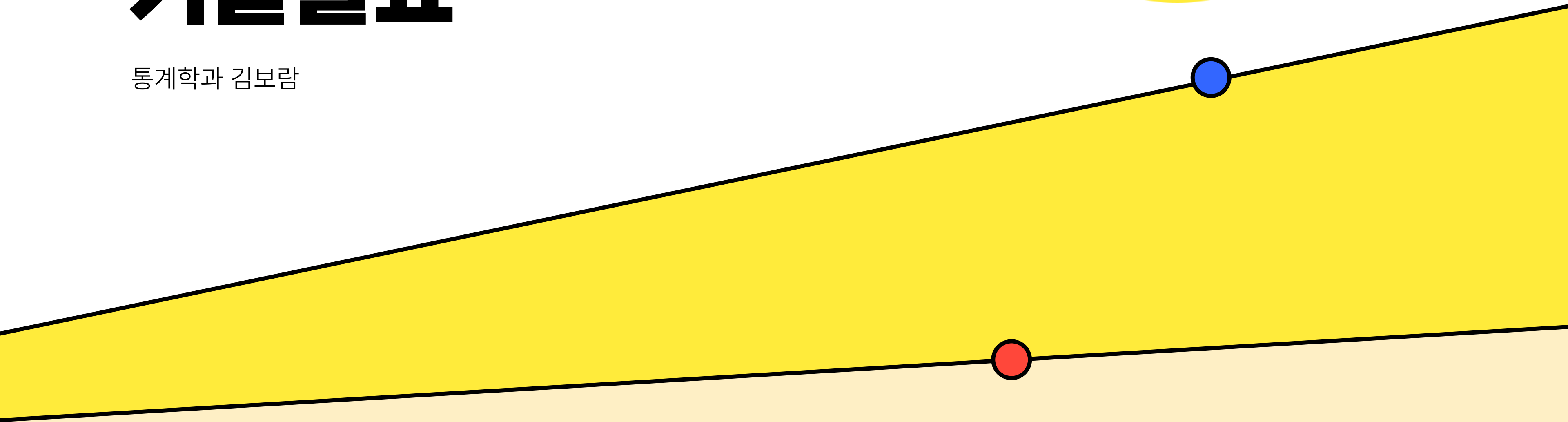


2023

고급시계열분석

기말발표

통계학과 김보람



목차



- 01 데이터 설명**
- 02 시계열 분석**
- 03 예측**
- 04 Train/Test**
- 05 결론**

데이터 설명

카카오 3년 선차트



데이터 설명

2023년 1월 ~ 11월 주식 데이터

해당 데이터는 11개의 설명 변수와
1월 ~ 11월의 226개의 열로 구성

해당 데이터에서 '일자'와 '종가' 데이터 사용

데이터 출처: KRX(정보데이터시스템)
<http://data.krx.co.kr>

카카오

	일자	종가	대비	등락률	시가	고가	저가	거래량	거래대금	시가총액	상장주식수
	<chr>	<int>	<int>	<dbl>	<int>	<int>	<int>	<int>	<dbl>	<dbl>	<int>
1	2023/01/02	52700	-400	-0.75	53600	53800	52400	887667	46979376500	2.347313e+13	445410387
2	2023/01/03	53300	600	1.14	52400	53500	51400	1420569	74588286800	2.374037e+13	445410387
3	2023/01/04	55700	2400	4.50	53200	56000	53100	2241411	123346180300	2.480936e+13	445410387
4	2023/01/05	57700	2000	3.59	55800	58200	55700	3046064	175103778900	2.570018e+13	445410387
5	2023/01/06	57200	-500	-0.87	57200	58000	56500	1420345	81326211100	2.547747e+13	445410387
6	2023/01/09	61100	3900	6.82	58700	61200	58300	3482961	208443993900	2.721457e+13	445410387

데이터 설명

주식 데이터이므로 공휴일의 거래 값이 없다.
1월 ~ 11월까지 108개의 결측값이 있다.

1	2023/01/01	NA
2	2023/01/02	52700
3	2023/01/03	53300
4	2023/01/04	55700
5	2023/01/05	57700
6	2023/01/06	57200
7	2023/01/07	NA
8	2023/01/08	NA
9	2023/01/09	61110

TYPE 1

1	2023/01/02	52700
2	2023/01/03	53300
3	2023/01/04	55700
4	2023/01/05	57700
5	2023/01/06	57200
6	2023/01/09	61110
7	2023/01/10	60700
8	2023/01/11	61900
9	2023/01/12	61500

TYPE 2

데이터 설명

시계열 데이터 NA값 결측값 처리

1. 결측값 채우기

na.locf : 앞쪽 값

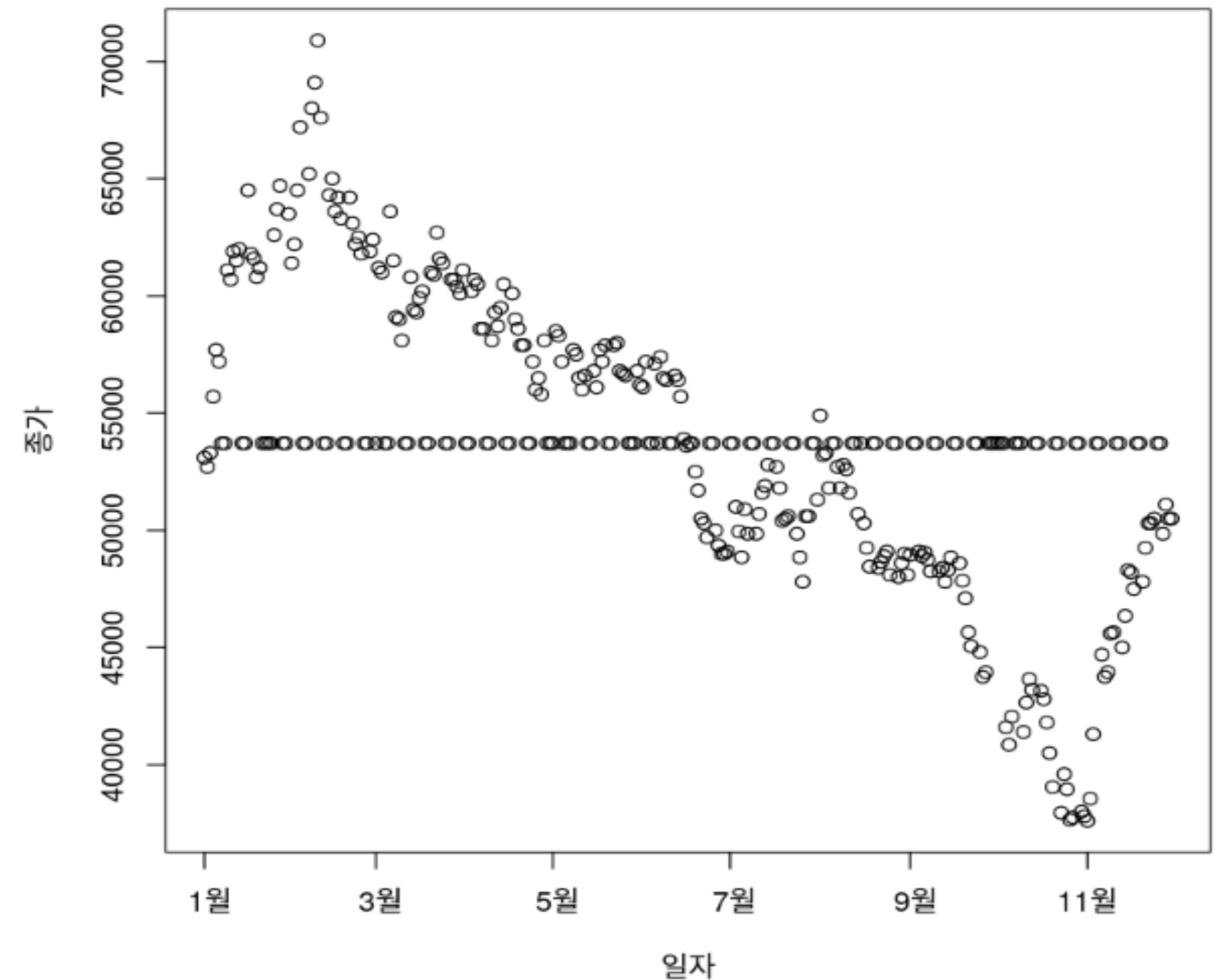
zoo:na.approx : 선형 보간

zoo:na.aggregate : 평균값, 중간값

2. 시계열 예측 모델 사용

3. 결측값 삭제

데이터 결측값(평균값 처리)



시계열 분석

x축: 일자(시간)
y축: 종가

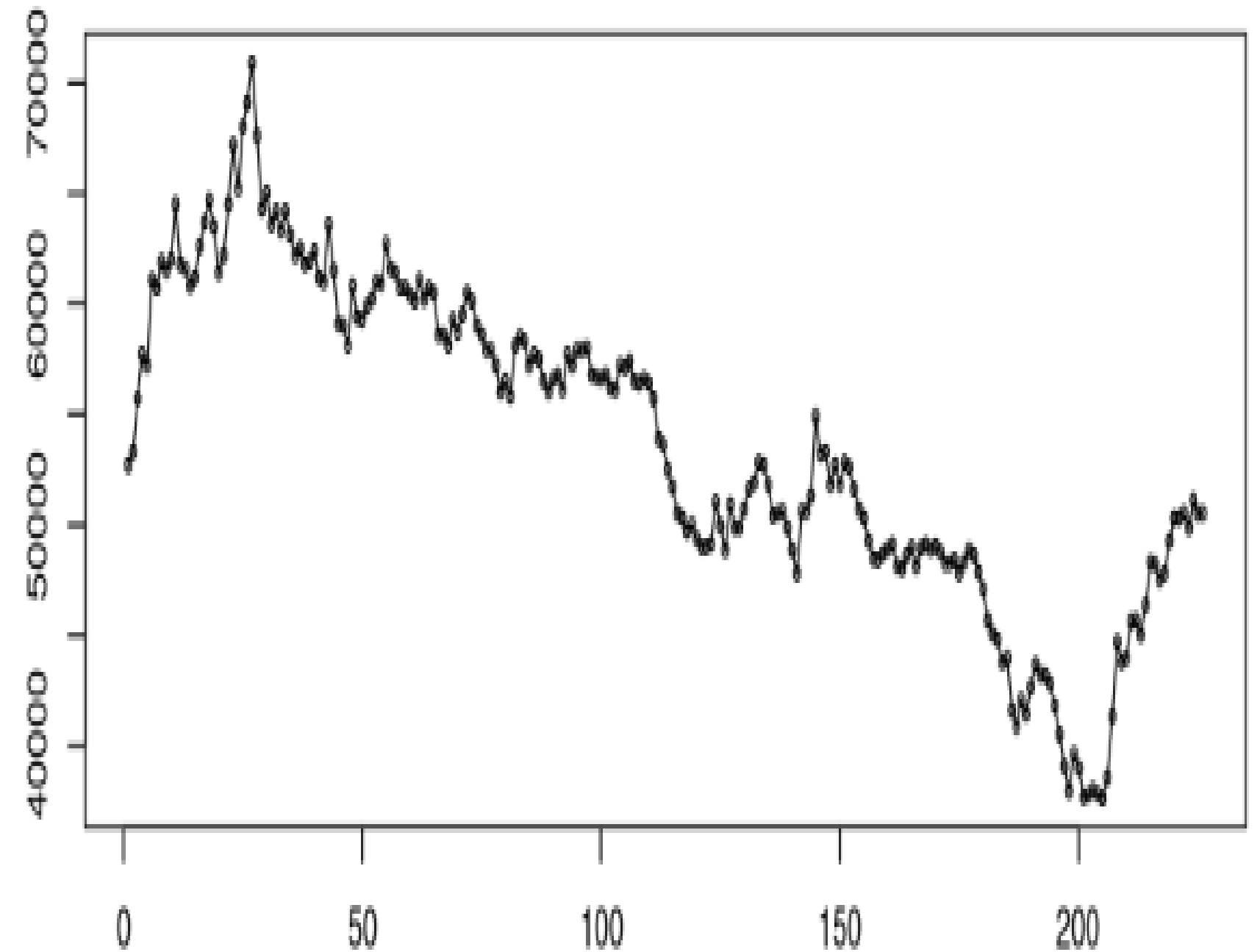
kakao의 하락세.... 추세가 보인다.

