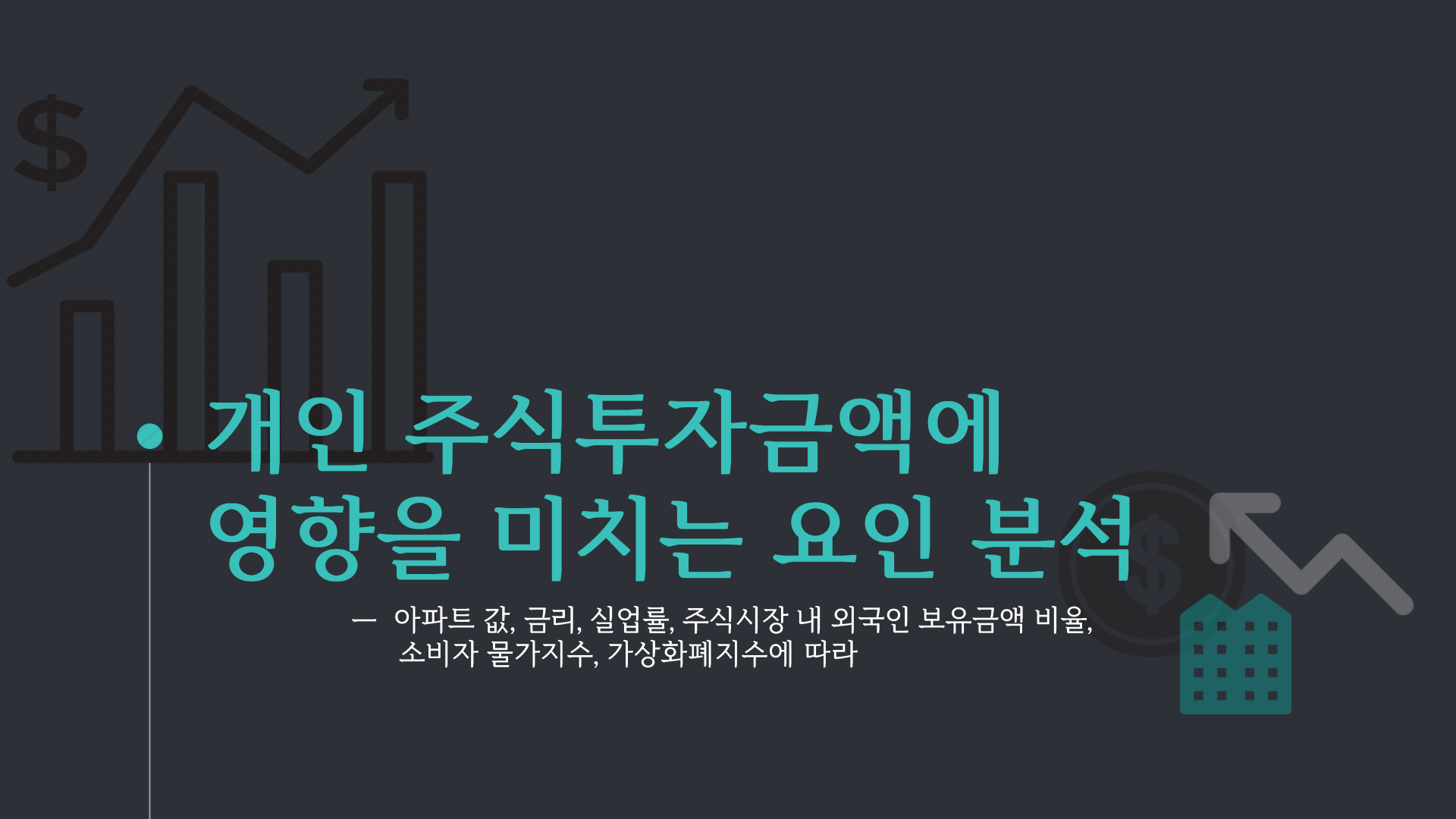


# R 개미

김은영  
박수민  
이성수  
이은지



# • 개인 주식투자금액에 영향을 미치는 요인 분석

- 아파트 값, 금리, 실업률, 주식시장 내 외국인 보유금액 비율,  
소비자 물가지수, 가상화폐지수에 따라

## ● INDEX

주제선정 이유



데이터 수집



데이터 분석



변수 설명



데이터 전처리



정리 및 예측



1

주제 선정 이유

## 주식·코인의 늪 빠진 2030 '영끌 빚투'

주식·가상화폐 등 자산시장 주역된 MZ세대 자산형성 높은 벽에 대출받아 '인생역전' 노려 다중채무자... 따른 집값 급등이 젊은층의 좌절이나 위험자산 선호 현상...

## 주식·코인에 빠진 대학생들... "인생 바꿀 수 있는 사다리가 없기에"

장기간 지속되고 있는 취업난에 최근 몇년간 아파트를 포함한 집값마저 천정부지로 치솟으면서 대학생들이 단기간에 큰 돈을 벌겠다는 기대심리에 주식과 코인 ...

## '내겐 너무 비싼 집값' 분노의 MZ세대, 주식·코인 '올인'

주식과 가상화폐  
흔까지 끌어모으

X세대보다 삶 만족도 높지만 뛰는 집값에 좌절... 행복은 더 멀리

대학생 지모(24)씨는 "친구들끼리 모이면 결국 담은 코인이나 주식이라는 얘기를 많이 한다"고 말했다. 단초를 부동산에서 찾아볼 수 있다. 통계청 가계동향조사에...

## 2030세대 10명중 9명, 주식에 쫓혔다

지난해 불어닥친 '동학 개미 운동' 이후 개인 투자자의 주식 열풍이 거세진 데다 초저금리와 집값 급등 속에서 주식 투자에라도 뛰어들지 않으면 '벼락 거지'가 될...

“

정말 아파트 값이  
오를수록 사람들은  
주식에 뛰어들까?

## ● 가설

# ○ 아파트 값은 개인의 주식 투자금액에 영향을 줄 것이다

- 그 외에 영향을 주는 다른 요인들이 또 있을까?



2

## 변수 설명



## ● 변수 설명

### 종속변수

- ❖ stock : 개인 주식 투자 금액 지수

### 독립변수

- ❖ house : 전국 아파트 실거래가 지수
- ❖ y\_job : 청년 실업률
- ❖ job : 전체 실업률
- ❖ moneyrate : 기준금리
- ❖ coin : 가상화폐 투자 공포 탐욕 지수
- ❖ foreigner : 주식시장 내 외국인 보유금액 비율
- ❖ customerprice : 소비자 물가지수

\* 주

stock(2017.11기준) / house(2017.11기준) / customerprice(2015년 기준)

기간 : 201710 ~ 202102 총 41개월의 데이터

3

## 데이터 수집

## ● 데이터 수집 : 출처

E- 나라지표

한국 부동산원

국가 통계 포털

한국은행 경제통계시스템

두나무 데이터밸류 연구소

# 데이터 수집 : STOCK(개인 주식 투자 금액 지수)

**KOSIS** 국가통계포털  
Korean Statistical Information Service

증권·파생상품시장통계		시점 2021	내려받기 및 인쇄	English
보고목적	한국거래소 증권·파생상품시장의 종합적인 동향을 파악하여 건전한 투자유도, 각종 시책의 수립평가 등을 위한 기초 자료로 제공			
조사주기	월			
자료보고방법	항정집계			
보고체계	한국거래소 통계팀에서 각 시장별 통계자료를 취합하여 작성			
통계종류	일반통계			

**KOSIS** 통계청

KRX 상장주식 투자자별 거래실적

통계설명자료 온라인간행물

출처 : 한국거래소, 「증권·파생상품시장통계」 [단보기]

자료갱신일: 2021-05-07 / 수록기간: 월, 년 2001 ~ 2021.03 / 자료문의처 : 02-3774-8937 / 기능문의: KOSIS Q&A 게시판

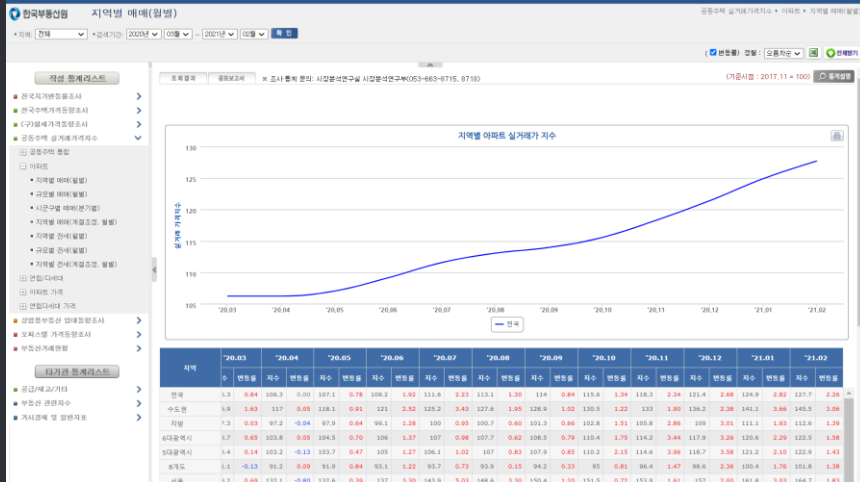
일괄설정 +

할목 [1/1] 거래구분별 [2/2] 투자자별 [104/104] 시점 [6/215]

