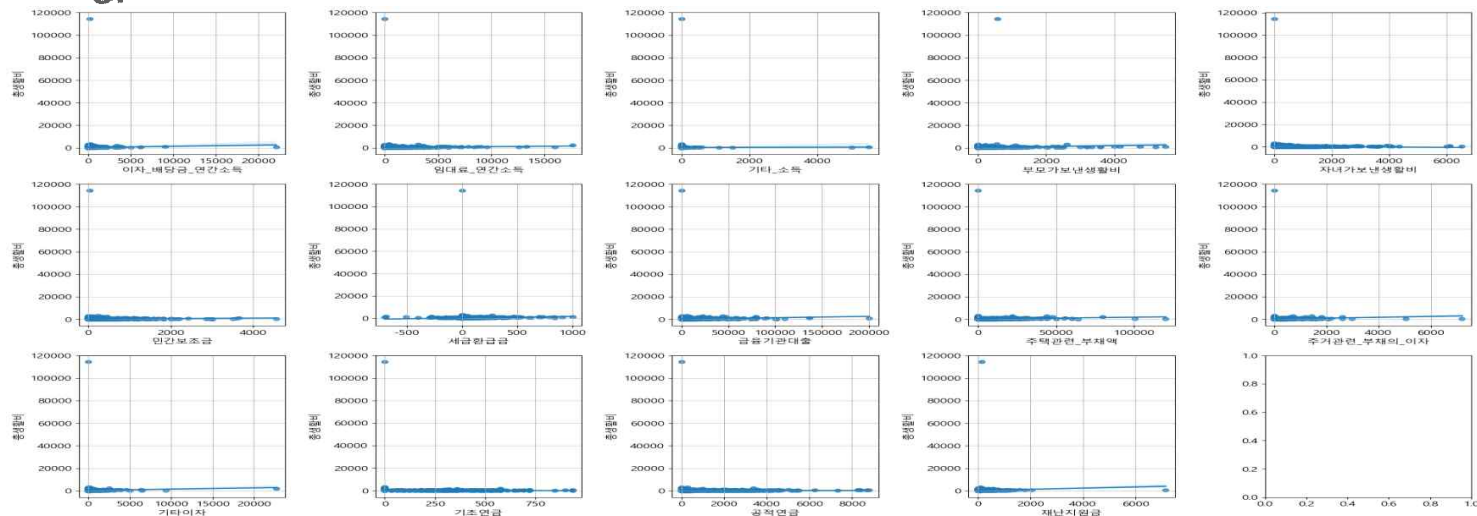
도수분포표 데이터의 분포와 패턴 파악



각 변수의 최대값으로 인해 도수분포가 비정상적인 형태를 띄고있음
특정 범위에 크게 치우쳐져 있을 것으로 예측 가능

02 데이터 전처리

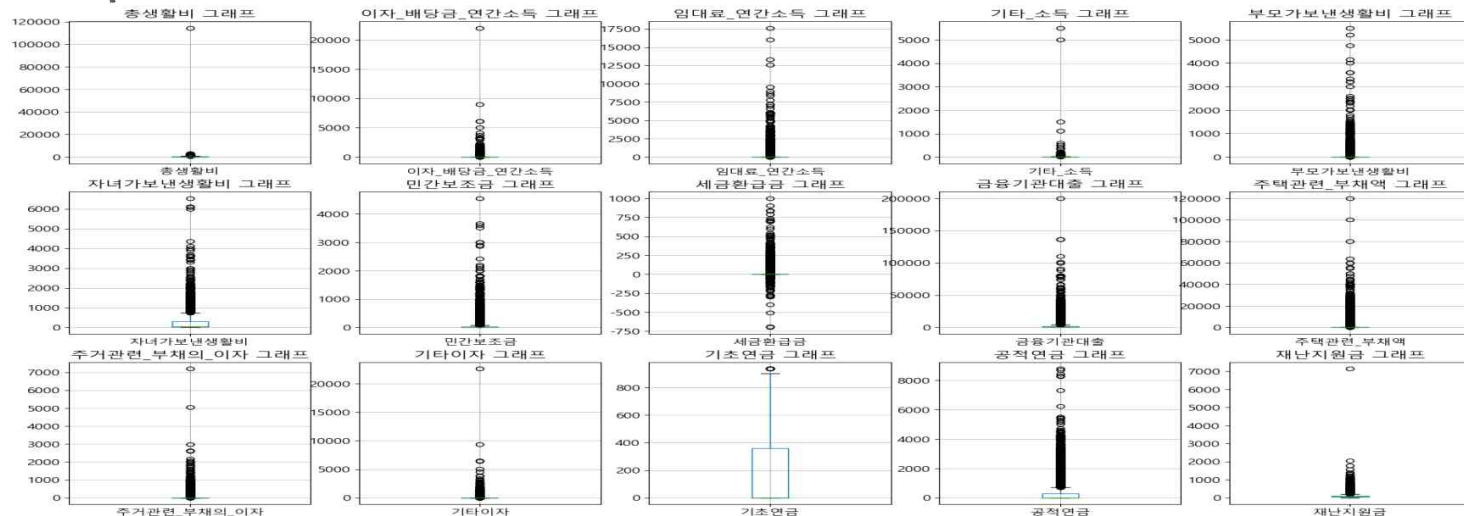
Regplot 산점도, 회귀선 시각화



종속변수인 '총생활비'의 최대값으로 인해 데이터의 분포 확인 어려움

02 데이터 전처리

Boxplot 데이터의 특성 파악



02 데이터 전처리

이상치 제거

총생활비: 100,000 이상, 이자 배당금 연간소득: 20,000 이상, 기타소득 : 4,900이상, 금융기관대출 : 190,000이상

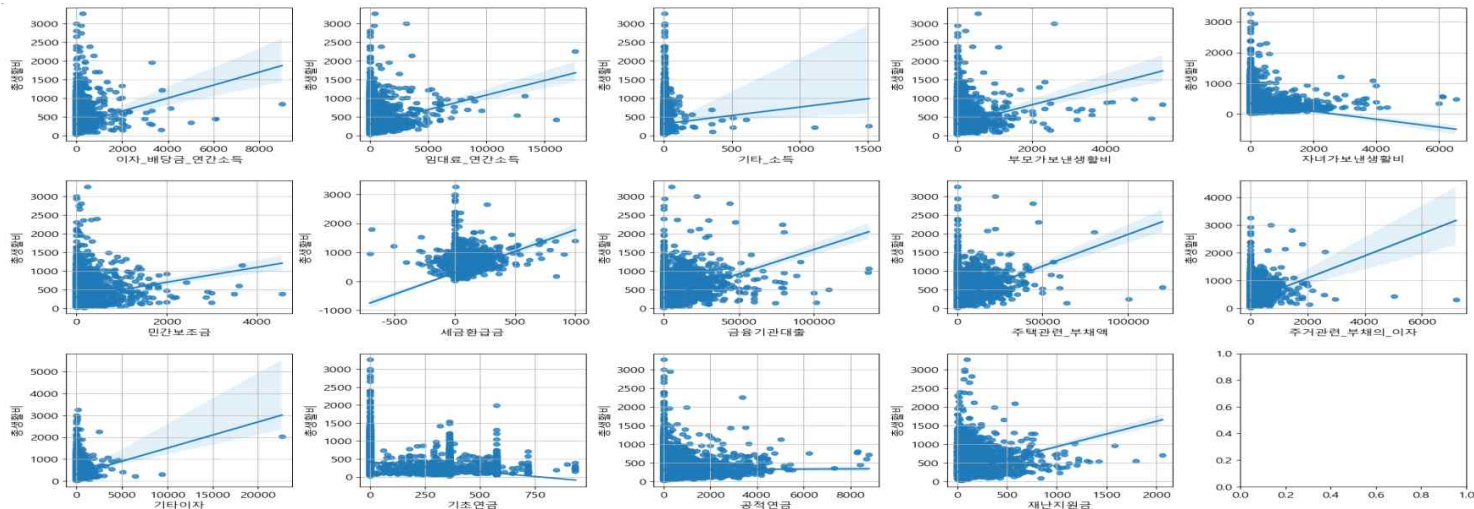
재난지원금 : 7,000이상의 값들을 이상치로 판단하고 제거

	총생 활비	이자_배 당금_연 간소득	임대료_연 간소득	기타_소 득	부모가 보낸생 활비	자녀가 보낸생 활비	민간보 조금	세금 환급 금	금융기 관대출	주택관련_부 채액	주거관 련_부채 의_이자	기타이 자	기초 연금	공적 연금	재난 지원금
0	275	0.0	0	7.0	0.0	500	360.0	0.0	10000	0.0	0	350.0	0	0	25
1	239	0.0	0	7.0	0.0	380	10.0	0.0	0	0.0	0	0.0	360	0	20
2	156	0.0	80	7.0	0.0	0	285.0	0.0	14000	20500.0	720	0.0	360	540	20
3	541	140.0	0	9.0	0.0	0	20.0	0.0	0	0.0	0	0.0	576	0	125
4	85	50.0	0	7.0	0.0	420	3.0	0.0	0	0.0	0	0.0	360	0	35
...
7860	384	0.0	0	1.0	50.0	0	0.0	70.0	3000	3000.0	96	0.0	0	0	100
7861	715	0.0	660	1.0	40.0	0	0.0	90.0	40000	0.0	0	1560.0	0	0	100
7862	595	0.0	4000	1.0	40.0	0	0.0	0.0	25947	25947.0	67	0.0	0	0	120
7863	562	40.0	0	0.0	0.0	0	0.0	0.0	12000	12000.0	360	0.0	0	0	30
7864	514	0.0	0	1.0	0.0	100	0.0	0.0	1880	0.0	0	22.0	0	0	50