계절성분은 없어 보인다.

200시차 이후로 값이 증가하는 모양
→ train/test로 나누어서
예측해보면 어떨까?

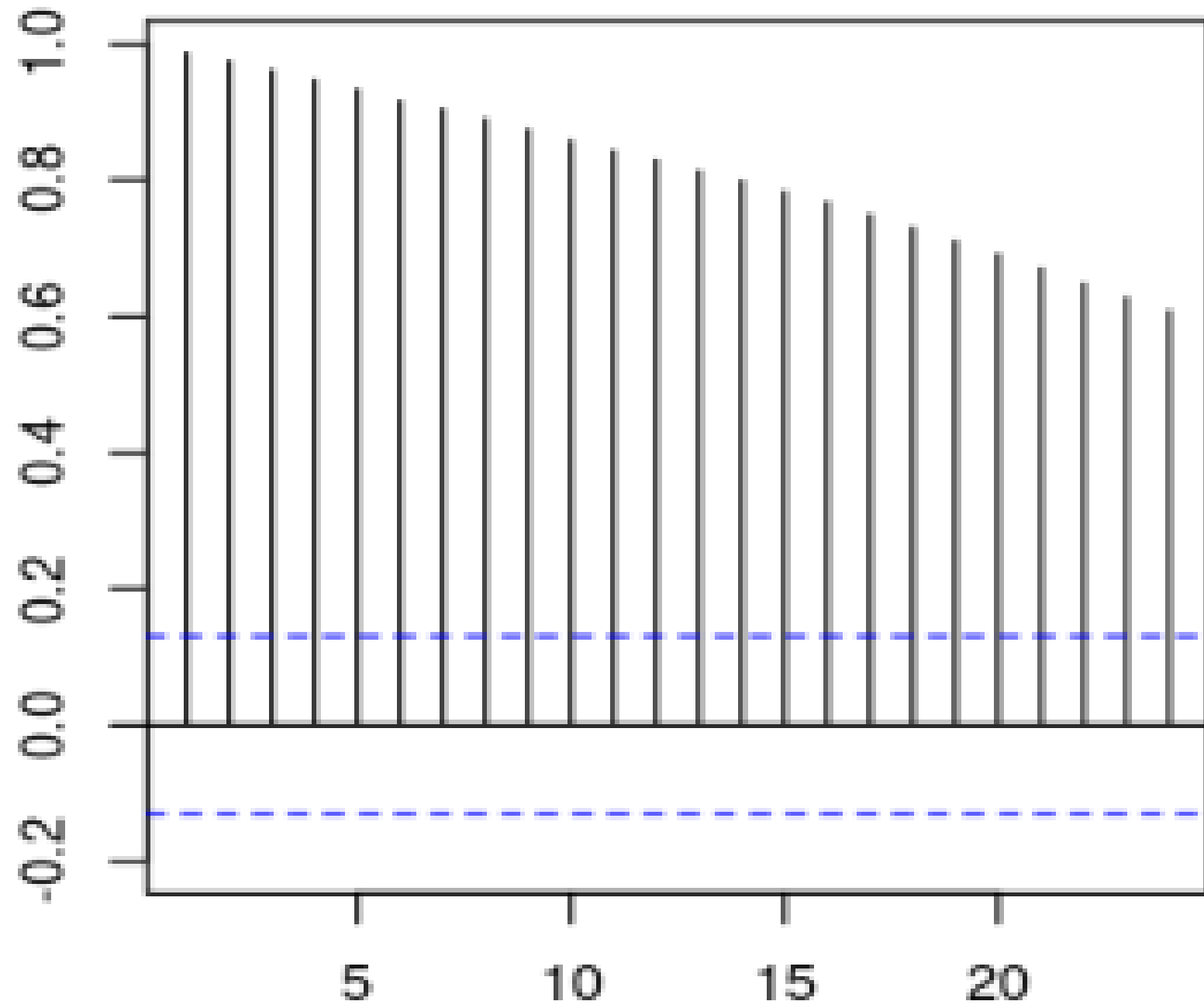
분산이 있어보인다. 분산 안정화

시도표

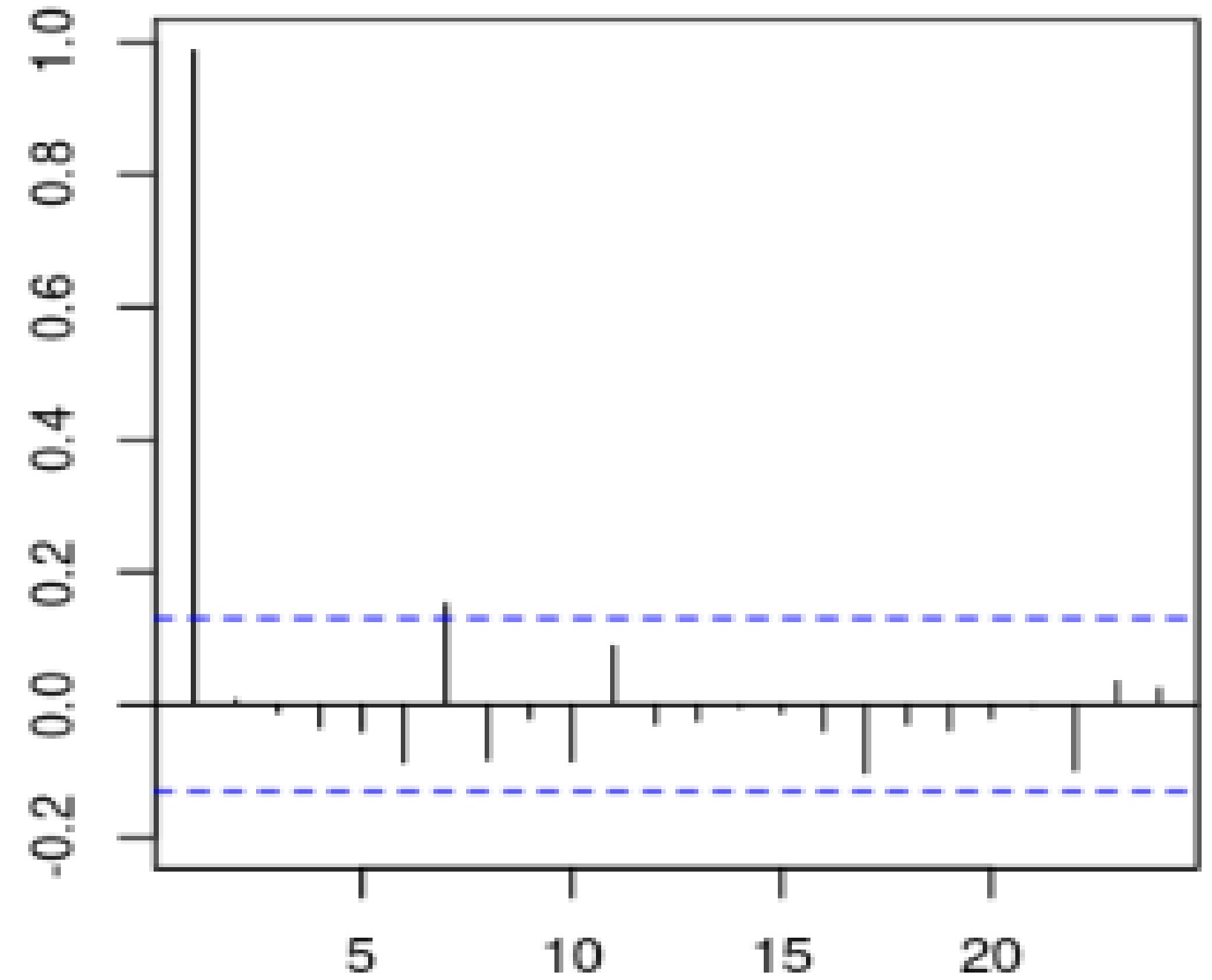


시계열 분석

ACF

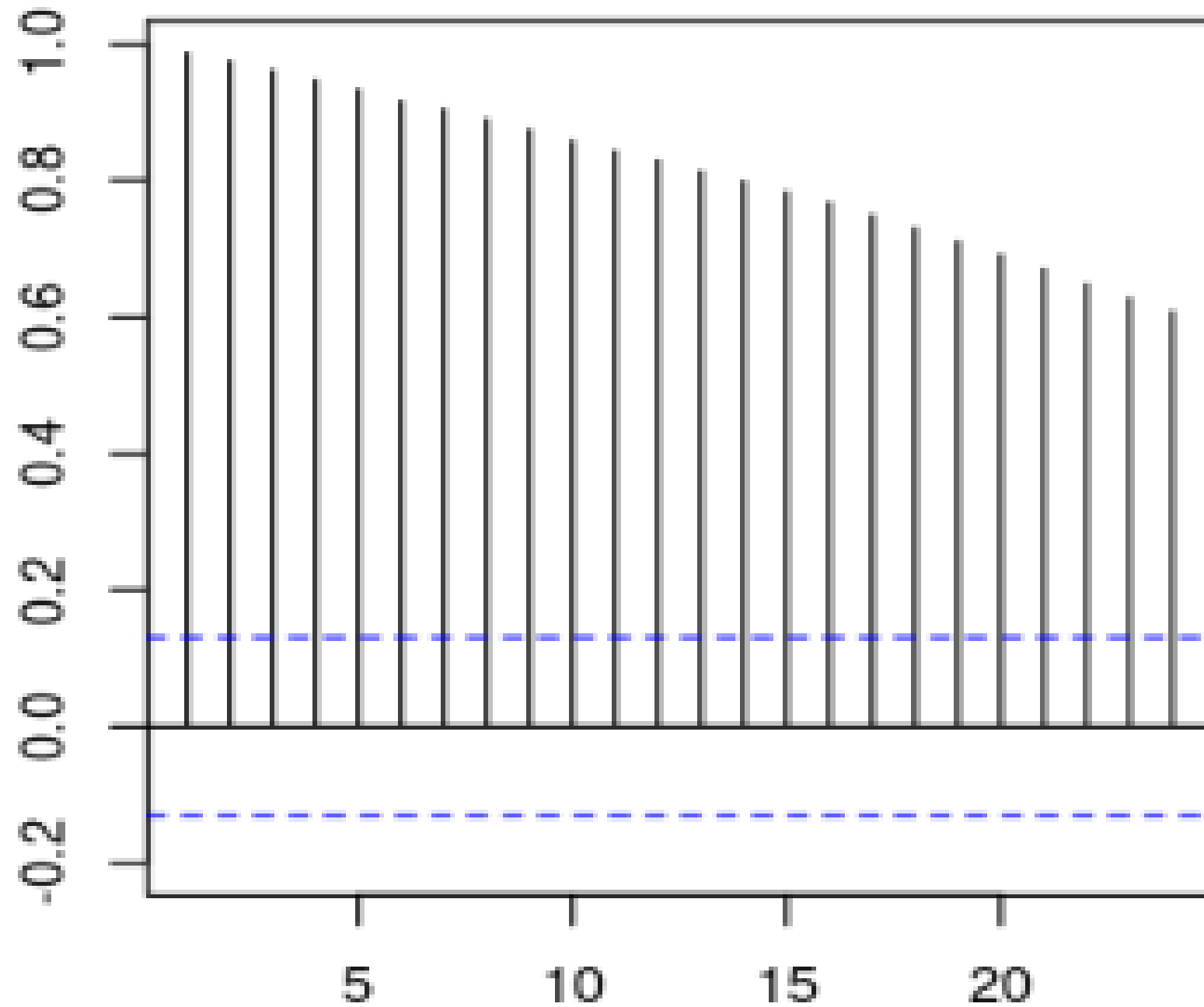


PACF



시계열 분석

ACF



ACF가 천천히 감소하고 있다.

확률적 추세가 있어보인다.

차분이 필요해보인다.

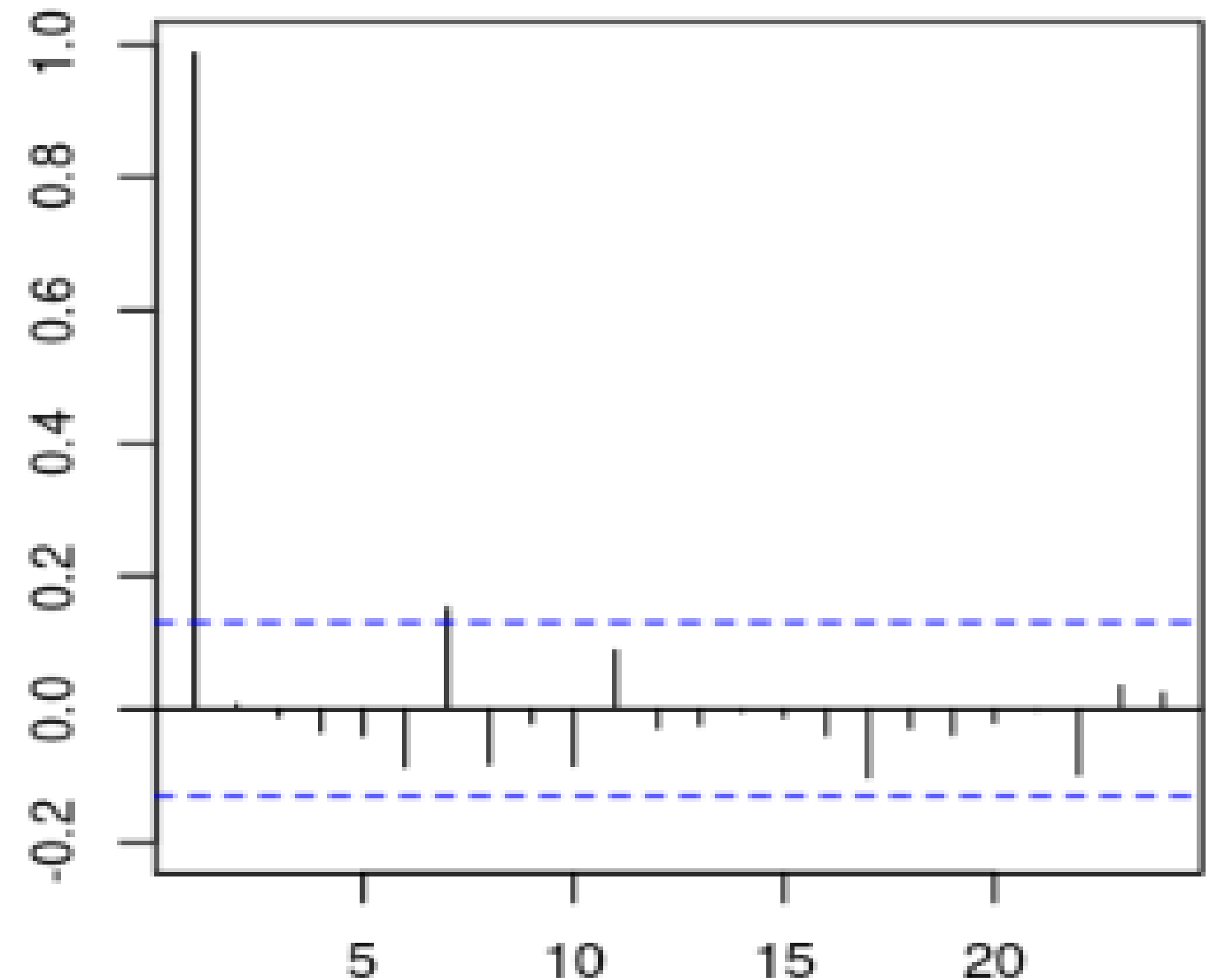
시계열 분석

PACF는 6차시에 살짝 튀어나오긴 했지만 무시할 만 하다.

1차시에 유효하다.

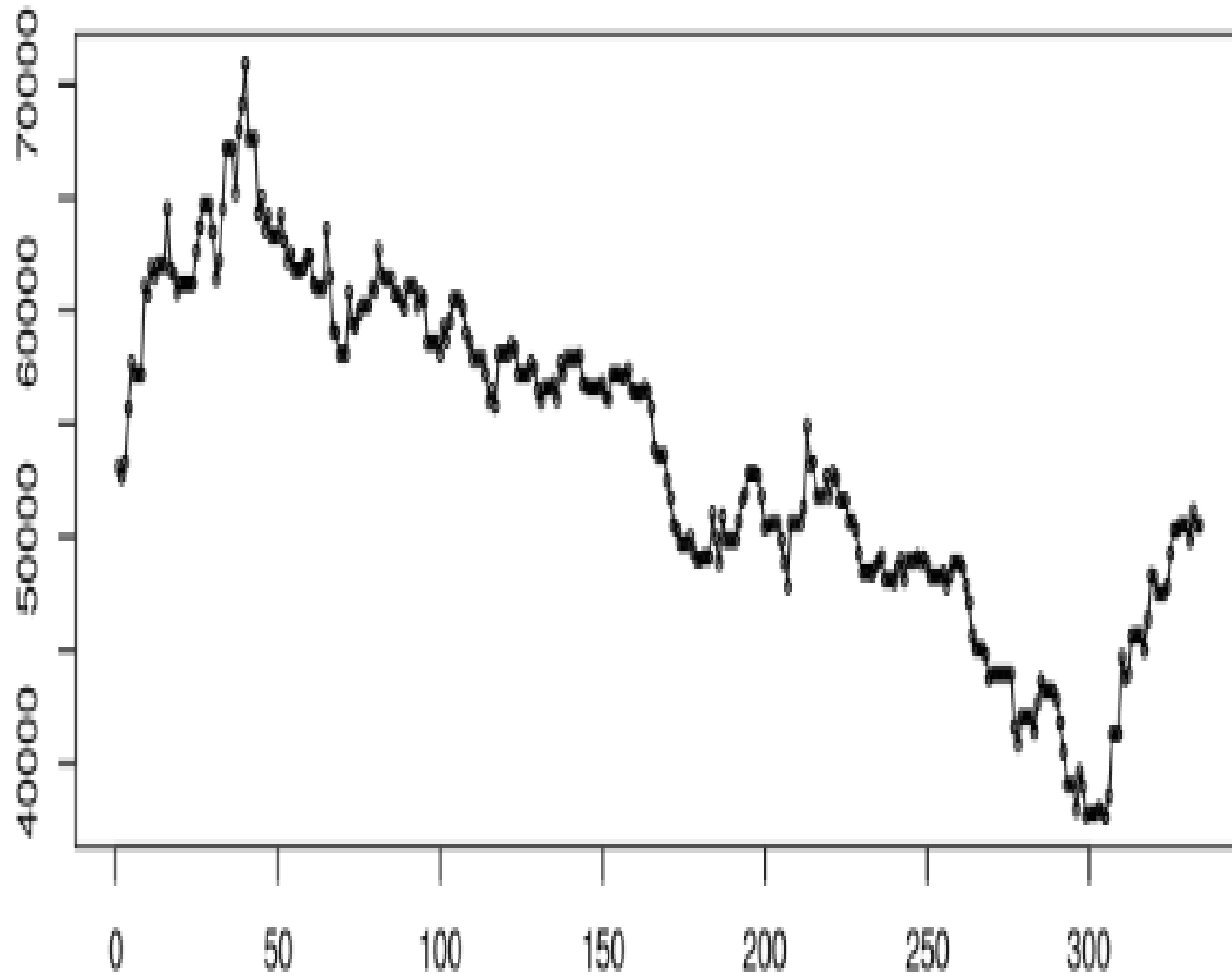
AR(1)모형을 생각해 볼 수 있을 것 같다.