주식정보 주주정보 발행전환 분석 차트 부가기능설정 ☆스크랩 다운로드 인쇄 도움말

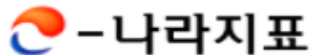
거래구분별(1)	투자자별(1)	투자자별(2)	2021. 03	2021. 02	2021. 01	2020. 12	2020. 11	2020. 10
거래량	총거래량	매도 (천주)	81,763,038	82,455,515	67,951,529	67,951,569	66,183,093	52
		매수 (천주)	81,763,038	82,455,515	67,951,529	67,951,569	66,183,093	52
		순매수 (천주)	0	0	0	0	0	
	금융투자	매도 (천주)	416,654	462,867	738,154	494,982	477,249	
		매수 (천주)	392,938	451,056	657,693	538,094	467,571	
		순매수 (천주)	-23,715	-11,811	-80,461	43,113	-9,678	
	보험	매도 (천주)	46,506	38,505	83,958	59,902	56,551	
		매수 (천주)	47,620	35,856	44,723	47,492	43,260	
		순매수 (천주)	1,114	-2,650	-38,236	-12,410	-13,291	
투자신탁	매도 (천주)		290,362	250,510	283,850	287,190	213,064	

# 데이터 수집 : HOUSE(전국 아파트 실거래가 지수)



통계명	공동주택 실거래가지수
최초작성연도	(매매) 2006년, (전세) 2014년
통계종류	일반 가동통계
발적근거	<ul style="list-style-type: none"> <li>승인번호 : 제11672호</li> <li>승인일자 : 2009년 12월 21일</li> <li>주택법 제88조, 제89조 및 동법 시행령 제11조</li> </ul>
작성목적	전국의 재고 아파트 및 연립·다세대용 대상으로 하여 실제 거래되어 신고된 아파트 및 연립·다세대용 거래 가격기준 및 변동률을 파악하여 정확한 시장동향 정보를 국민에게 제공하고 정부 정책 수립에 참고 자료로 활용
작성주기	매월

# 데이터 수집 : JOB, Y\_JOB(전체 실업률, 청년 실업률)



## 지표설명

### 고용지표 개념

○ 조사대상월 15일 현재 만 15세 이상 인구 중 경제활동인구를 대상으로 고용지표를 조사

※ **취업자** : 가. 조사대상주간에 수업을 목적으로 1시간 이상 일한 자

나. 동업가구나 가족원이 운영하는 농장이나 사업체의 수업을 위해 주당 18시간이상 일한 무급가족종사자

다. 직업 또는 사업체를 가지고 있으나 일시적인 병 또는 사고, 연가, 교육, 노사분규 등의 사유로 일하지 못한 일시 휴직자

**실업자** : 조사대상주간에 수입있는 일을 하지 않았고, 지난 4주간 일자리를 찾아 적극적으로 구직활동을 하였던 사람으로서 일자리가 주어지면 즉시 취업이 가능한 사람

**실업률** :  $\text{실업률}(\%) = (\text{실업자} / \text{경제활동인구}) \times 100$

**청년실업률** :  $\text{청년실업률}(\%) = (15 \sim 29 \text{세 실업자} / 15 \sim 29 \text{세 경제활동인구}) \times 100$

### 지표의의 및 활용도

○ 국민의 경제활동 속, 국민의 취업, 실업 등과 같은 특성을 조사하여 거시경제 분석과 인력자원의 개발정책 수립에 필요한 기초 자료를 제공

## 취업자 수/실업률 추이

☆ 14

· 지표 담당 : 기획재정부 경제기획관 044-215-2818

· 최신 갱신일 : 2021-04-28 (입력 예정일 : 2021-05-14)

그래프/통계표

의미분석

지표정보

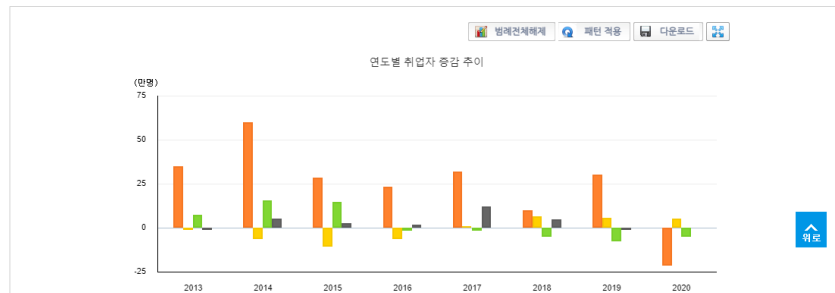
정책정보

의견 및 질문

## 그래프/통계표

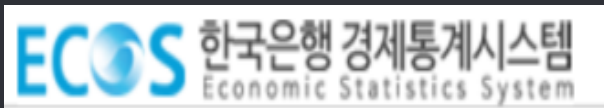
### ○ 그래프

- 연도별 취업자 증감 추이



위로

## 데이터 수집 : MONEYRATE(기준금리)



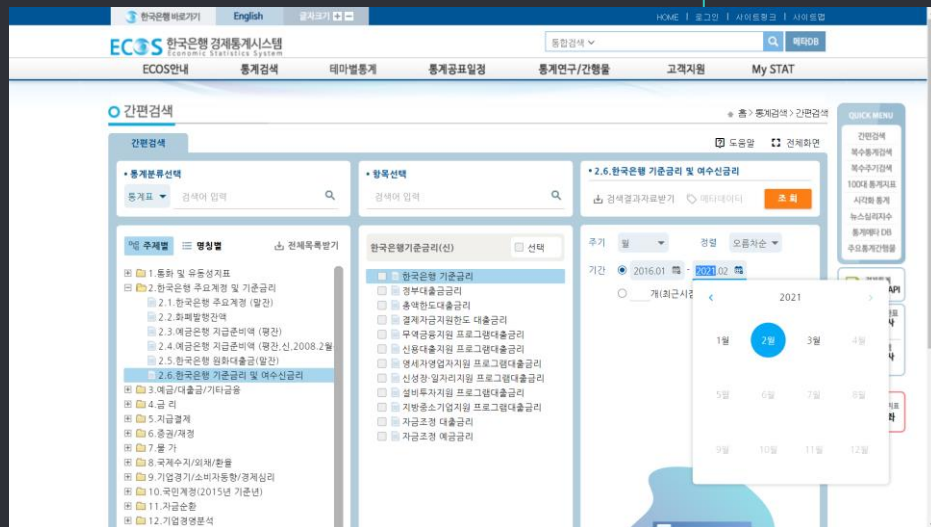
### 2.6. 한국은행 기준금리 및 여수신금리

통계담당 : 한국은행 통화정책국 정책총괄팀(02-759-4679)

주 : 1) 한국은행 기준금리의 경우 2008년 2월까지의 월평균치, 2008.3월부터는 한국은행 기준금리

2) 숙박한도대출금리의 경우 2013.4.11일 이후 지원부분별도 대출금리 차등화

3) 2020년 통화정책방향 결정회의 개최 예정일: 1.17(금), 2.27(목), 4.9(목), 5.28(목), 7.16(목), 8.27(목), 10.14(수), 11.26(목)



## ● 데이터 수집 : COIN(가상화폐 투자 공포 탐욕 지수)

Dunamu DataValue



### 디지털자산 공포-탐욕 지수

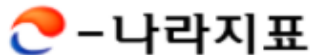
#### Methodology Book

version 1.0

Digital Asset Fear and Greed Index(디지털자산 공포-탐욕 지수)는 업비트 거래소의 암호화폐 가격과 거래량 지표를 가공하여 암호화폐 시장의 심리상황 파악, 효율적 가격 식별 등의 목적을 위해 개발되었다. 당일 디지털자산 공포-탐욕 지수는 현재 시점을 기준으로 24시간 동안의 데이터를 수집하여 5분마다 계산한다. 공포지수의 기준 날짜는 업비트 거래소와 동일하게 9:00 KST 를 기준으로 변경된다.