7859 rows × 15 columns

02 데이터 전처리

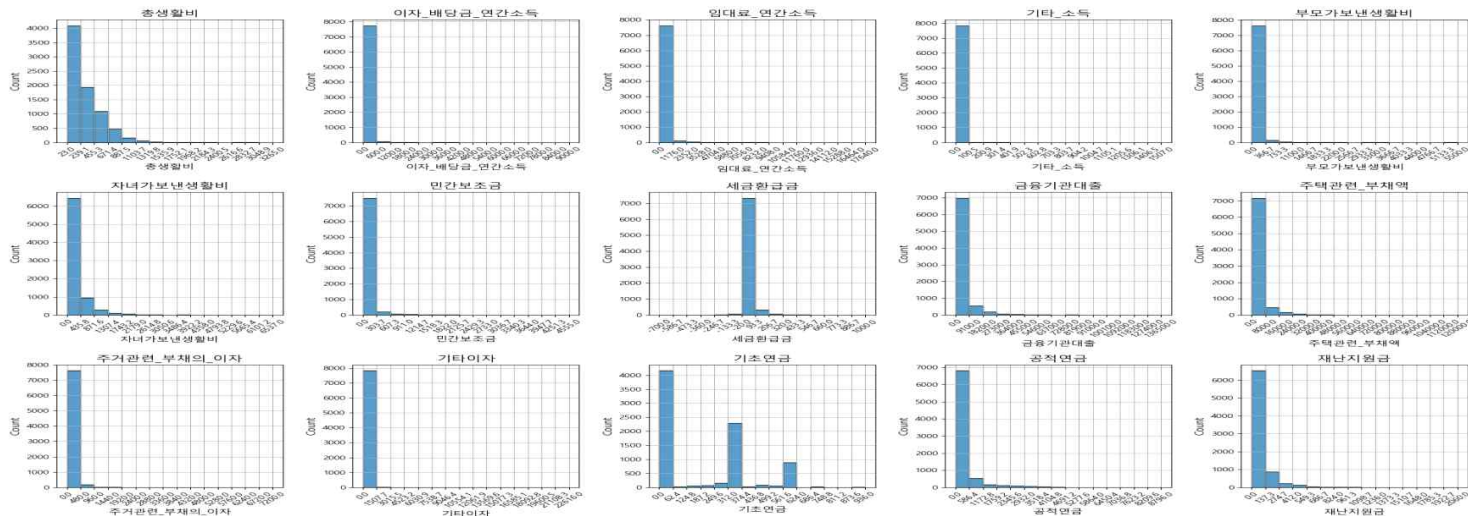
이상치 제거후 Regplot 확인



이상치 제거 후 데이터의 분포 확인 가능

02 데이터 전처리

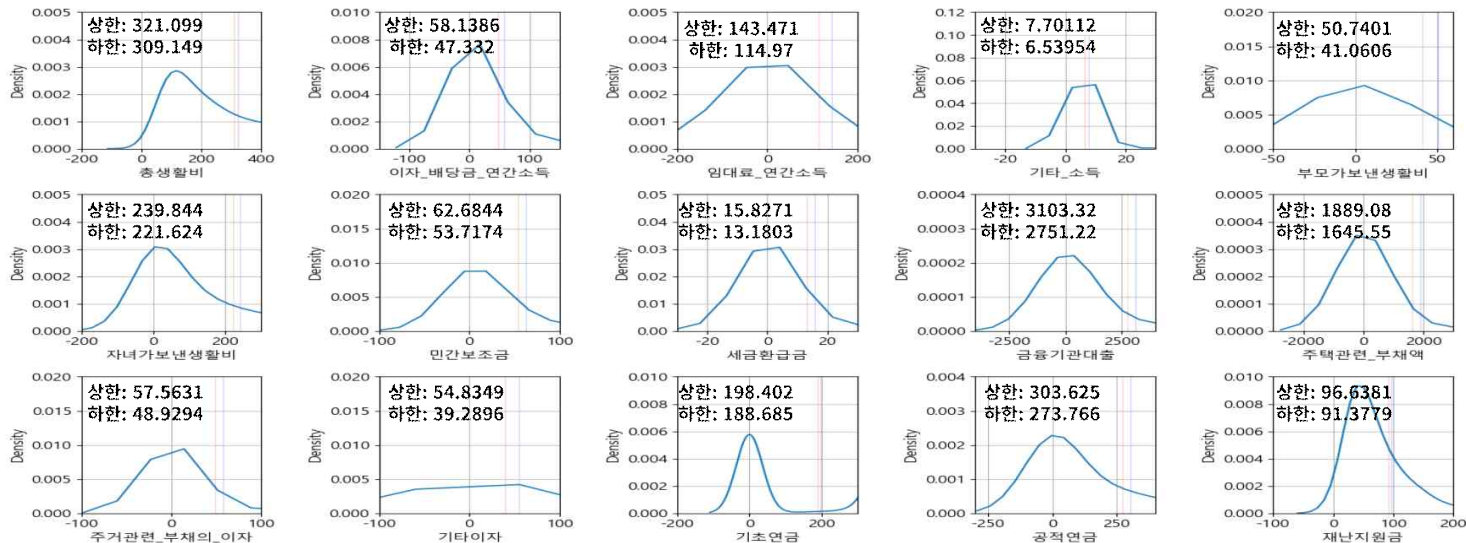
이상치 제거후 도수분포표 확인



이상치 제거 후 분포도가 개선된 모습
 총생활비의 경우 절반 이상의 가구가 월 239만원 이하에 분포되어있음

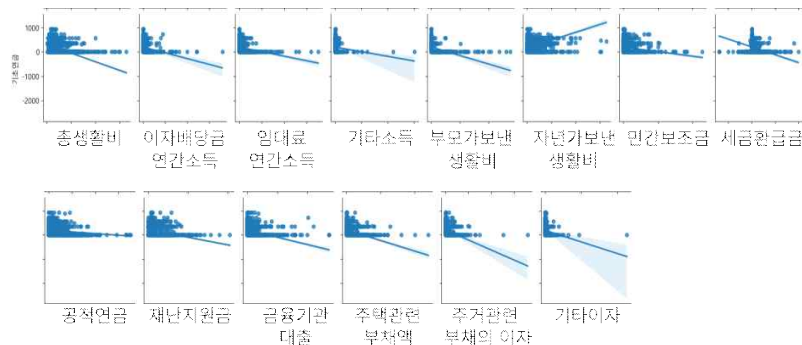
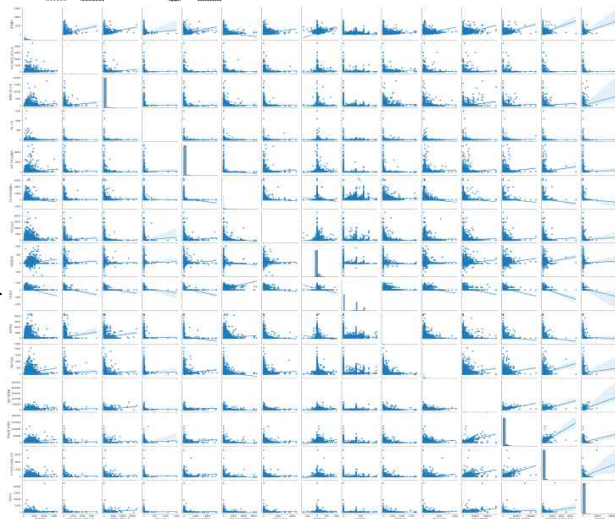
02 데이터 전처리

신뢰구간 표본데이터를 기반으로 추정할 수 있는 모집단의 평균 값의 구간



02 데이터 전처리

산점도 행렬 변수들의 관계와 패턴 파악



'기초연금'이라는 변수의 특성상 대부분의 변수들과
반비례함을 확인 가능

02 데이터 전처리

정규분포 가정 $|왜도| < 3, |첨도| < 8$ 또는 $|왜도| < 2, |첨도| < 4$

왜도 : 분포의 비대칭도

구분	내용
정규분포	첨도 = 0
위로 뽀족함	첨도 > 0
아래로 뽀족함	첨도 < 0

첨도 : 확률분포의 뽀족한정도

구분	내용
정규분포	왜도 = 0
왼쪽으로 치우침	왜도 > 0
오른쪽으로 치우침	왜도 < 0

변수명	정규분포 왜도	정규분포 첨도
0 이자_배당금_연간소득	16.080	398.172
1 임대료_연간소득	11.470	199.716
2 기타_소득	38.127	1832.707
3 부모가보낸생활비	12.077	204.786
4 자녀가보낸생활비	4.408	36.134
5 민간보조금	8.621	110.507
6 세금환급금	5.110	60.736
7 기초연금	0.517	-1.189
8 공적연금	4.467	28.361
9 재난지원금	4.887	38.836
10 금융기관대출	5.733	52.463
11 주택관련_부채액	5.941	64.295
12 주거관련_부채의_이자	12.302	313.154
13 기타이자	38.963	2256.060