PACF

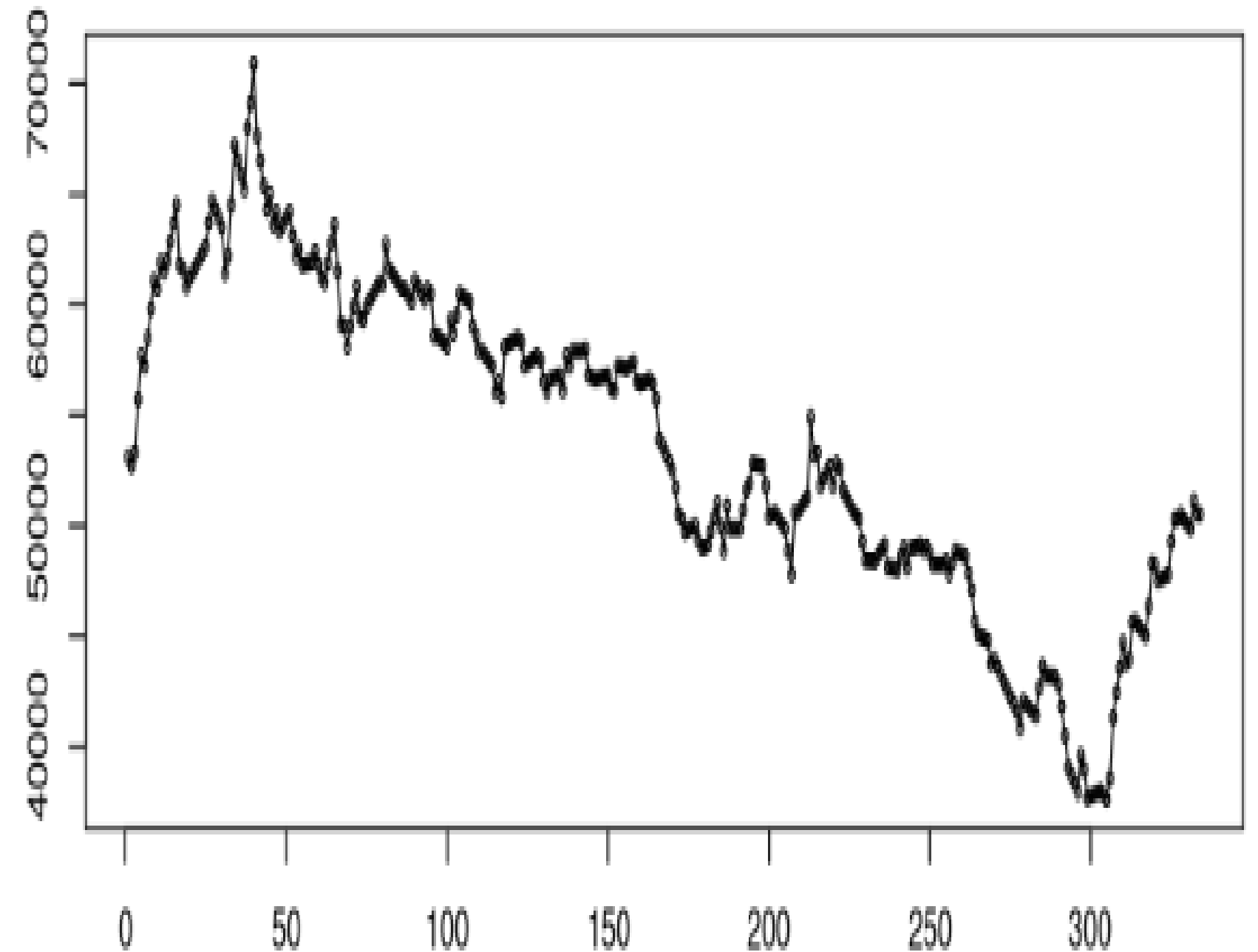


시계열 분석

앞쪽 값으로 결측값 채운 시도표



선형 보간으로 결측값 채운 시도표



시계열 분석

studentized Breusch-Pagan test

```
data: lm(df2 ~ t)
BP = 2.7869, df = 1, p-value = 0.09504
```

studentized Breusch-Pagan test

```
data: lm(log_df ~ t)
BP = 15.269, df = 1, p-value = 9.323e-05
```

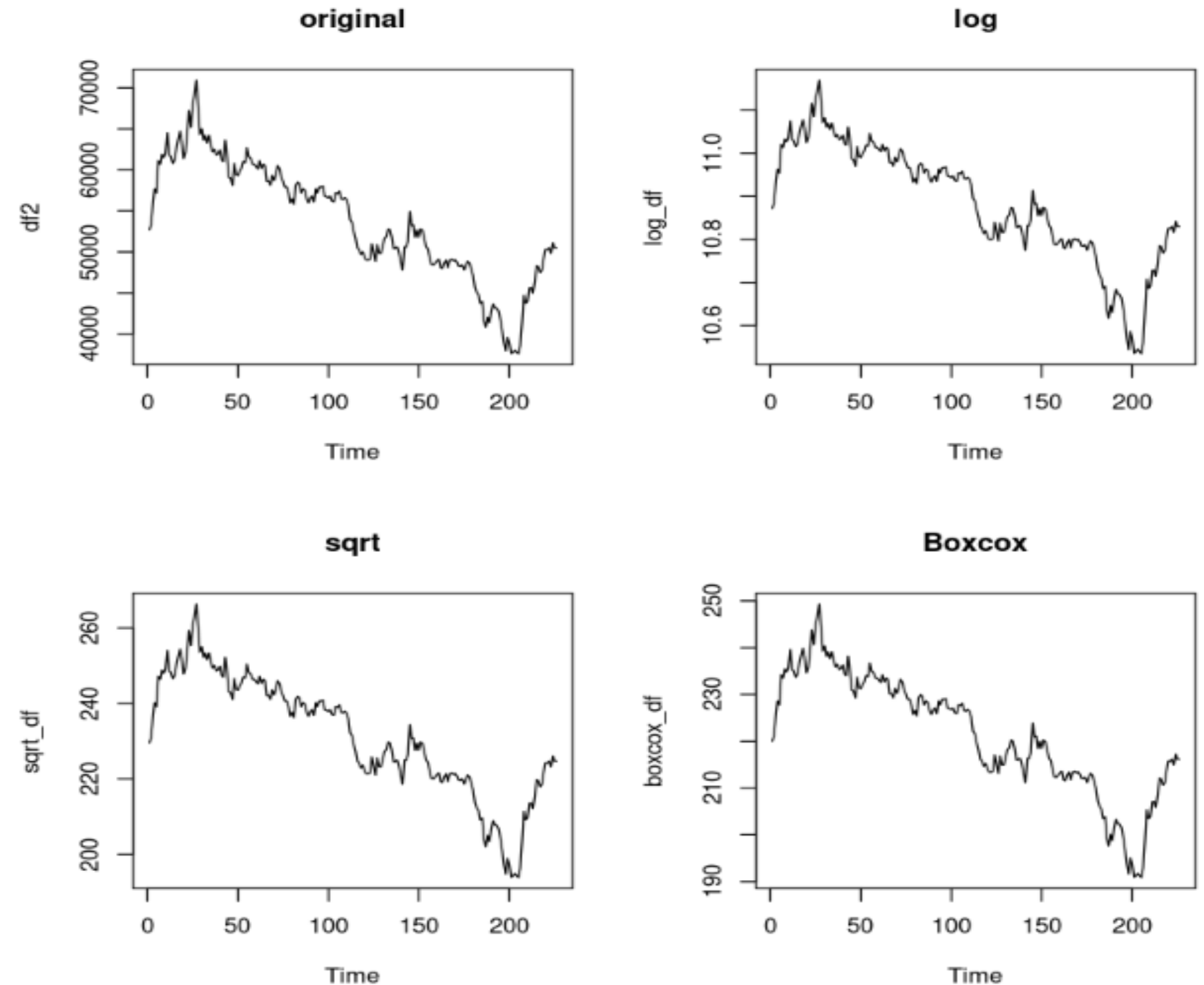
studentized Breusch-Pagan test

```
data: lm(sqrt_df ~ t)
BP = 8.2031, df = 1, p-value = 0.004182
```

studentized Breusch-Pagan test

```
data: lm(boxcox_df ~ t)
BP = 9.3125, df = 1, p-value = 0.002276
```

변수 변환



시계열 분석

ACF 그림에서 확인한 확률적 추세를

단위근 검정을 통해 재확인

1, 2, 3 시차 모두

P-value > 0.05

차분 필요

단위근 검정

Title:
Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
Lag Order: 1
STATISTIC:
Dickey-Fuller: -1.14
P VALUE:
0.6323

Description:
Wed Dec 13 22:57:07 2023 by user:

Title:
Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
Lag Order: 2
STATISTIC:
Dickey-Fuller: -1.1591
P VALUE:
0.6252

시계열 분석

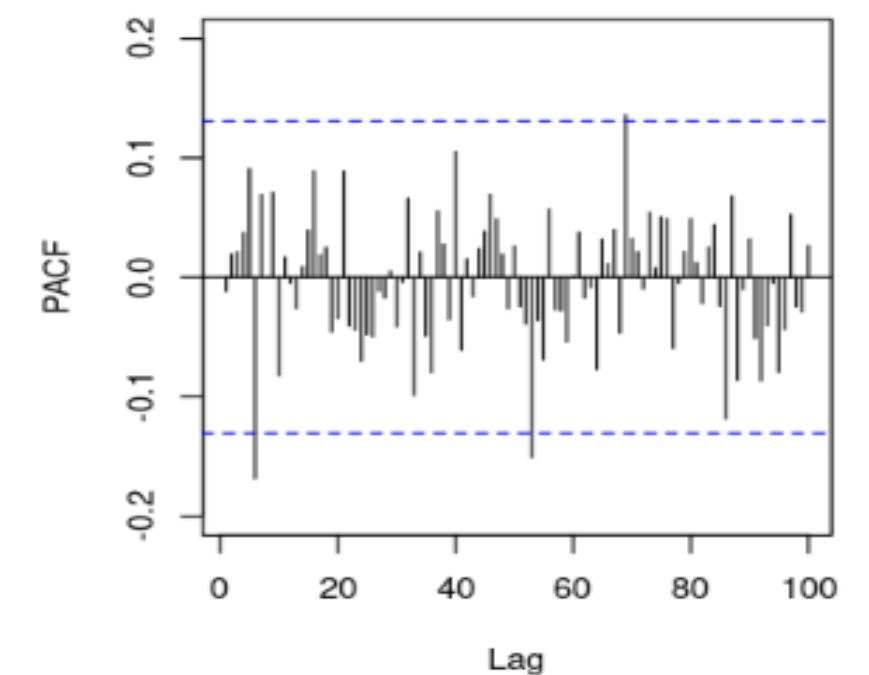
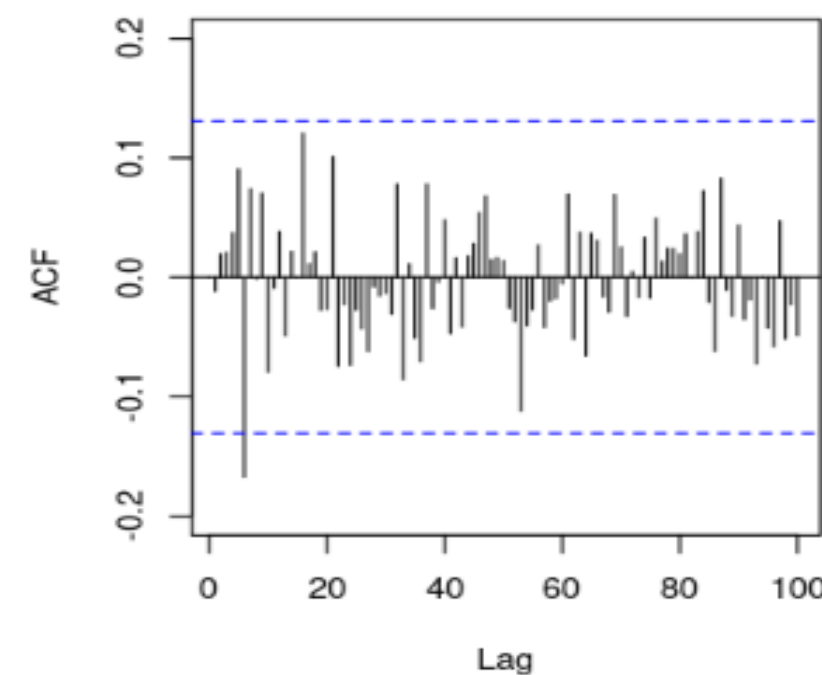
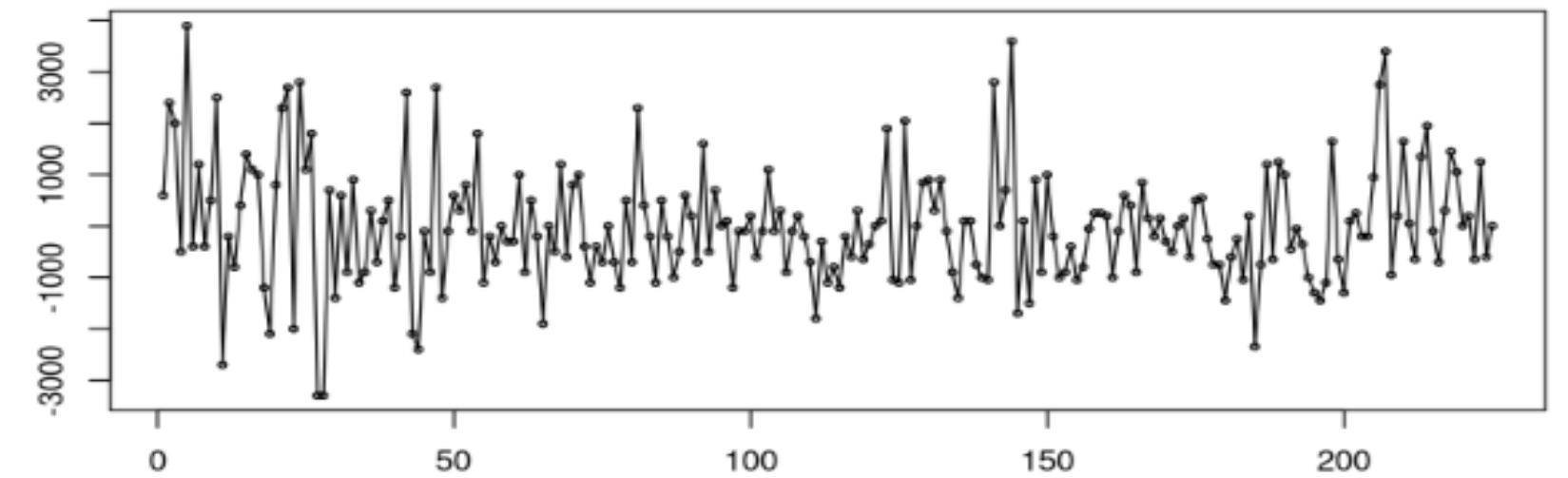
차분을 통해 평균=0이 되었음