# 데이터 수집 : FOREIGNER(주식시장 내 외국인 보유금액 비율)



외국인 증권투자 현황 ☆ 11

자료 담당 : 금융위원회, 금융시장분석과, 02-2100-2856  
최근 갱신일 : 2021-04-16 (일렉 매핑일 : 2021-05-31)

그래프/통계표   의미분석   지표정보   정책정보   의견 및 질문

통계표  
~ 외국인 증권투자 현황

[단위 : 조원, %]

		2013	2014	2015	2016	2017	2018	2019	2020	2020 10월	2020 11월	2020 12월	2021 01월	2021 02월	2021 03월
보유금액	외국인 보유금액	430.6	422.4	420.5	480.8	634.7	509.7	593.2	764.3	584.8	675.2	764.3	787.9	797.5	805.3
	유가증권시장	418.8	406.3	401.0	461.0	597.0	480.6	563.5	722.9	555.2	642.0	723.1	749.5	760.1	766.5
	코스닥시장	11.8	16.0	20.0	20.0	37.0	25.4	25.4	38.6	29.6	33.5	38.4	37.6	36.5	38.2
시가총액대비(%)	외국인 보유금액	33.0	31.6	29.1	31.8	33.6	31.3	33.3	31.4	30.4	30.8	31.4	31.6	31.6	31.3
	유가증권시장	35.3	34.1	32.2	35.2	37.2	35.8	38.2	36.5	35.7	35.9	36.5	36.5	36.6	35.9
	코스닥시장	9.9	11.2	9.9	10.1	13.3	11.1	10.5	10.0	9.6	9.6	10.0	10.1	9.8	9.8

출처 : 금융감독원 '외국인 투자자 증권매매 동향', 한국거래소 '시장정보' 통계(주식) 투자참고 외국인보유량 추이

통계표 목록  
기본통계표: 외국인 증권투자 현황 (연, 2001 ~ 2020)

## 지표정보

### 작성방법

투자등록한 외국인의 보유주식의 조사

### 자료출처

출처 : '외국인 투자자의 증권매매 동향'(매 익월 중순), 한국거래소  
통계생산기관 : 금융감독원-금융투자 서비스국, 한국거래소  
통계주기 : 매월

## 데이터 수집 : CUSTOMERPRICE(소비자 물가지수)

**KOSIS** 국가통계포털  
Korean Statistical Information Service

조사목적	상품과 서비스의 가격을 조사하여 소비자물가지수를 작성하고 그 결과를 정부 재정, 금융정책의 기초자료로 이용하고, 가격수지, 국민소득계정 등 다른 경제지표의 디플레이터로 사용
조사주기	월
자료수집방법	면접조사
조사체계	조사대상처 → 지방통계청 및 사무소 → 통계청
통계종류	지정통계

### 1) 지출목적별 소비자물가지수(품목 포함, 2015=100)

자료갱신일: 2021-05-04 / 수록기간: 월, 분기, 년 1965.01 ~ 2021.04 / 자료문의처 : 042-481-2533 / 기능문의: KOSIS Q&A 게시판

일괄설정 +

항목 [1/1]

시도별 [1/17]

지출목적별 [13/585]

시점 [6/957]

(단위 : 2015=100)

주석정보

주소정보

행렬전환

분석

차트

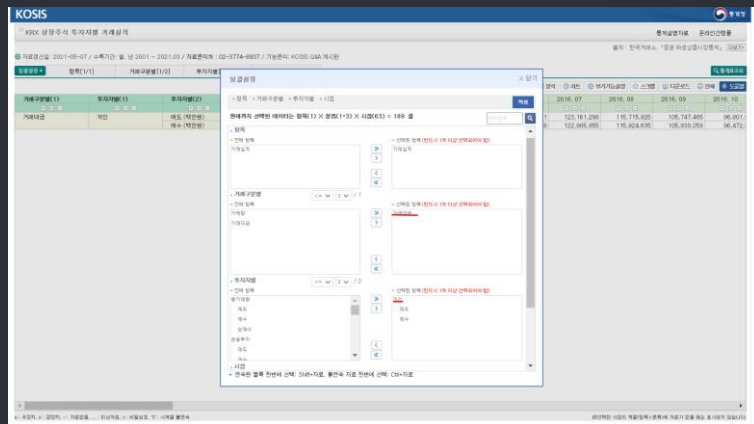
부가기능

시도별	지출목적별	2021. 04	2021. 03	2021. 02	2021. 01	2020. 12
전국	0 총지수	107.39	107.16	107.00	106.47	105.67
	01 식료품·비주류음료	120.34	120.61	120.97	118.04	115.79
	02 주류 및 담배	103.87	103.43	103.54	103.58	103.38
	03 의류 및 신발	104.72	105.05	105.34	105.12	105.43
	04 주택·수도·전기 및 연료	104.44	104.48	104.18	103.60	103.58
	05 가전용품 및 가사 서비스	107.43	107.19	107.94	108.53	107.45
	06 보건	104.69	104.91	104.81	104.83	103.82

4

## 데이터 전처리

# 데이터 전처리 : STOCK(개인 주식 투자 금액 지수)



A	B	C	D	E	F	G	H	
거래구분별(1)	투자자별(1)	투자자별(2)	2017. 10	2017. 11	2017. 12	2018. 01	2018. 02	2018. 03
거래대금	개인	매도 (백만원)	98,863,839	191,351,506	160,638,726	247,689,728	152,057,343	
		매수 (백만원)	99,473,851	190,008,318	155,507,676	246,764,203	155,403,931	
		평균	99168845	190679912	158073201	247226965.5	153730637	

개인 투자자의 매도/매수 금액 평균을 구한 후 지수로 변환

## ● 데이터 전처리 : HOUSE(전국 아파트 실거래가 지수)

지 역		2017년 10월		2017년 11월		2017년 12월		2018년 01월	
		지수	변동률	지수	변동률	지수	변동률	지수	변동률
전국		100.289	0.08	100	-0.29	99.692	-0.31	99.936	
수도권		100.286	0.46	100	-0.29	99.971	-0.03	100.79	
지방		100.288	-0.23	100	-0.29	99.462	-0.54	99.236	-
6대광역시		100.294	-0.08	100	-0.29	99.674	-0.33	99.561	-
5대광역시		100.186	-0.13	100	-0.19	99.619	-0.38	99.445	-
8개도		100.383	-0.34	100	-0.38	99.306	-0.69	99.026	-
서울		99.567	0.97	100	0.44	100.884	0.88	102.905	
	도심권	98.999	0.24	100	1.01	100.578	0.58	103.101	
	동북권	99.847	0.93	100	0.15	100.324	0.32	101.442	
	서북권	100.218	0.67	100	-0.22	100.155	0.16	101.61	
	서남권	99.781	0.92	100	0.22	100.695	0.69	102.044	
	동남권	98.725	1.23	100	1.29	102.144	2.14	106.49	
부산		100.064	-0.3	100	-0.06	99.128	-0.87	98.651	-
대구		99.932	0.37	100	0.07	99.928	-0.07	100.402	
인천		100.763	0.15	100	-0.76	99.912	-0.09	100.07	
광주		99.831	0.21	100	0.17	100.731	0.73	101.024	
대전		100.957	0.28	100	-0.95	100.266	0.27	100.093	-
울산		100.579	-1.02	100	-0.58	98.759	-1.24	98.029	-
세종		101.224	2.21	100	-1.21	99.576	-0.42	99.888	
경기		100.63	0.21	100	-0.63	99.395	-0.61	99.588	
강원		100.222	0.14	100	-0.22	99.583	-0.42	99.738	
충북		99.591	-0.96	100	0.41	99.08	-0.92	98.111	-
충남		100.288	-0.62	100	-0.29	99.186	-0.81	99.032	-
전북		100.897	0.07	100	-0.89	100.041	0.04	100.364	

## ● 데이터 전처리 : COIN(가상화폐 투자 공포 탐욕 지수)

2021/05/09	76.92	13972.08	11236.89
2021/05/08	79.37	13944.28	11146.83
2021/05/07	74.02	13237.82	10499.82
2021/05/06	77.74	13325.70	10615.85
2021/05/05	78.43	13162.01	10348.97
2021/05/04	70.35	12497.18	9533.52
2021/05/03	77.90	12929.46	9708.84
2021/05/02	68.41	12175.49	8683.32
2021/05/01	71.21	12231.69	8672.97
2021/04/30	66.71	11835.06	8186.45

C	D	피벗 테이블 필드
균 : 공포지수		보고서에 추가할 필드 선택
61.48		검색
75.42		<input checked="" type="checkbox"/> 날짜
81.06		<input checked="" type="checkbox"/> 공포지수
47.27		<input type="checkbox"/> UBM
34.87		<input type="checkbox"/> UBAI
30.95		<input type="checkbox"/> 분기
51.29		<input type="checkbox"/> ...
47.99		아래 영역 사이에 필드를 끌어 놓으십시오.
37.42		필터
47.91		행
38.34		열
43.15		

	A	B	C	D
1	년도	월	평균 공포지수	
2	2017년	10월	61.07	
3	2017년	11월	76.54	
4	2017년	12월	86.69	
5	2018년	1월	49.89	
6	2018년	2월	32.13	
7	2018년	3월	28.41	
8	2018년	4월	48.24	
9	2018년	5월	48.24	
10	2018년	6월	36.56	
11	2018년	7월	47.43	
12	2018년	8월	37.76	
13	2018년	9월	42.96	
14	2018년	10월	46.88	
15	2018년	11월	34.78	
16	2018년	12월	32.71	
17	2019년	1월	38.96	
18	2019년	2월	48.28	
19	2019년	3월	53.55	
20	2019년	4월	67.11	
21	2019년	5월	74.71	
22	2019년	6월	69.95	
23	2019년	7월	48.03	
24	2019년	8월	48.75	