'기초연금'을 제외한 나머지 변수들은
정규분포를 따르지 않음

각 변수의 첨도 값이 대체로 큰 것으로
보아 위쪽으로 뽀족한 분포를 보이는
것을 알 수 있음

03 상관 분석

스피어만 상관분석 & 결과검정

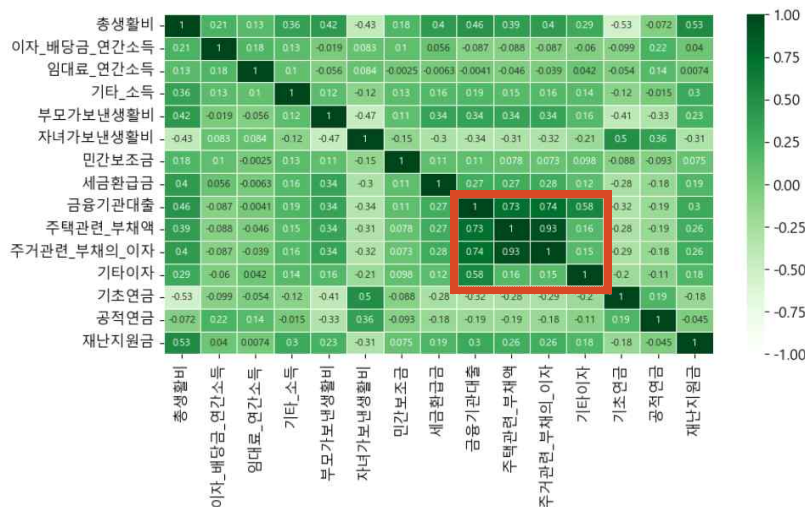
	statistic	pvalue	result
fields			
총생활비 vs 이자_배당금_연간소득	0.214765	1.136305e-82	True
이자_배당금_연간소득 vs 임대료_연간소득	0.175856	1.292815e-55	True
임대료_연간소득 vs 기타_소득	0.099732	7.854510e-19	True
기타_소득 vs 부모가보낸생활비	0.120537	7.929718e-27	True
부모가보낸생활비 vs 자녀가보낸생활비	-0.467324	0.000000e+00	True
자녀가보낸생활비 vs 민간보조금	-0.153996	6.558783e-43	True
민간보조금 vs 세금환급금	0.110892	6.234430e-23	True
세금환급금 vs 금융기관대출	0.266333	1.000648e-127	True
금융기관대출 vs 주택관련_부채액	0.730055	0.000000e+00	True
주택관련_부채액 vs 주거관련_부채의_이자	0.932065	0.000000e+00	True
주거관련_부채의_이자 vs 기타이자	0.153365	1.437161e-42	True
기타이자 vs 기초연금	-0.203579	2.677688e-74	True
기초연금 vs 공적연금	0.193484	3.711177e-67	True
공적연금 vs 재난지원금	-0.045009	6.565812e-05	True
재난지원금 vs 총생활비	0.529540	0.000000e+00	True

데이터가 정규분포를 따르지 않기 때문에
스피어만 상관분석 진행

상관분석의 검정 결과 모든 변수들 간의 p-value값이
0.05 이하로 **통계적으로 유의**한 것으로 나타남

03 상관 분석

상관분석 시각화(Heatmap) 변수들의 상관관계 파악



'금융기관대출', '주택관련 부채',
'주거관련 부채의 이자' 변수들의 강한
양의 상관관계 확인

서로 영향을 미치는 변수들 제외 후
다시 스피어만 상관분석 진행

03 상관분석

상관분석 시각화(Heatmap) 변수들의 상관관계 파악



'주택관련 부채액', '주거관련 부채의 이자'를
제외한 후 스피어만 상관분석 진행

'금융기관대출'도 높은 상관관계를 보였지만
'주택관련 부채액', '주거관련 부채의 이자'를
제외하였더니 상관관계수가 감소하여
제외하지 않음

03 상관분석

회귀분석 결과보고 서로 영향을 미치는 변수들 제외한 후 회귀분석 진행

종속변수	독립변수	비표준화 계수		표준화 계수	t	유의확률	공선성 통계량
		B	표준오차	베타(β)			VIF
총생활비	민간보조금	0.0861	0.009	0	9.703*	0	1.02891
	이자_배당금_연간소득	0.0901	0.009	0	10.112*	0	1.03794
	기초연금	-0.2715	0.01	0	-27.753*	0	1.35677
	기타이자	0.0145	0.01	0	1.474*	0.14	1.24154
	공적연금	0.0141	0.009	0	1.569*	0.117	1.04843
	기타_소득	0.0154	0.009	0	1.744*	0.081	1.00368
	금융기관대출	0.2025	0.01	0	19.370*	0	1.47918
	부모가보낸생활비	0.0967	0.009	0	10.699*	0	1.07123
	재난지원금	0.1833	0.009	0	20.164*	0	1.12183
	세금환급금	0.2112	0.009	0	23.159*	0	1.14699
	자녀가보낸생활비	-0.0258	0.009	0	-2.758*	0.006	1.12745
	임대료_연간소득	0.0852	0.009	0	9.276*	0	1.10112
R(0.392), R ² (0.391), F(422.2), 유의확률(0.00), Durbin-Watson(1.876) (*은 0.05 미만을 의미)							

분석결과 회귀모형은 통계적으로
유의미한 것으로 나타났으며

(F= 422.2, p<0.05)

설명력은 39.2%(수정된 R제곱은
39.1%)로 나타남

Durbin-Watson값은 1.876으로 2에 근
사한 값을 보여 잔차의 독립성 가정

모든 변수들의 VIF값은 10이하로 다중공
선성의 문제 없음

04 주성분 분석 (PCA)

주성분 분석(PCA) 데이터의 특성을 보존하면서 데이터의 변수 수를 줄이는 방법

표준화 진행

	총생활비	이자_배당 금_연간소득	임대료_연 간소득	기타_소득	부모가 보낸생활비	자녀가 보낸생활비	민간보 조금	세금환 급금	금융기관대 출	주택관련_ 부채액	주거관 련_부채 의_이자	기타이자	기초연금	공적연금	재난지원금
count	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000	7865.000
mean	329.822	55.538	129.336	8.456	45.949	230.697	58.166	14.506	2958.797	1773.448	53.392	47.885	193.396	288.639	94.908
std	1314.994	347.927	644.463	87.830	218.872	411.962	202.689	59.831	8277.347	5527.462	195.466	358.612	219.695	675.031	143.003
min	23.000	0.000	0.000	0.000	0.000	0.000	0.000	-700.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
25%	121.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	35.000
50%	228.000	0.000	0.000	7.000	0.000	70.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	60.000
75%	435.000	20.000	0.000	8.000	0.000	300.000	30.000	0.000	1700.000	0.000	0.000	0.000	360.000	288.000	105.000
max	114446.000	22025.000	17640.000	5506.000	5500.000	6537.000	4555.000	1000.000	200000.000	120000.000	7200.000	22616.000	936.000	8796.000	7140.000