ACF와 PACF가 몇 개 튀어나온 부분이
있지만 WN으로 보인다.

AR(1)모형을 차분 → WN

차분

diff_df2



시계열 분석

모형 식별

원 데이터: AR(1)모형
차분데이터: WN

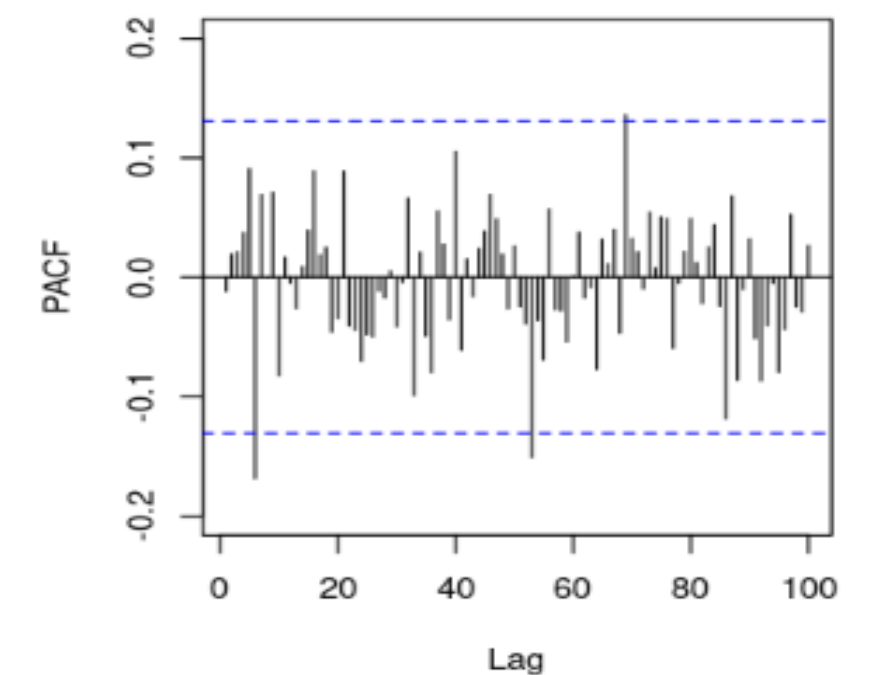
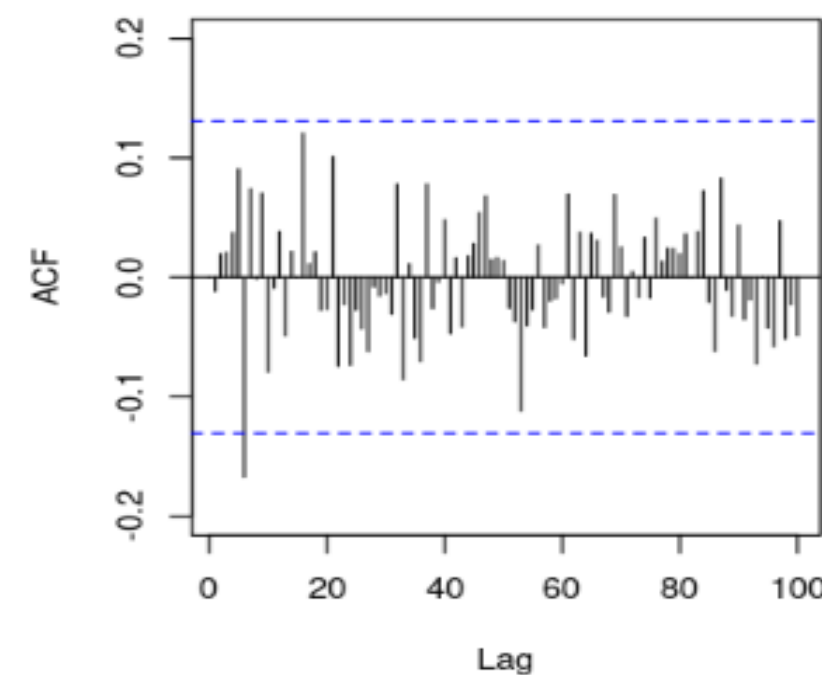
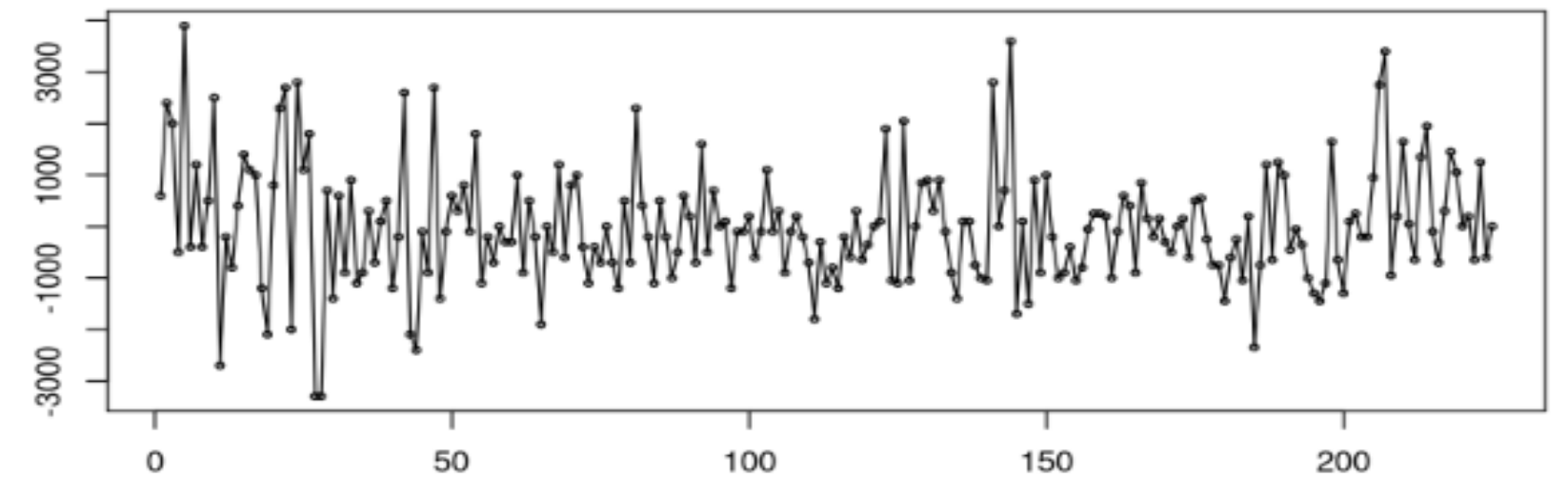
kakao: ARIMA(1,0,0)
• AIC: 3838.28