- 데이터 전처리 : 엑셀파일로 정리

data0



	A	B	C	D	E	F	G	H
1	stock	house	y_job	job	moneyrate	coin	foreigner	customerprice
2	52.01	100.289	8.6	3.2	1.25	61.07	34.6	103.39
3	100.00	100	9.2	3.1	1.5	76.54	33.8	102.62
4	82.90	99.692	9.2	3.3	1.5	86.69	33.6	102.99
5	129.66	99.936	8.7	3.7	1.5	49.89	33	103.42
6	80.62	100.356	9.8	4.6	1.5	32.13	32.9	104.21
7	91.46	100.361	11.6	4.5	1.5	28.41	32.7	104.10
8	111.86	100.467	10.7	4.1	1.5	48.24	32.8	104.29
9	107.91	100.205	10.5	4	1.5	48.24	32.7	104.34
10	80.17	100.125	9	3.7	1.5	36.56	32.7	104.13
11	63.62	100.093	9.3	3.7	1.5	47.43	33	103.93
12	63.48	101.162	10	4	1.5	37.76	32.8	104.85
13	61.77	102.605	8.8	3.6	1.5	42.96	32.3	105.65
14	60.38	102.915	8.4	3.5	1.5	46.88	33.1	105.46
15	57.33	102.213	7.9	3.2	1.75	34.78	32.3	104.71
16	50.94	101.45	8.6	3.4	1.75	32.71	32.1	104.35
17	68.65	100.99	8.9	4.5	1.75	38.96	32.3	104.24
18	53.78	100.575	9.5	4.7	1.75	48.28	32.1	104.69
19	60.78	99.913	10.8	4.3	1.75	53.55	32.2	104.49
20	73.87	99.62	11.5	4.4	1.75	67.11	32.4	104.87
21	66.86	99.476	9.9	4	1.75	74.71	32.3	105.05
22	57.50	99.789	10.4	4	1.75	69.95	32.7	104.88
23	69.41	100.204	9.8	3.9	1.5	48.03	33.5	104.56
24	60.58	100.517	7.2	3	1.5	48.75	33.3	104.81
25	54.74	100.891	7.3	3.1	1.5	42.83	33.6	105.20

read\_excel("data0.xlsx")

5

## 데이터 분석



## 데이터 구조 / 기초통계량 확인

- str()
- summary()

```
str(data0)
```

```
## tibble[,8] [41 x 8] (S3: tbl_df/tbl/data.frame)
## $ stock      : num [1:41] 52 100 82.9 129.7 80.6 ...
## $ house      : num [1:41] 100.3 100 99.7 99.9 100.4 ...
## $ y_job      : num [1:41] 8.6 9.2 9.2 8.7 9.8 11.6 10.7 10.5 9 9.3 ...
## $ job        : num [1:41] 3.2 3.1 3.3 3.7 4.6 4.5 4.1 4 3.7 3.7 ...
## $ moneyrate  : num [1:41] 1.25 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 ...
## $ coin       : num [1:41] 61.1 76.5 86.7 49.9 32.1 ...
## $ foreigner  : num [1:41] 34.6 33.8 33.6 33 32.9 32.7 32.8 32.7 32.7 33 ...
## $ customerprice: num [1:41] 103 103 103 103 104 ...
```

```
summary(data0)
```

##	stock	house	y_job	job
##	Min. : 50.94	Min. : 99.48	Min. : 7.000	Min. :3.000
##	1st Qu.: 61.77	1st Qu.:100.20	1st Qu.: 8.300	1st Qu.:3.400
##	Median : 80.62	Median :101.40	Median : 9.200	Median :4.000
##	Mean :117.22	Mean :105.02	Mean : 9.129	Mean :3.876
##	3rd Qu.:154.70	3rd Qu.:106.32	3rd Qu.: 9.900	3rd Qu.:4.200
##	Max. :324.88	Max. :127.69	Max. :11.600	Max. :5.700
##	moneyrate	coin	foreigner	customerprice
##	Min. :0.500	Min. :28.41	Min. :30.0	Min. :102.6
##	1st Qu.:0.750	1st Qu.:42.96	1st Qu.:31.6	1st Qu.:104.3
##	Median :1.500	Median :48.75	Median :32.7	Median :104.9
##	Mean :1.232	Mean :52.82	Mean :32.4	Mean :104.9
##	3rd Qu.:1.500	3rd Qu.:62.49	3rd Qu.:33.1	3rd Qu.:105.5
##	Max. :1.750	Max. :86.69	Max. :34.6	Max. :107.0

## house만을 이용한 단순선형 회귀분석

- result0 <- lm(stock~house, data=data0)
- summary(result0)

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -837.8239    78.4790  -10.68 3.89e-13 ***
## house        9.0938     0.7455   12.20 6.91e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34.95 on 39 degrees of freedom
## Multiple R-squared:  0.7923, Adjusted R-squared:  0.787
## F-statistic: 148.8 on 1 and 39 DF, p-value: 6.911e-15
```

$$\widehat{STOCK} = -837.82 + 9.09house$$

## 1차 다중회귀분석 변수 선택

- 설명력을 높이기 위해 house(전국 아파트 실거래가 지수)이외에 y\_job(청년 실업률), job(전체 실업률), moneyrate(기준금리) 변수를 추가로 사용

```
data1 <- data0[c('stock', 'house', 'y_job', 'job', 'moneyrate')]  
head(data1)
```

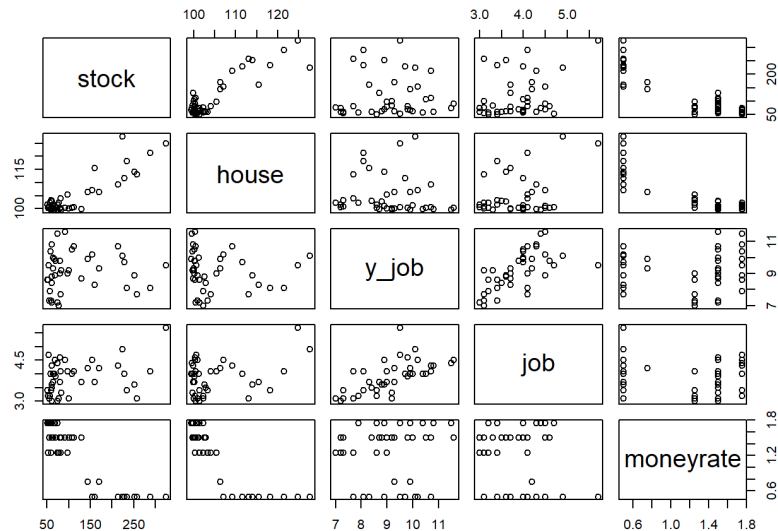
```
## # A tibble: 6 x 5  
##   stock house y_job  job moneyrate  
##   <dbl> <dbl> <dbl> <dbl>     <dbl>  
## 1  52.0  100.    8.6   3.2      1.25  
## 2  100    100    9.2   3.1      1.5  
## 3  82.9  99.7    9.2   3.3      1.5  
## 4  130.   99.9    8.7   3.7      1.5  
## 5  80.6  100.    9.8   4.6      1.5  
## 6  91.5  100.   11.6   4.5      1.5
```

## 변수간 상관관계 파악

- `cor()`
- `pairs()`

```
cor(data1)
```

##	stock	house	y_job	job	moneyrate
## stock	1.00000000	0.89013862	0.03593275	0.3376034	-0.8853730
## house	0.89013862	1.00000000	-0.09197738	0.3454285	-0.8535554
## y_job	0.03593275	-0.09197738	1.00000000	0.6592215	0.1054085
## job	0.33760341	0.34542847	0.65922152	1.0000000	-0.1827480
## moneyrate	-0.88537296	-0.85355544	0.10540847	-0.1827480	1.0000000



## 변수간 상관관계 파악

- `library(corrplot)`
- `corrplot()`

stock과 나머지 변수들 간의  
선형성 확인

: y\_job 과 job 중에서 상관계수가 더  
낮은 y\_job을 제거하기로 결정



stock : 개인 주식 투자 금액 지수  
house : 전국 아파트 실거래가 지수  
job : 전체 실업률  
moneyrate : 기준금리

## 다중 회귀분석 / 다중공선성 확인

- result1<-lm(stock~house+job+moneyrate,data=data1)
- summary(result1)

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -290.316    149.530  -1.942  0.05984 .
## house         4.359      1.305    3.340  0.00192 **
## job          12.509      8.618    1.451  0.15508
## moneyrate    -80.130     19.416  -4.127  0.00020 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.62 on 37 degrees of freedom
## Multiple R-squared:  0.8585, Adjusted R-squared:  0.847
## F-statistic: 74.83 on 3 and 37 DF, p-value: 8.949e-16
```

$$\widehat{STOCK} = -290.32 + 4.36house + 12.51job - 80.13moneyrate$$

- 다중공선성 확인

```
vif(result1)
```

```
##      house      job moneyrate
## 4.267768 1.198482 3.888395
```