변수들의 크기 차이가 클 경우 변수의 중요도가
잘못 계산될 수 있고, 결과적으로 주성분 분석의 해석이
어려워질 수 있기 때문에 표준화 과정 필요

```
model = pca()
fit = model.fit_transform(x_train_std_df)
topfeat = fit['topfeat']
topfeat
```

04 주성분 분석 (PCA)

주성분 분석(PCA)

	PC	feature	loading	type
0	PC1	금융기관대출	0.491171	best
1	PC2	자녀가보낸생활비	0.497598	best
2	PC3	공적연금	0.567770	best
3	PC4	기타_소득	0.758736	best
4	PC5	세금환급금	-0.454537	best
5	PC6	민간보조금	-0.635869	best
6	PC7	재난지원금	-0.509041	best
7	PC8	공적연금	-0.540182	best
8	PC9	부모가보낸생활비	0.771742	best
9	PC10	이자_배당금_연간소득	0.454611	best
10	PC11	기타이자	0.599527	best
11	PC12	기초연금	-0.762436	best
12	PC3	임대료_연간소득	0.530654	weak
13	PC1	주택관련_부채액	0.483626	weak
14	PC1	주거관련_부채의_이자	0.464062	weak

임대료 연간소득, 주택관련 부채액, 주거관련 부채의 이자가 설명력이 낮은 변수로 나타남

설명력이 낮은 세가지 변수 모두 제외한 후 회귀분석 진행

04 주성분 분석 (PCA)

회귀분석 결과보고 주성분 분석(PCA)을 통해 설명력이 약한 변수들 모두 제외한 후 회귀분석 진행

종속변수	독립변수	비표준화 계수		표준화 계수	t	유의확률	공선성 통계량
		B	표준오차	베타(β)			VIF
총생활비	민간보조금	0.0844	0.009	0	9.459*	0	1.02785
	이자_배당금_연간소득	0.0956	0.009	0	10.694*	0	1.03494
	기초연금	-0.277	0.01	0	-28.217*	0	1.35581
	기타이자	0.0192	0.01	0	1.948*	0.051	1.23846
	공적연금	0.0213	0.009	0	2.366*	0.018	1.04103
	기타_소득	0.0157	0.009	0	1.766*	0.077	1.00368
	금융기관대출	0.2212	0.01	0	21.439*	0	1.43914
	부모가보낸생활비	0.0943	0.009	0	10.379*	0	1.06943
	재난지원금	0.183	0.009	0	20.024*	0	1.12105
	세금환급금	0.2092	0.009	0	22.827*	0	1.14429
	자녀가보낸생활비	-0.0228	0.009	0	-2.433*	0.015	1.12592
R(0.386), R^2 (0.385), F(448.0), 유의확률(0.00), Durbin-Watson(1.869) (*은 0.05 미만을 의미)							

분석결과 회귀모형은 통계적으로
유의미한 것으로 나타났으며

($F = 448.0$, $p < 0.05$)

설명력은 38.6%

(수정된 R제곱은 38.5%)로 나타남

Durbin-Watson값은 1.869으로 2에
근사한 값을 보여 잔차의 독립성 가정

모든 변수들의 VIF값은 10이하로
다중공선성의 문제 발생하지 않음

05 회귀분석/결과보고

모든 경우의 수 회귀분석 진행 주성분 분석(PCA)을 통해 발견한 설명력 약한 변수들을 제외할 수 있는 모든 경우의수에 대하여 회귀분석 진행

변수 설정 및 결과	
Case1	weak로 나온 3개의 독립변수를 처리하지 않고 모든 변수를 사용한 MLR 분석
	R(0.397), $R^2(0.396)$, F(369.5), 유의확률(0.00), Durbin-Watson(1.880)
Case2	weak 중 하나인 임대료_연간소득만 제외 후 MLR 분석
	R(0.389), $R^2(0.388)$, F(384.4), 유의확률(0.00), Durbin-Watson(1.872)
Case3	weak 중 하나인 주택관련_부채액만 제외 후 MLR 분석
	R(0.394), $R^2(0.393)$, F(392.3), 유의확률(0.00), Durbin-Watson(1.876)
Case4	weak 중 하나인 주거관련_부채의_이자만 제외 후 MLR 분석
	R(0.397), $R^2(0.396)$, F(397.9), 유의확률(0.00), Durbin-Watson(1.880)
Case5	weak 중 임대료_연간소득 및 주택관련_부채액 2가지 weak변수 제외 후 MLR 분석
	R(0.394), $R^2(0.386)$, F(412.3), 유의확률(0.00), Durbin-Watson(1.869)
Case6	weak 중 임대료_연간소득 및 주거관련_부채의_이자 2가지 weak변수 제외 후 MLR 분석
	R(0.389), $R^2(0.388)$, F(416.5), 유의확률(0.00), Durbin-Watson(1.872)
Case7	weak 중 주택관련_부채액 및 주거관련_부채의_이자 2가지 weak변수 제외 후 MLR 분석
	R(0.392), $R^2(0.391)$, F(422.2), 유의확률(0.00), Durbin-Watson(1.876)
Case8	weak 변수 3가지 전부 제외 후 MLR 분석 시행
	R(0.386), $R^2(0.385)$, F(448.0), 유의확률(0.00), Durbin-Watson(1.869)