diff_kakao: ARIMA(1,1,0)
• AIC: 3889.91

auto: ARIMA(0,1,0)
• AIC: 3816.16

모형 식별

diff_df2

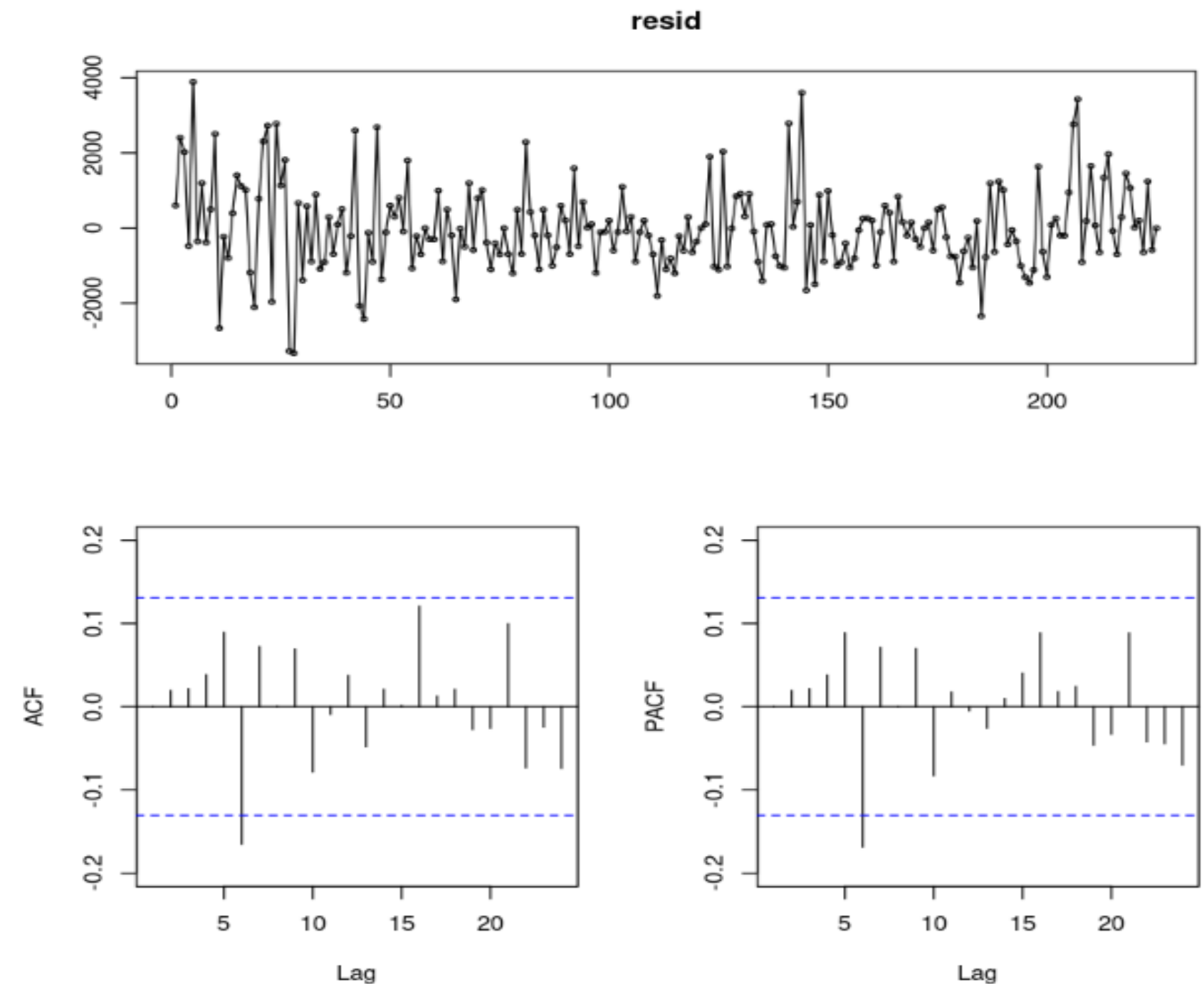


시계열 분석

AR(1) 모형의 잔차 분석

6시차에 튀어나온 부분이 있어 보이지만
WN으로 보인다.

잔차 분석



시계열 분석

포트맨토 검정

$P\text{-value} > 0.05$ 이므로

$\rho_1 = \dots = \rho_k = 0$

즉, WN이다.

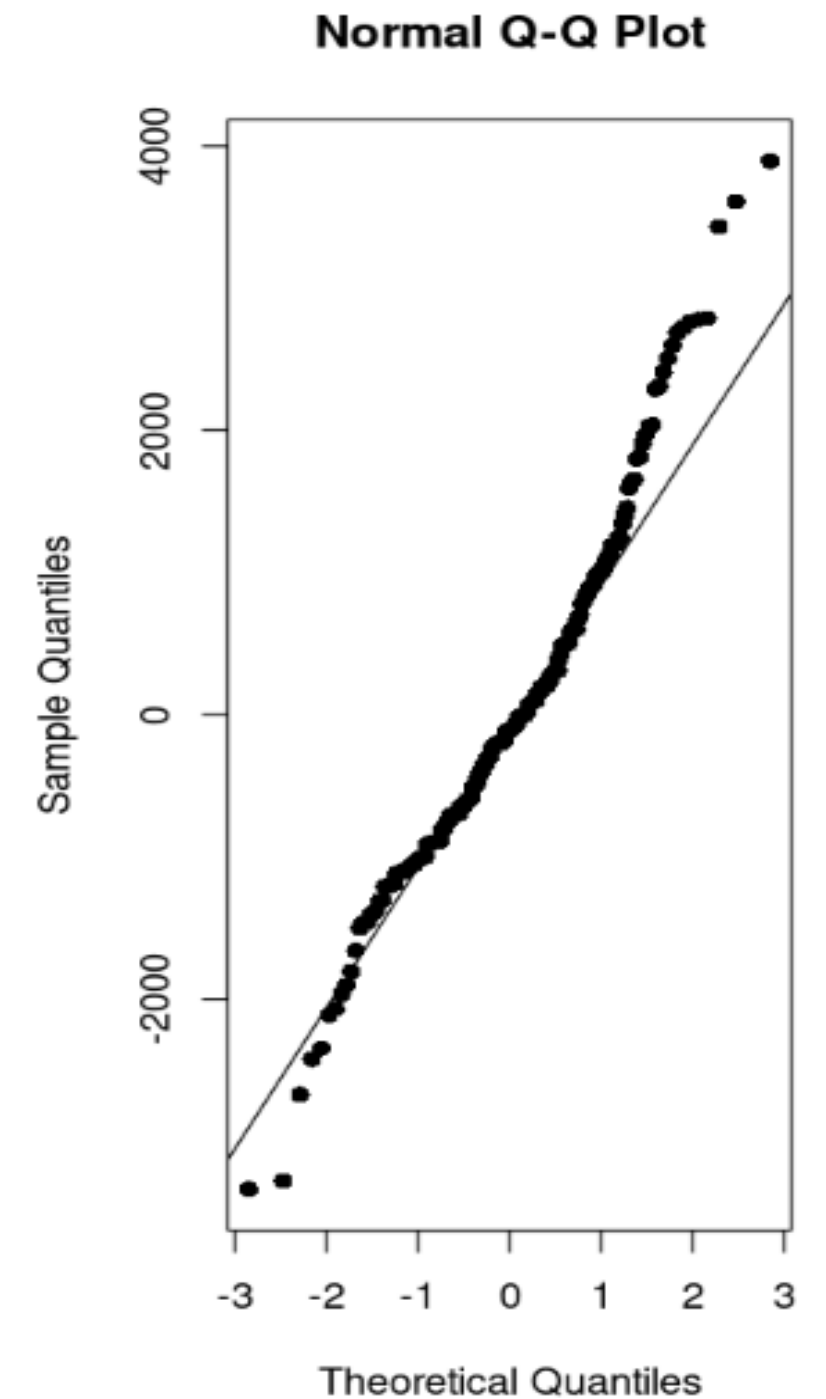
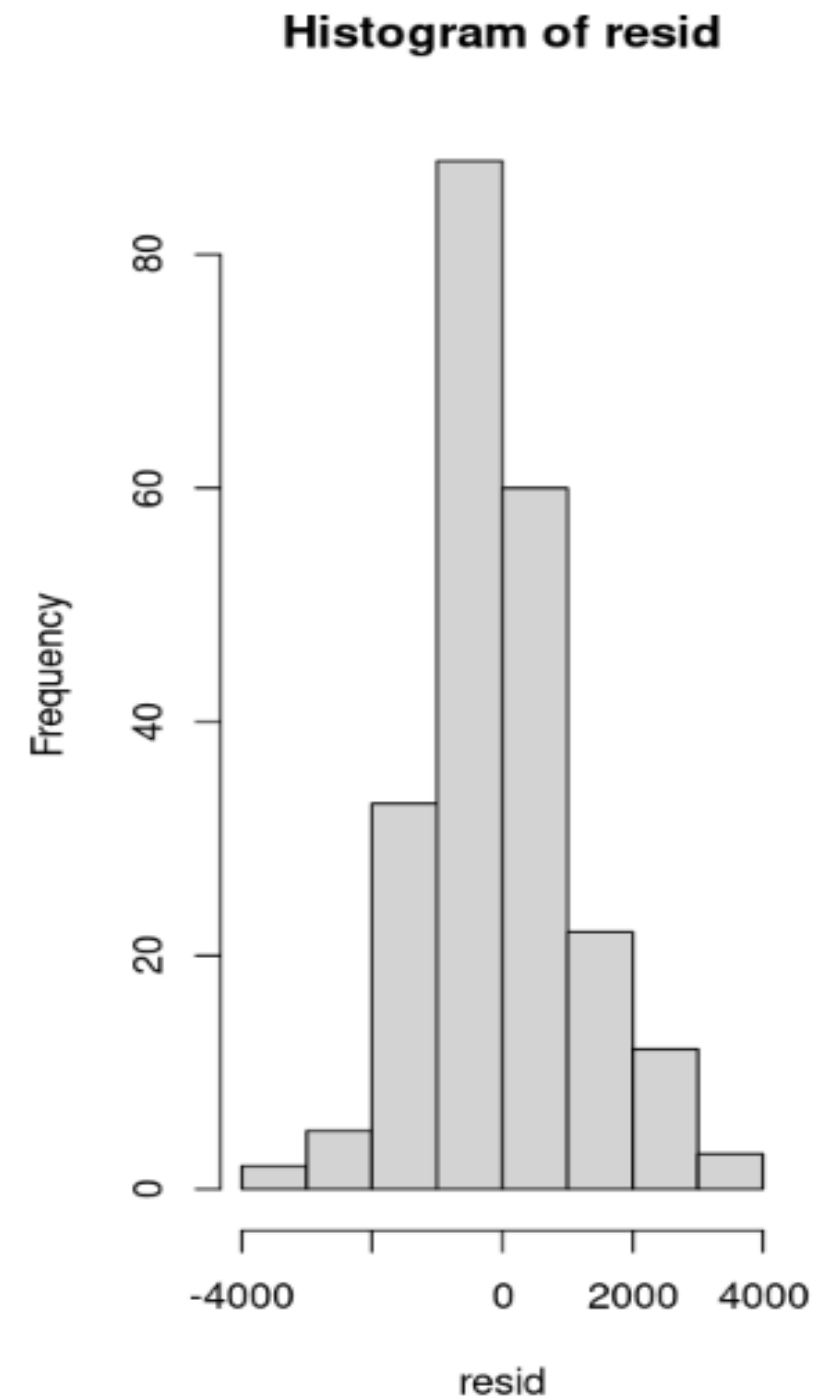
잔차 분석(포트맨토 검정)			
lags	statistic	df	p-value
6	8.877486	5	0.1140518
12	13.696612	11	0.2502380
18	18.704037	17	0.3457823
24	23.960522	23	0.4059972

시계열 분석

정규 분포가 아니다.

데이터의 이상점으로 인해 꼬리가 길다.