: 양호

## 회귀분석의 가정

선형성 (변수 간 상관관계에서 확인)

독립성

정규성

등분산성

## 독립성 확인

- $H_0$  : 잔차들이 서로 독립적이다.
- $H_1$  : 잔차들이 서로 독립적이지 않다.

```
durbinWatsonTest(result1)
```

```
## lag Autocorrelation D-W Statistic p-value  
## 1 0.1899558 1.46302 0.026  
## Alternative hypothesis: rho != 0
```

-> 새로운 변수 추가 필요성 확인

- 새로운 변수 투입  
coin(가상화폐 지수), foreigner(주식시장 내 외국인 보유금액 비율),  
customerprice(소비자 물가지수)

```
data2 <- data0[c('stock', 'house', 'job', 'moneyrate', 'coin', 'foreigner', 'customerprice')]  
head(data2)
```

```
## # A tibble: 6 x 7  
## stock house job moneyrate coin foreigner customerprice  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 52.0 100. 3.2 1.25 61.1 34.6 103.  
## 2 100 100 3.1 1.5 76.5 33.8 103.  
## 3 82.9 99.7 3.3 1.5 86.7 33.6 103.  
## 4 130. 99.9 3.7 1.5 49.9 33 103.  
## 5 80.6 100. 4.6 1.5 32.1 32.9 104.  
## 6 91.5 100. 4.5 1.5 28.4 32.7 104.
```

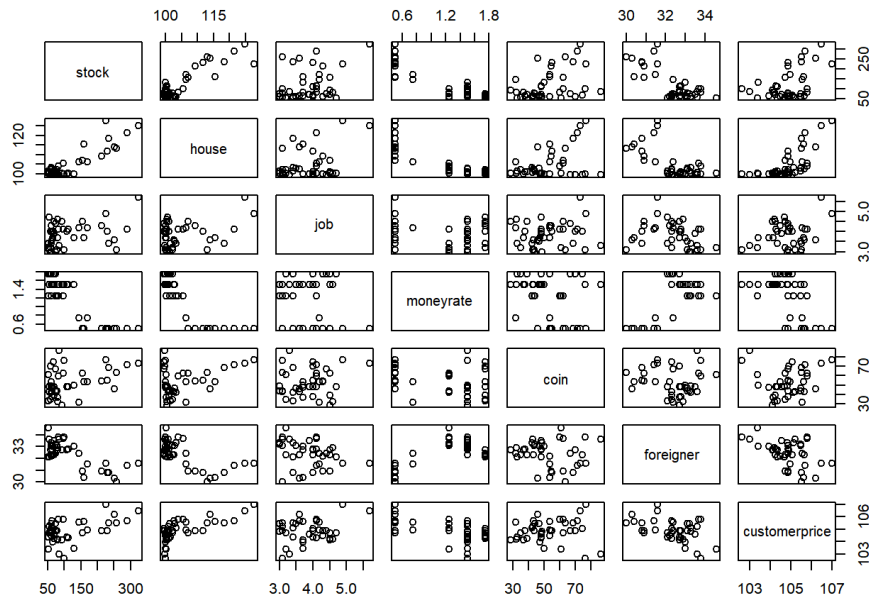


## 변수간 상관관계 파악

- `cor()`
- `pairs()`

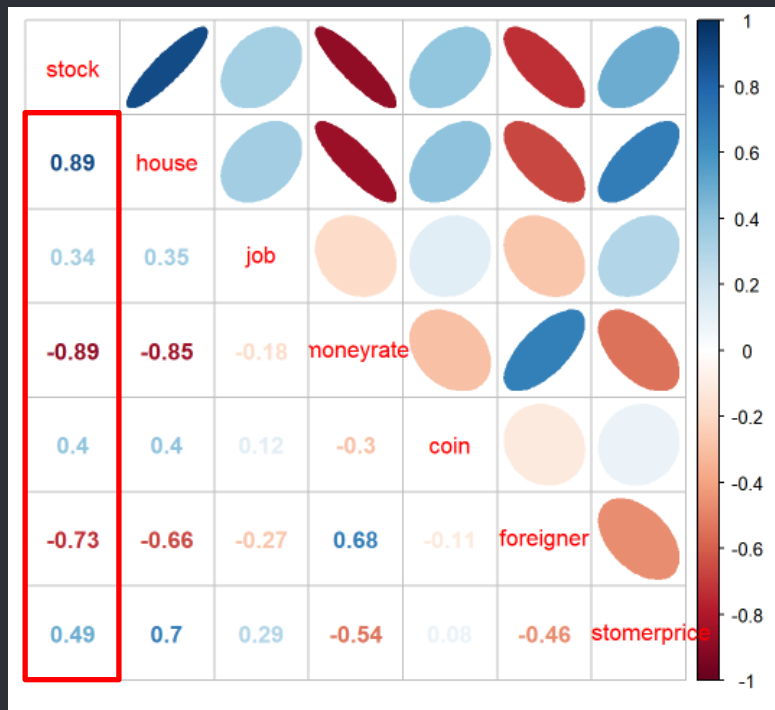
```
cor(data2)
```

```
##          stock      house      job  moneyrate      coin
## stock      1.000000    0.8901386  0.3376034 -0.8853730  0.39663695
## house      0.8901386    1.0000000  0.3454285 -0.8535554  0.40062227
## job        0.3376034    0.3454285  1.0000000 -0.1827480  0.12074959
## moneyrate  -0.8853730   -0.8535554  -0.1827480  1.0000000  -0.29976986
## coin       0.3966389    0.4006223  0.1207495  -0.2997699  1.00000000
## foreigner  -0.7335571   -0.6619969  -0.2713717  0.6808660  -0.11384521
## customerprice 0.4941167  0.6978144  0.2920704  -0.5400510  0.08401998
##          foreigner customerprice
## stock      -0.7335571    0.49411665
## house      -0.6619969    0.69781444
## job        -0.2713717    0.29207039
## moneyrate  -0.6808660   -0.54005104
## coin       -0.1138452    0.08401998
## foreigner   1.0000000   -0.46463703
## customerprice -0.4646370    1.00000000
```



## 변수간 상관관계 파악

- corrplot()



## 회귀분석

stock : 개인 주식 투자 금액 지수  
house : 전국 아파트 실거래가 지수  
job : 전체 실업률  
moneyrate : 기준금리  
coin : 가상화폐 투자 공포 탐욕 지수  
foreigner : 주식시장 내 외국인 보유금액 비율  
customerprice : 소비자 물가지수

- `result2 <- lm(stock ~house+job+moneyrate+coin+foreigner+customerprice,data=data2)`
- `summary(result2)`

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1733.9250   686.2334   2.527  0.01633 *
## house         5.2799     1.4874   3.550  0.00115 **
## job          11.0681     7.7314   1.432  0.16139
## moneyrate    -61.8394    18.4350  -3.354  0.00196 **
## coin          0.3026     0.3402   0.889  0.38001
## foreigner    -13.0977     5.5496  -2.360  0.02415 *
## customerprice -16.4941     6.9799  -2.363  0.02398 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.3 on 34 degrees of freedom
## Multiple R-squared:  0.8975, Adjusted R-squared:  0.8794
## F-statistic: 49.62 on 6 and 34 DF, p-value: 2.12e-15
```

$$\widehat{stock} = 1733.93 + 5.28house + 11.07job - 61.84moneyrate + 0.3coin - 13.1foreigner - 16.49customerprice$$

## stepAIC

- library(MASS)
- result3 <- MASS::stepAIC(result2)

```
result3 <- MASS::stepAIC(result2)
```

```
## Start: AIC=274.42  
## stock ~ house + job + moneyrate + coin + foreigner + customerprice  
##  
##           Df Sum of Sq  RSS   AIC  
## - coin      1     547.1 24059 273.36  
## <none>                23512 274.42  
## - job       1    1417.3 24930 274.82  
## - foreigner  1    3852.0 27364 278.64  
## - customerprice 1    3861.7 27374 278.65  
## - moneyrate  1    7781.5 31294 284.14  
## - house     1    8713.9 32226 285.35  
##  
## Step: AIC=273.36  
## stock ~ house + job + moneyrate + foreigner + customerprice  
##  
##           Df Sum of Sq  RSS   AIC  
## <none>                24059 273.36  
## - job       1    1411.8 25471 273.70  
## - foreigner  1    3435.9 27495 276.84  
## - customerprice 1    5343.7 29403 279.59  
## - moneyrate  1    7514.7 31574 282.51  
## - house     1   12980.6 37040 289.05
```

: coin 탈락

## stepAIC 결과

- summary(result3)

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1861.609    669.045   2.782 0.008635 **
## house         5.839      1.344   4.346 0.000114 ***
## job          11.047      7.708   1.433 0.160699
## moneyrate    -60.595    18.327  -3.306 0.002192 **
## foreigner    -12.130     5.426  -2.236 0.031849 *
## customerprice -18.433     6.611  -2.788 0.008513 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.22 on 35 degrees of freedom
## Multiple R-squared:  0.8951, Adjusted R-squared:  0.8801
## F-statistic: 59.74 on 5 and 35 DF, p-value: 3.799e-16
```

$$\widehat{stock} = 1861.61 + 5.84\text{house} + 11.05\text{job} - 60.96\text{moneyrate} - 12.13\text{foreigner} - 18.43\text{customerprice}$$