모든 경우의 수를 가정하여 회귀분석을 진행한 결과 독립변수들을 모두 포함하여 회귀분석을 진행한 **Case1**에서 가장 높은 설명력이 나타남

스피어만 상관분석을 통해 상관관계가 있는 변수들을 발견했지만 VIF값이 정상범위를 나타내므로 모든 변수들 포함하는
Case 1 채택

05 회귀분석/결과보고

Case 1 회귀분석 결과보고

종속변수	독립변수	비표준화 계수		표준화 계수	t	유의확률	공선성 통계량
		B	표준오차	베타(β)			VIF
총생활비	민간보조금	0.0856	0.009	0	9.682*	0	1.02894
	이자_배당금_연간소득	0.0917	0.009	0	10.333*	0	1.03931
	기초연금	-0.2649	0.01	0	-27.099*	0	1.36073
	기타이자	0.0277	0.01	0	2.755*	0.006	1.3171
	공적연금	0.017	0.009	0	1.888*	0.059	1.05018
	기타_소득	0.0133	0.009	0	1.517*	0.129	1.00487
	금융기관대출	0.1322	0.014	0	9.705*	0	2.44304
	부모가보낸생활비	0.0927	0.009	0	10.269*	0	1.07398
	재난지원금	0.1813	0.009	0	19.978*	0	1.12695
	세금환급금	0.2052	0.009	0	22.515*	0	1.15123
	자녀가보낸생활비	-0.0255	0.009	0	-2.744*	0.006	1.12748
	임대료_연간소득	0.0957	0.009	0	10.359*	0	1.12699
	주택관련_부채액	0.098	0.015	0	6.676*	0	2.82219
	주거관련_부채의_이자	0.0006	0.013	0	0.050*	0.96	2.15651
R(0.397), R^2 (0.396), F(369.5), 유의확률(0.00), Durbin-Watson(1.880) (*은 0.05 미만을 의미)							

분석결과 회귀모형은 통계적으로
유의미한 것으로 나타났으며

($F=369.5$, $p<0.05$)

설명력은 39.7%

(수정된 R제곱은 39.6%)로 나타남

Durbin-Watson값은 1.880으로 2에
근사한 값을 보여 잔차의 독립성 가정

모든 변수들의 VIF값은 10이하로
다중공선성의 문제 발생하지 않음

05 회귀분석/결과보고

1. 기초연금 회귀계수: -0.2649

기초연금은 총 생활비와 반비례 하는 것으로 보임.
만 65세 이상, 일정소득 이하 가구에 지급하는 기초연금의 특성상 기초연금을 받는 가구 일수록 소득이 적을것이며
그렇기 때문에 생활비 또한 낮은 것으로 예측가능

2. 세금환급금 회귀계수: 0.2052

일반적으로 소비가 많은 경우에 세액공제 혜택을 많이
받을 수 있으므로 세금환급금은 비례하는 것으로 볼 수 있
음.

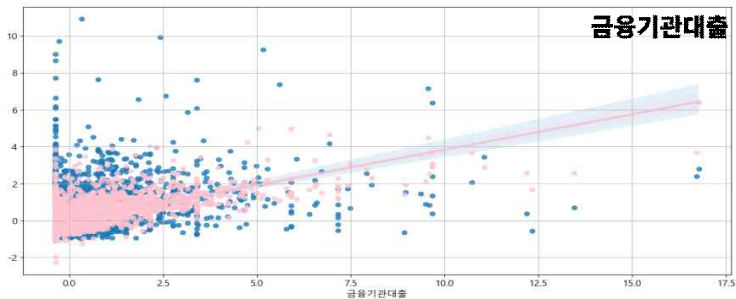
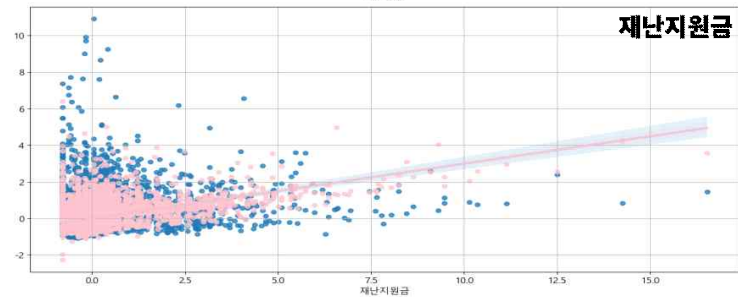
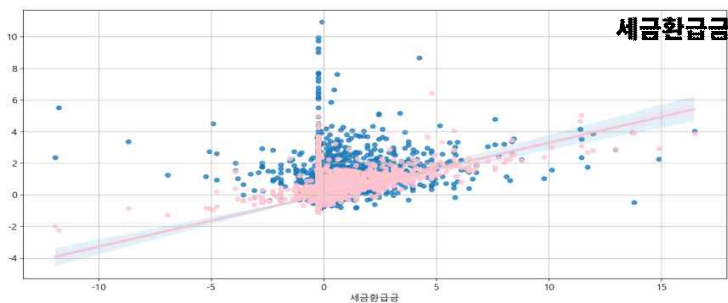
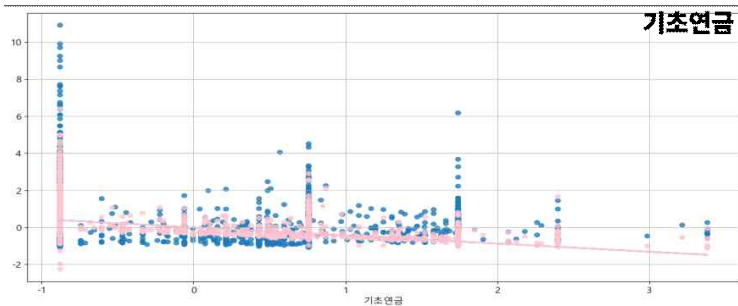
3. 재난지원금 회귀계수: 0.1813