잔차 분석(정규분포)



예측

n=9

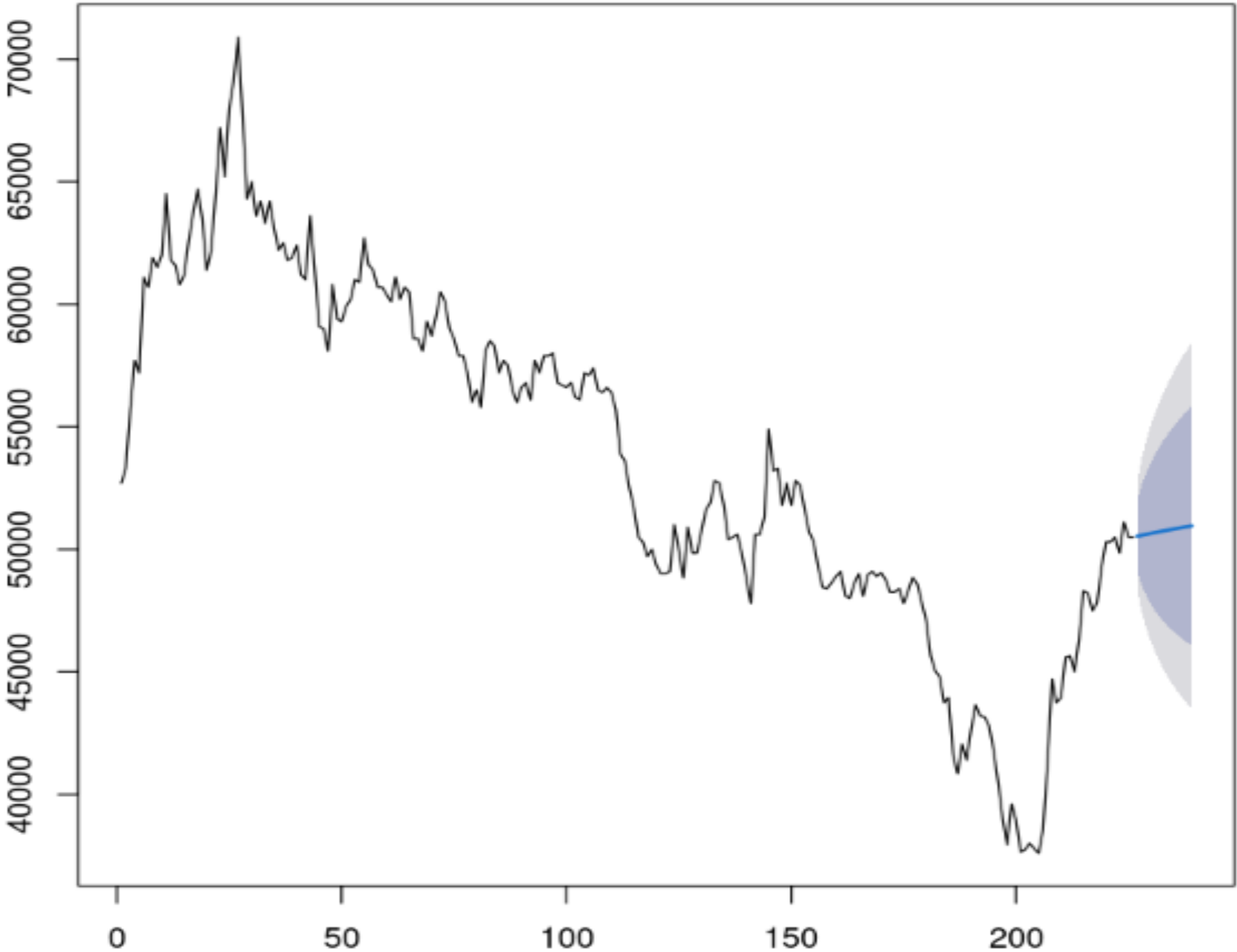
yhat

y

227	50538.54	49,700
228	50576.47	50,800
229	50613.80	50,800
230	50650.54	50,500
231	50686.71	50,500
232	50722.31	51,700
233	50757.34	52,100
234	50791.83	51,800
235	50825.77	50,900
		54,400