- 다중공선성 확인

```
car::vif(result3)
```

##	house	job	moneyrate	foreigner	customerprice
##	5.773948	1.223543	4.421304	1.994274	2.001409

: 양호

## 독립성 확인

- $H_0$  : 잔차들이 서로 독립적이다.
- $H_1$  : 잔차들이 서로 독립적이지 않다.

```
durbinWatsonTest(result3)
```

```
## lag Autocorrelation D-W Statistic p-value  
## 1 -0.003869169 1.86609 0.238  
## Alternative hypothesis: rho != 0
```

: 독립성 확인

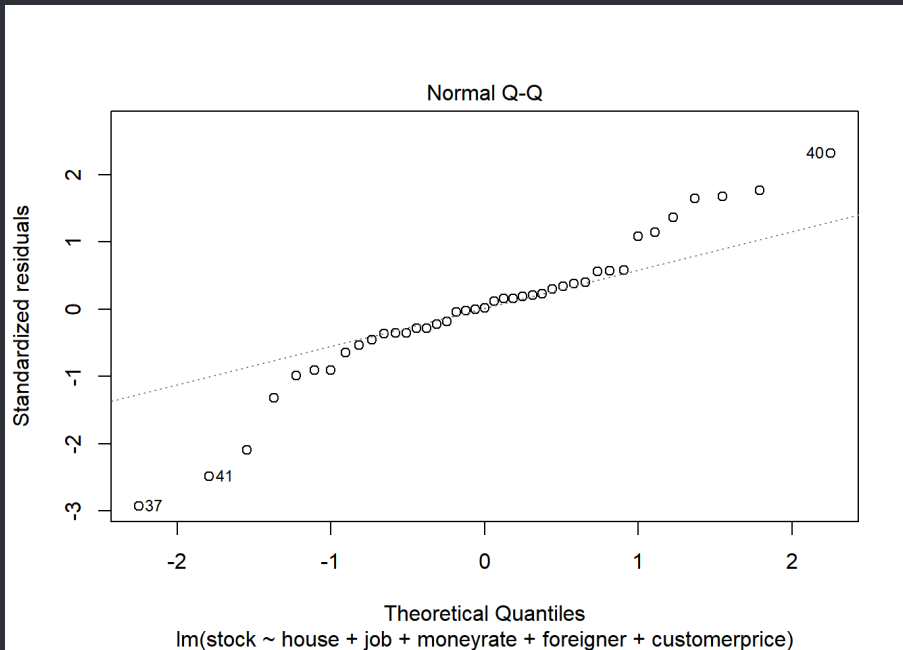
## 정규성 확인

- $H_0$  : 잔차들이 정규성을 띄고 있다.
- $H_1$  : 잔차들이 정규성을 띄고 있지 않다.

```
shapiro.test(resid(result3))
```

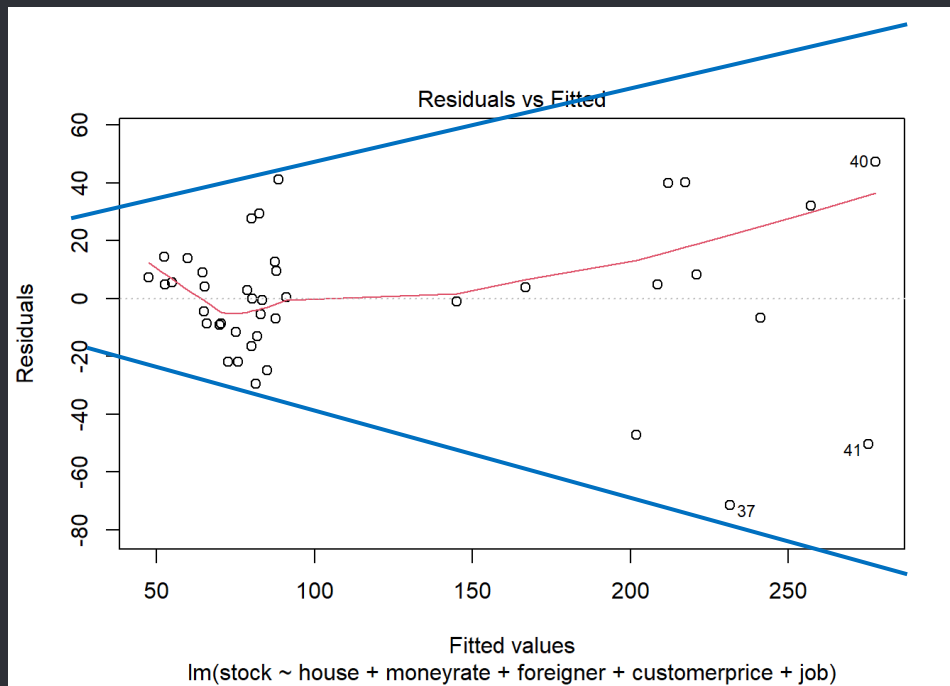
```
##  
## Shapiro-Wilk normality test  
##  
## data:  resid(result3)  
## W = 0.95588, p-value = 0.1127
```

: 정규성 확인



## 등분산성 확인

○ `plot(result3,1)`



-> 추가 분석 필요

: ! 잔차들이 커지는 경향



## Log 변환

stock : 개인 주식 투자 금액 지수  
house : 전국 아파트 실거래가 지수  
job : 전체 실업률  
moneyrate : 기준금리  
coin : 가상화폐 투자 공포 탐욕 지수  
foreigner : 주식시장 내 외국인 보유금액 비율  
customerprice : 소비자 물가지수

- log\_lm <- lm(log(stock)~house+moneyrate+coin+job+foreigner+customerprice, data=data2)
- result4 <- stepAIC(log\_lm)

```
## Start: AIC=-122.59
## log(stock) ~ house + moneyrate + coin + job + foreigner + customerprice
##
##           Df Sum of Sq  RSS   AIC
## - coin      1   0.04013 1.5056 -123.48
## <none>                 1.4654 -122.59
## - house      1   0.14792 1.6134 -120.65
## - job         1   0.17718 1.6426 -119.91
## - foreigner   1   0.19998 1.6654 -119.34
## - customerprice 1   0.26777 1.7332 -117.71
## - moneyrate   1   1.01907 2.4845 -102.94
##
## Step: AIC=-123.48
## log(stock) ~ house + moneyrate + job + foreigner + customerprice
##
##           Df Sum of Sq  RSS   AIC
## <none>                 1.5056 -123.48
## - foreigner      1   0.17305 1.6786 -121.02
## - job             1   0.17665 1.6822 -120.93
## - house           1   0.26825 1.7738 -118.76
## - customerprice   1   0.37276 1.8783 -116.41
## - moneyrate       1   0.99434 2.4999 -104.69
```

: coin 탈락

## 회귀분석

- summary(result4)

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  21.11940    5.29253   3.990 0.000321 ***
## house        0.02655    0.01063   2.497 0.017372 *
## moneyrate    -0.69702    0.14498  -4.808 2.86e-05 ***
## job          0.12357    0.06098   2.026 0.050387 .
## foreigner    -0.08608    0.04292  -2.006 0.052668 .
## customerprice -0.15395    0.05230  -2.944 0.005725 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2074 on 35 degrees of freedom
## Multiple R-squared:  0.8805, Adjusted R-squared: 0.8634
## F-statistic: 51.58 on 5 and 35 DF, p-value: 3.64e-15
```

$$\log(\widehat{stock}) = 21.12 + 0.03\text{house} - 0.7\text{moneyrate} + 0.12\text{job} - 0.09\text{foreigner} - 0.15\text{customerprice}$$

## 독립성 확인

- $H_0$  : 잔차들이 서로 독립적이다.
- $H_1$  : 잔차들이 서로 독립적이지 않다.