재난지원금은 코로나19 시기 생계안정과 소비촉진을 위해
지급된 지원금으로 지급액이 많은 가구일수록 총생활비가
증가한 것으로 볼 수 있음. 수급자들의 혜택에 따른 소비량
의 상승으로 예측할 수 있음.

4. 금융기관대출 회귀계수: 0.1322

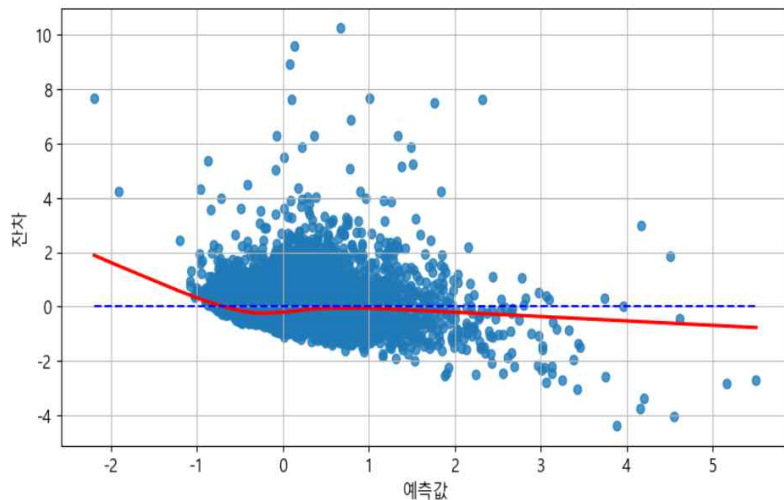
금융기관대출은 이자와 원금을 상환할 수 있다는 경제적
능력에 대한 판단 후 진행되기 때문에 높은 금융기관대출
액은 높은 자본력과 수입의 크기를 나타낸다고 볼 수 있음.
이와 같은 이유로 금융기관대출과 총생활비는 비례하는 것
으로 보여짐.

05 회귀분석/결과보고



06 잔차 분석

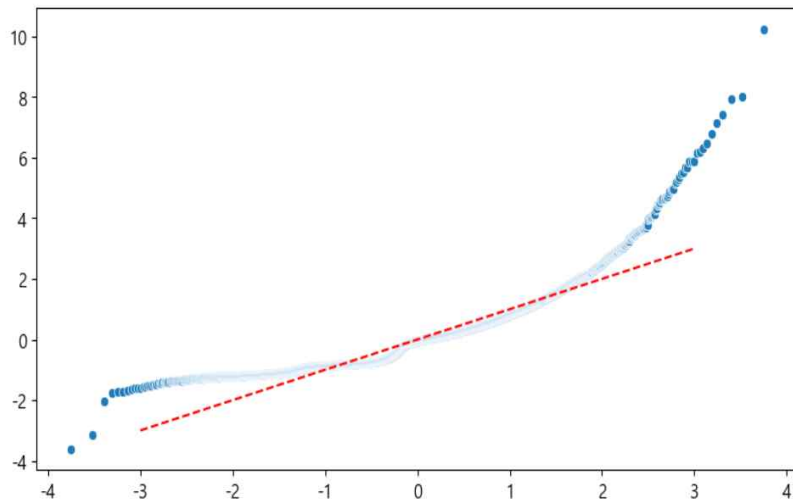
잔차의 선형성 확인



잔차의 추세선이 점선에서 크게 벗어나지 않는 것으로 보아 잔차의 선형성을 확인할 수 있음

06 잔차 분석

잔차의 정규성 확인 (Q-Q plot)



특정 구간에서 잔차값이 정규분포를 벗어난 것을 확인할 수 있음

오른쪽 끝이 길게 뻗어져있는 것으로 보아 평균보다 큰 값이 많음을 예측할 수 있음