예측

Forecasts from ARIMA(1,0,0) with non-zero mean



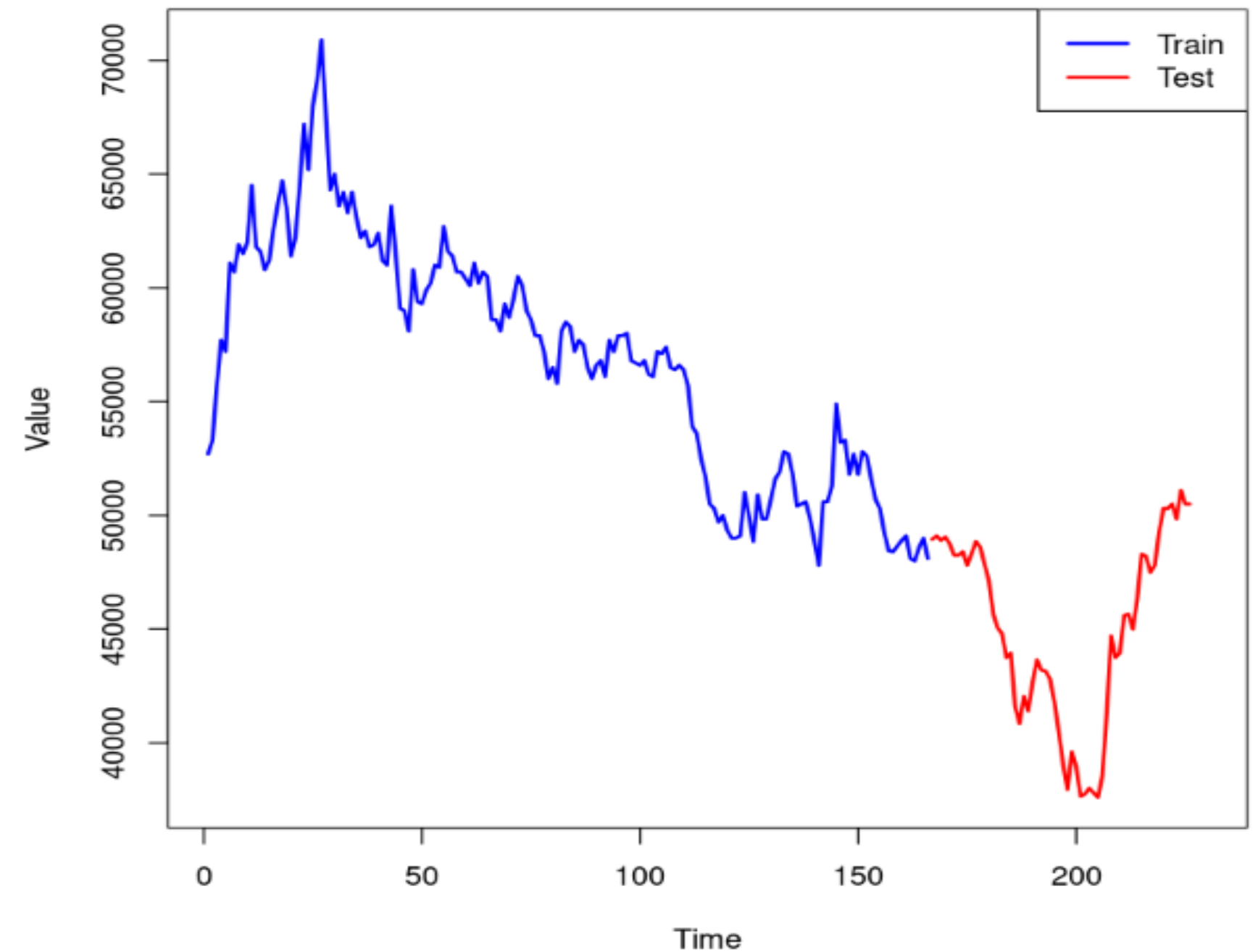
Train / Test

Train : 1 ~ 8월 166개 데이터

Test : 9 ~ 11월 60개 데이터

Train / Test Data

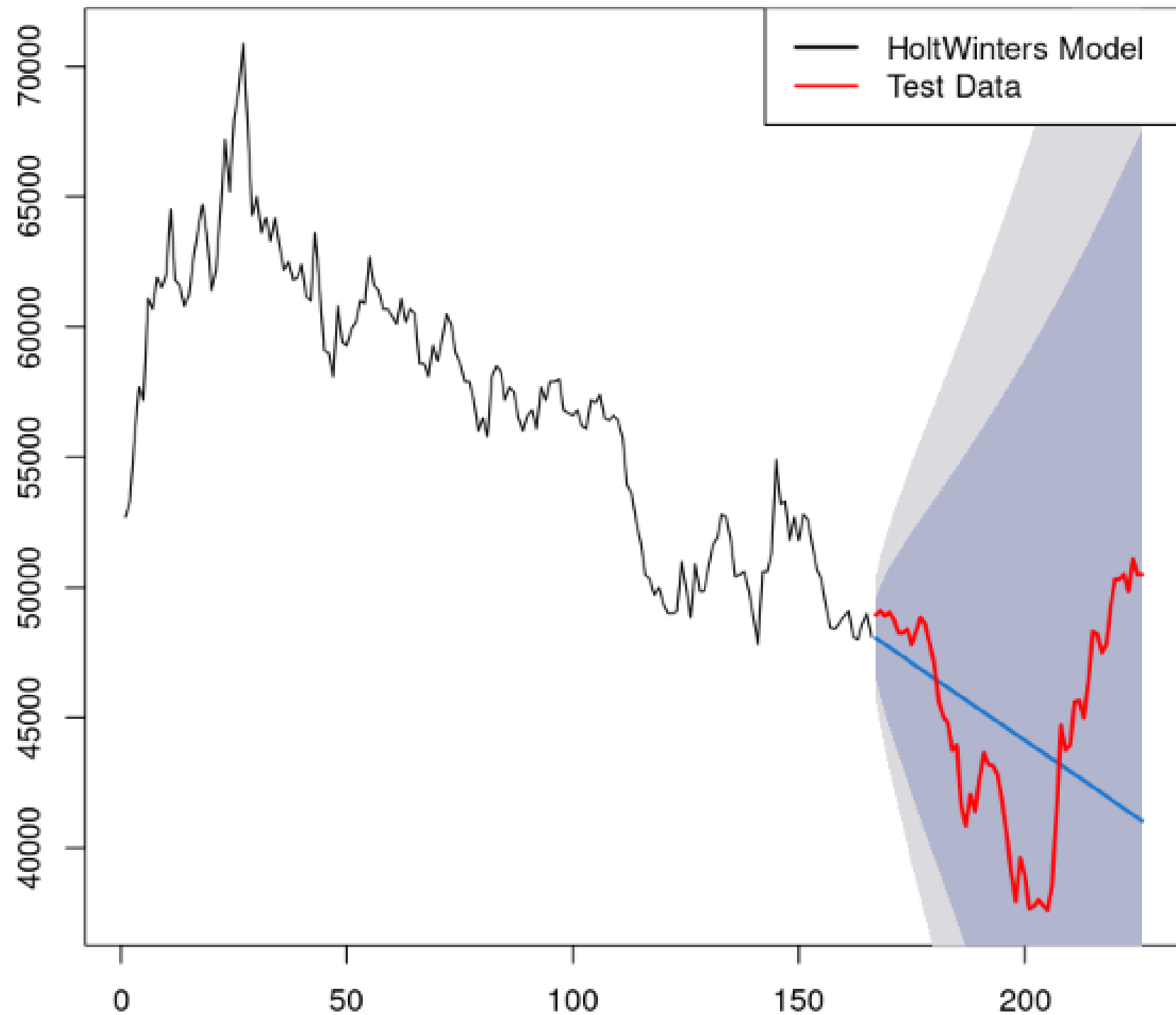
Train, Test Time Series



Train / Test

이중지수평활

HoltWinters Model and Test Data



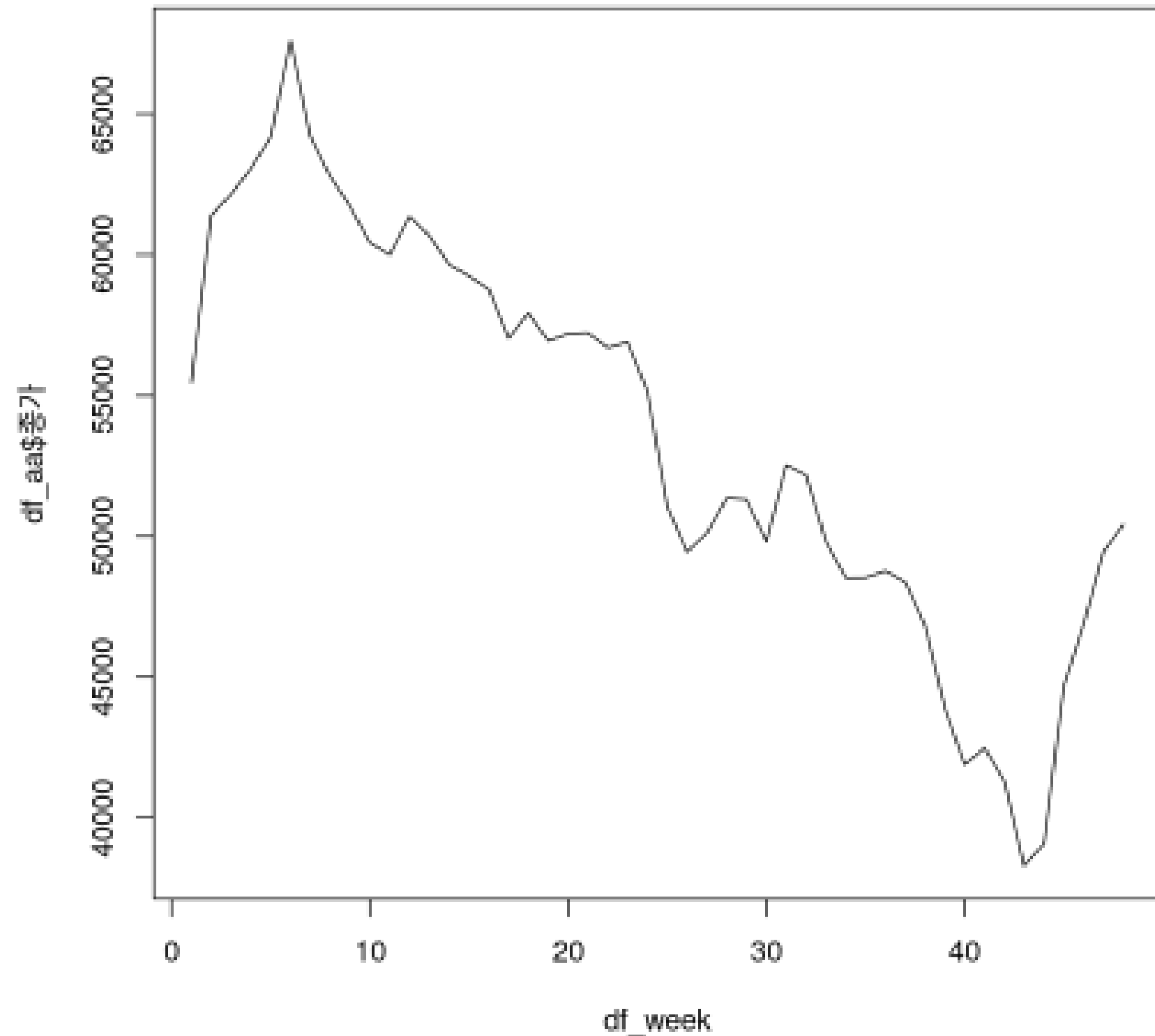