```
durbinWatsonTest(result4)
```

```
## lag Autocorrelation D-W Statistic p-value  
## 1 0.09880391 1.646206 0.06  
## Alternative hypothesis: rho != 0
```

: 독립성 확인

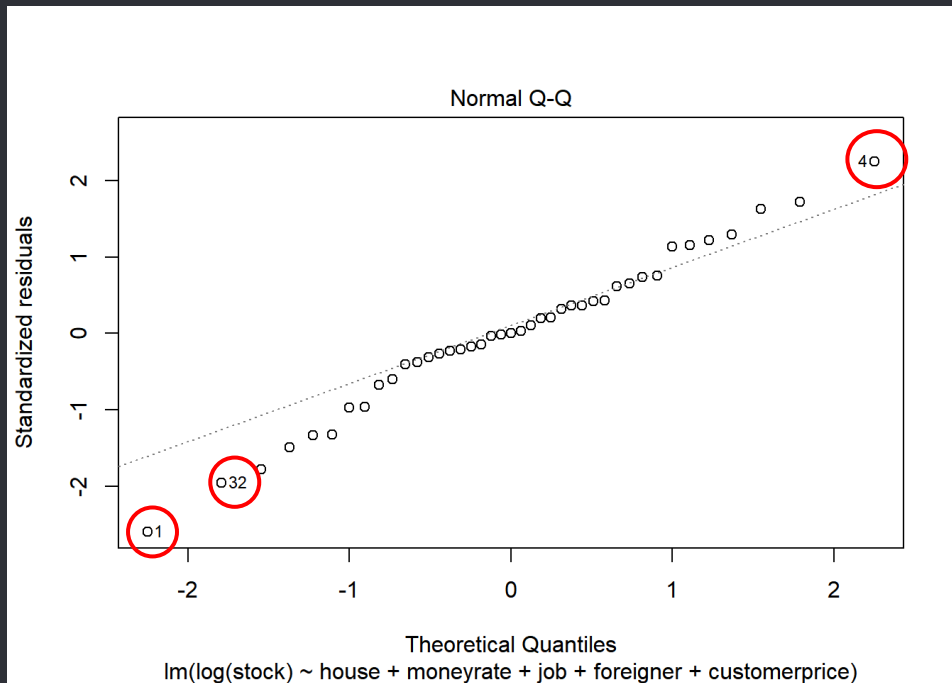
## 정규성 확인

- $H_0$  : 잔차들이 정규성을 띄고 있다.
- $H_1$  : 잔차들이 정규성을 띄고 있지 않다.

```
shapiro.test(resid(result4))
```

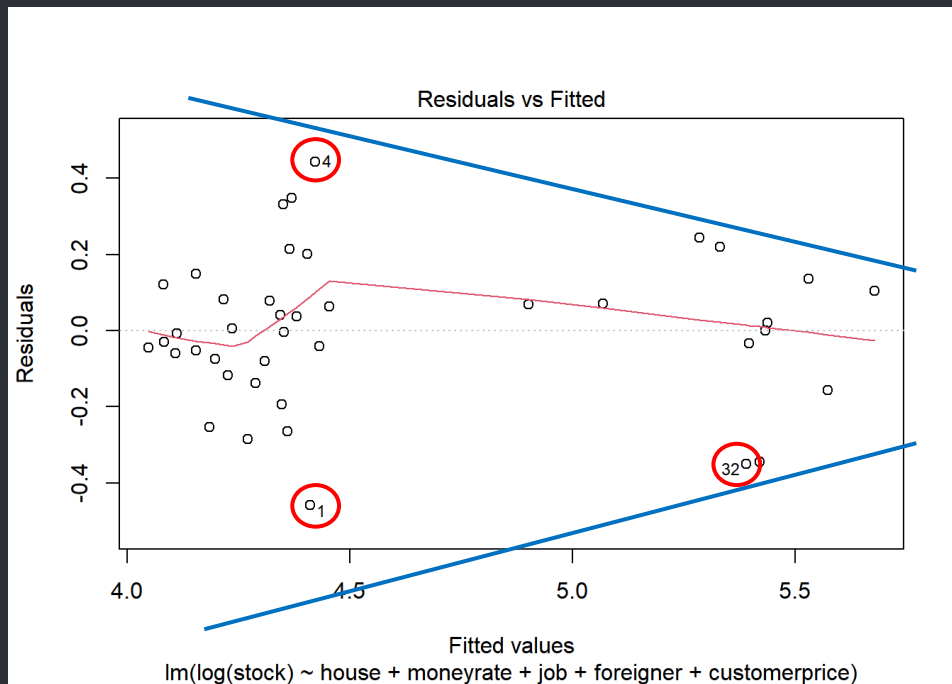
```
##  
## Shapiro-Wilk normality test  
##  
## data:  resid(result4)  
## W = 0.98645, p-value = 0.8989
```

: 정규성 확인



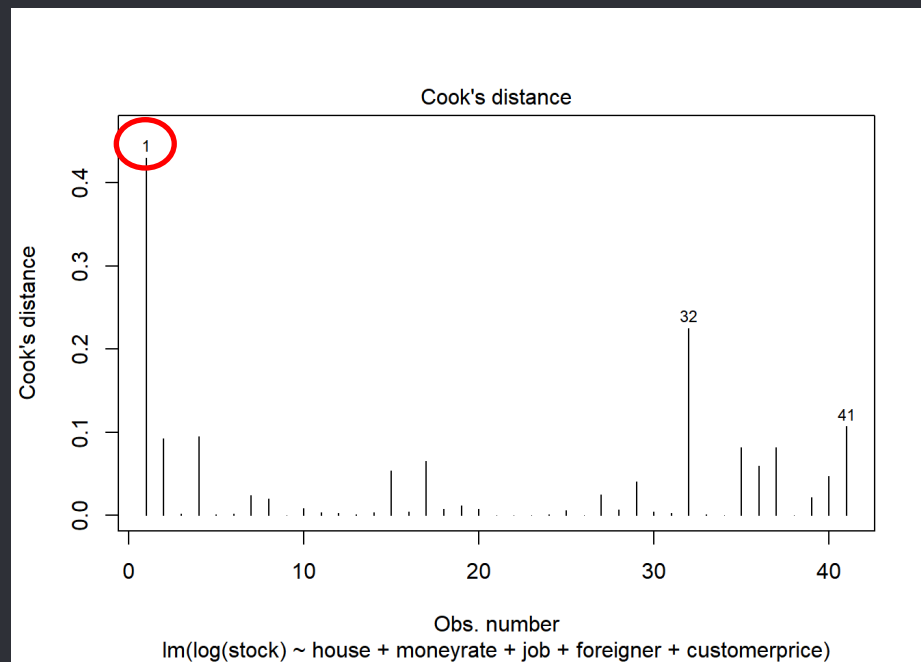
## 등분산성 확인

○ `plot(result4,1)`



## • 쿡의 거리

- `plot(result4,4)`



## 이상치 제거

- data5 <- data2[-1,]
- log\_lm <- lm(log(stock)~house+moneyrate+coin+job+foreigner+customerprice, data=data5)
- result5 <- stepAIC(log\_lm)

stock : 개인 주식 투자 금액 지수  
house : 전국 아파트 실거래가 지수  
job : 전체 실업률  
moneyrate : 기준금리  
coin : 가상화폐 투자 공포 탐욕 지수  
foreigner : 주식시장 내 외국인 보유금액 비율  
customerprice : 소비자 물가지수

```
## Start: AIC=-127.17
## log(stock) ~ house + moneyrate + coin + job + foreigner + customerprice
##
##           Df Sum of Sq  RSS   AIC
## - foreigner    1  0.03735 1.2104 -127.92
## - coin          1  0.04161 1.2146 -127.78
## <none>                    1.1730 -127.17
## - job           1  0.14649 1.3195 -124.46
## - house         1  0.17079 1.3438 -123.73
## - customerprice 1  0.39975 1.5728 -117.44
## - moneyrate     1  1.22821 2.4013 -100.52
##
## Step: AIC=-127.92
## log(stock) ~ house + moneyrate + coin + job + customerprice
##
##           Df Sum of Sq  RSS   AIC
## - coin          1  0.02976 1.2401 -128.946
## <none>                    1.2104 -127.918
## - job           1  0.16221 1.3726 -124.887
## - house         1  0.20869 1.4191 -123.555
## - customerprice 1  0.43306 1.6434 -117.683
## - moneyrate     1  1.59704 2.8074 -96.264
##
## Step: AIC=-128.95
## log(stock) ~ house + moneyrate + job + customerprice
##
##           Df Sum of Sq  RSS   AIC
## <none>                    1.2401 -128.946
## - job           1  0.15955 1.3997 -126.105
## - house         1  0.33045 1.5706 -121.497
## - customerprice 1  0.54619 1.7863 -116.348
## - moneyrate     1  1.56767 2.8078 -98.259
```

: coin, foreigner 탈락

## 회귀분석

- summary(result5)

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  22.271420   4.724361   4.714 3.79e-05 ***
## house        0.029296   0.009593   3.054 0.004299 **
## moneyrate    -0.832063   0.125092  -6.652 1.08e-07 ***
## job          0.117164   0.055215   2.122 0.040992 *
## customerprice -0.192305   0.048980  -3.926 0.000386 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1882 on 35 degrees of freedom
## Multiple R-squared:  0.8981, Adjusted R-squared:  0.8865
## F-statistic: 77.15 on 4 and 35 DF, p-value: < 2.2e-16
```

$$\log(\widehat{stock}) = 22.27 + 0.03\text{house} - 0.83\text{moneyrate} + 0.12\text{job} - 0.19\text{customerprice}$$

- 다중공선성 확인

```
car::vif(result5)
```

```
##          house      moneyrate          job customerprice
##          5.648998      3.996053      1.177713      1.981859
```

: 양호



## 독립성

- $H_0$  : 잔차들이 서로 독립적이다.
- $H_1$  : 잔차들이 서로 독립적이지 않다.

```
durbinWatsonTest(result5)
```

```
## lag Autocorrelation D-W Statistic p-value  
## 1 0.129751 1.714692 0.148  
## Alternative hypothesis: rho != 0
```

: 독립성 확인

## 정규성 확인

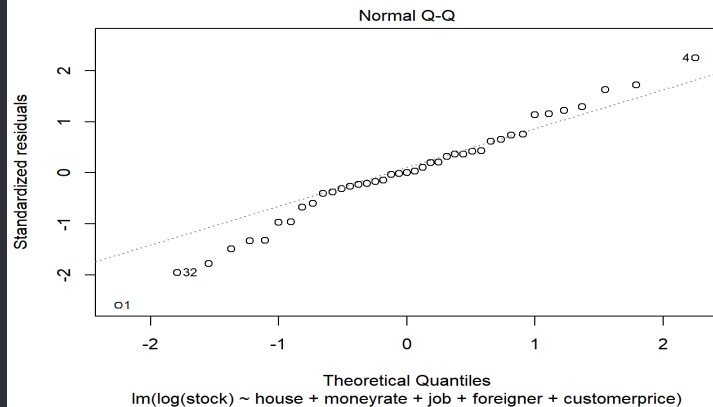
- $H_0$  : 잔차들이 정규성을 띄고 있다.
- $H_1$  : 잔차들이 정규성을 띄고 있지 않다.

```
shapiro.test(resid(result5))

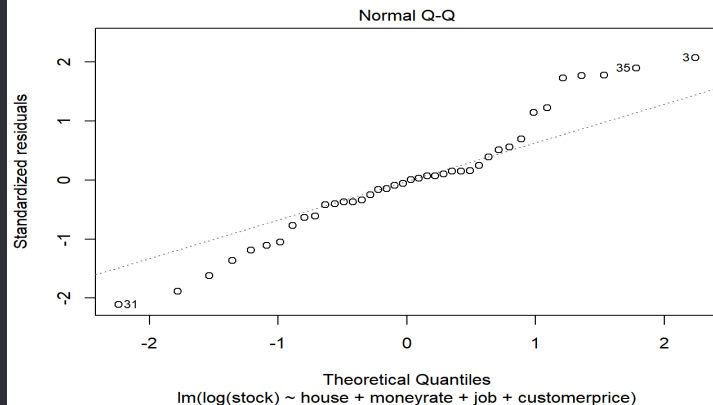
##
## Shapiro-Wilk normality test
##
## data:  resid(result5)
## W = 0.95566, p-value = 0.1188
```

여전히  
정규성을 가지고  
있음

result4  
\* 이상치 제거전



result5  
\* 이상치 제거후

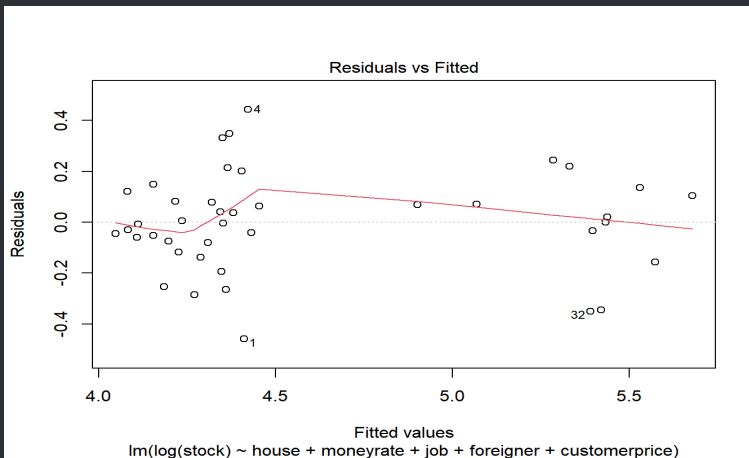


## 등분산성

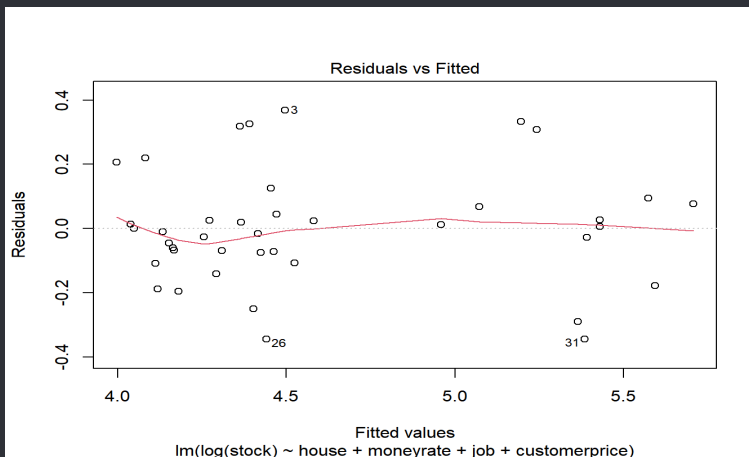
plot(result5,1)

log 취하고  
이상치까지 제거 후  
등분산성을  
만족하는 것을 확인

result4  
\* 이상치 제거전



result5  
\* 이상치 제거후



6

## 정리 및 예측

## 표준화 회귀계수

lm.beta::lm.beta(result5)

```
lm.beta::lm.beta(result5)

##
## Call:
## lm(formula = log(stock) ~ house + moneyrate + job + customerprice,
##     data = data5)
##
## Standardized Coefficients::
##      (Intercept)      house      moneyrate      job customerprice
##      0.0000000      0.3915723     -0.7173274      0.1242327     -0.2981838
```

#개인의\_주식투자에\_가장\_큰\_영향을\_끼치는\_것은  
#금리이다

**-0.7173274**  
moneyrate

**0.3915723**  
house

**-0.2981838**  
job

**0.1242327**  
customerprice

## 엑셀 FORECAST.ETS 함수로 값 예측

- FORECAST.ETS()  
지수평활법 알고리즘으로 시계열 데이터에서 미래의 값을 예측

```
new_data = data.frame(house,moneyrate,job,customerprice)
```

house

	date	expect
1	2021년 3월	127.3211
2	2021년 4월	130.1539
3	2021년 5월	132.9868

job

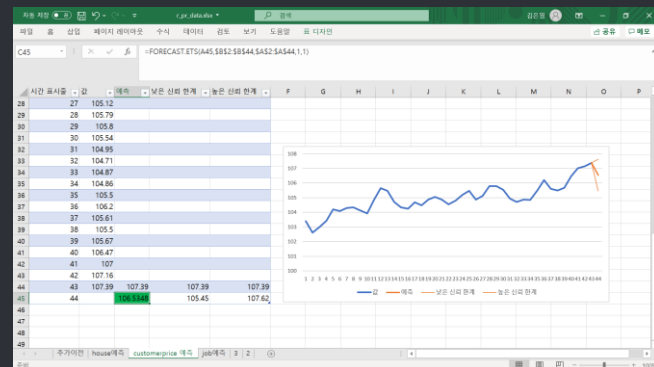
	date	expect
1	2021년 3월	4.3
2	2021년 4월	4
3	2021년 5월	4.2

moneyrate

	date	expect
1	2021년 3월	4.3
2	2021년 4월	4
3	2021년 5월	4.2

customerprice

	date	expect
1	2021년 3월	107.16
2	2021년 4월	107.39
3	2021년 5월	106.5348



## 예측

- `predict(최종회귀모형, new_data)`

```
yearmonth = c("2021년 3월", "2021년 4월", "2021년 5월")
expect = exp(predict(result5, newdata = new_data, interval = "confidence"))

total = data.frame(yearmonth, expect)
total
```

##	yearmonth	fit	lwr	upr
## 1	2021년 3월	240.2778	190.0841	303.7256
## 2	2021년 4월	241.1467	181.1553	321.0048
## 3	2021년 5월	316.1729	226.6664	441.0238

이미 공개된 2021년 3월의 주식시장 개인 거래실적 지수 변환 값 221  
-> 예측 값의 95% 신뢰수준 사이에 들어있기 때문에,  
이 모델이 상당히 신뢰할 만하다고 판단할 수 있다

Thanks!

ANY QUESTIONS?