auto.ARIMA(0,1,0)

ARIMA Forecast with Test Data

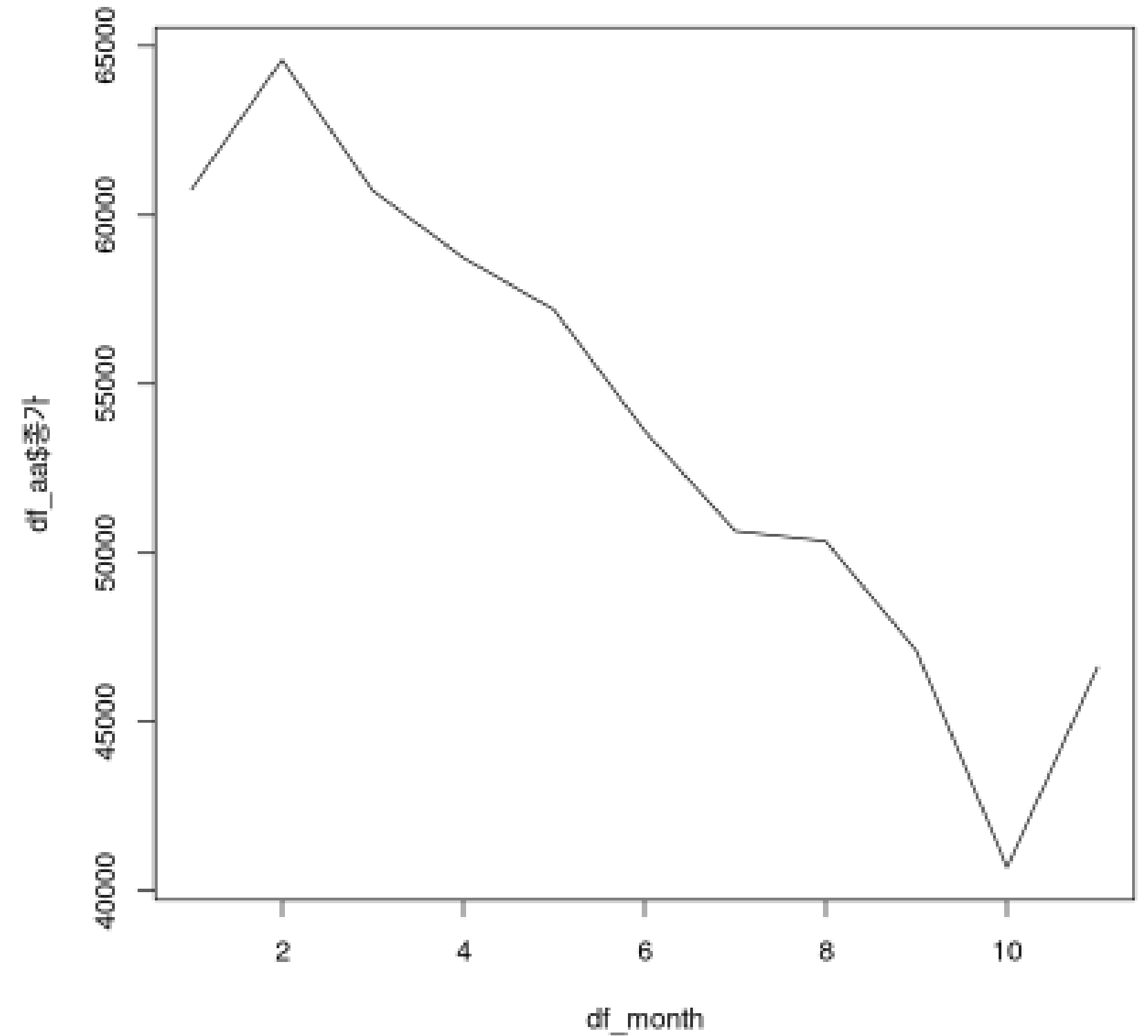


Train / Test

주별 데이터



월별 데이터



결론

주식 데이터의 시계열 분석은

불확실성이 많아 분석이 어렵다.

카카오 데이터의 경우

하락장이라서 AR모형이 적합됐지만,

장기예측을 하기에는 어려움이 있다.

KaKao



결론

(번외..)

카카오 주식 매입했던 시점에서는,

하늘까지 치솟는 줄 알았어요...

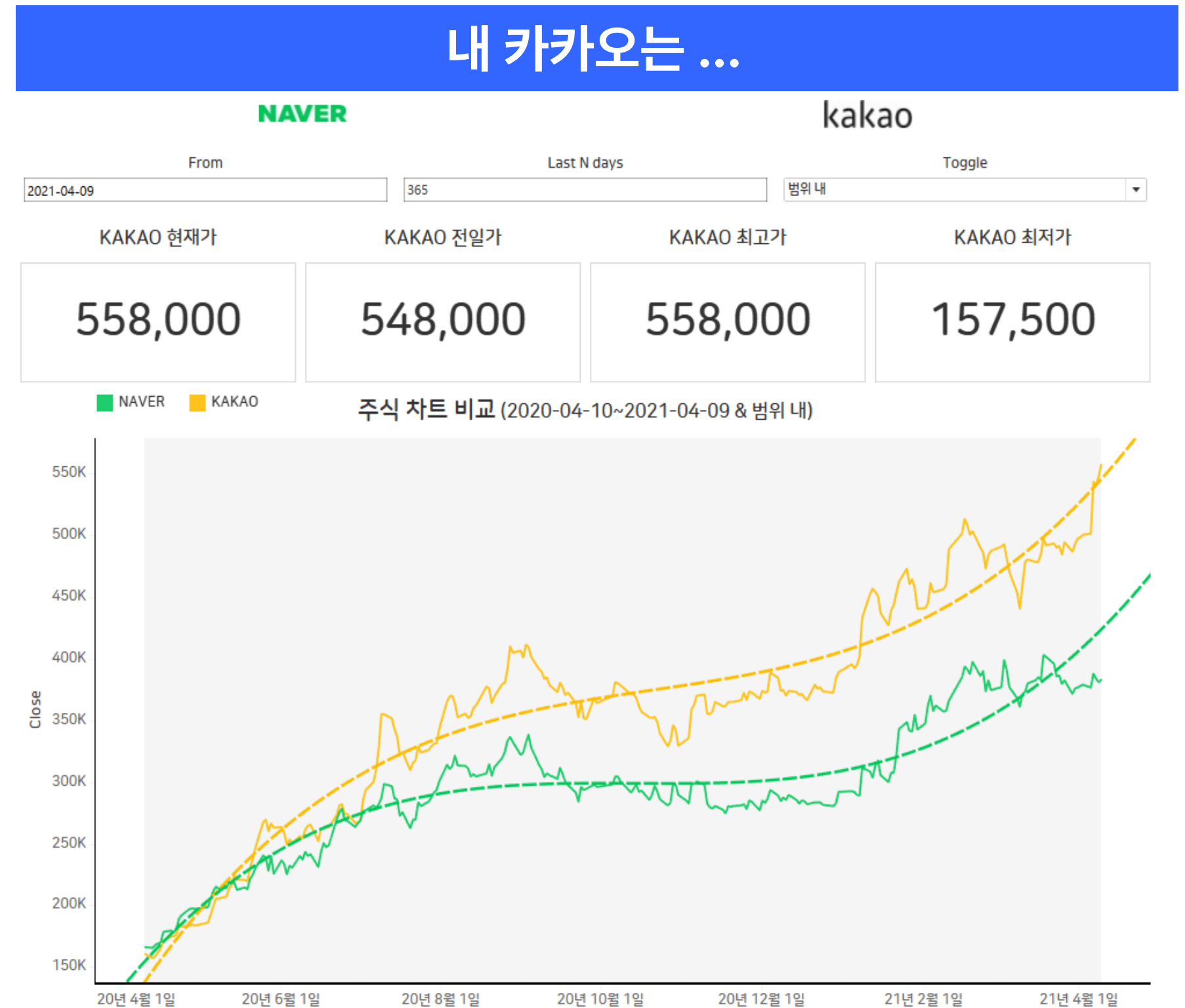
주식은 다른 시계열 데이터에 비해,

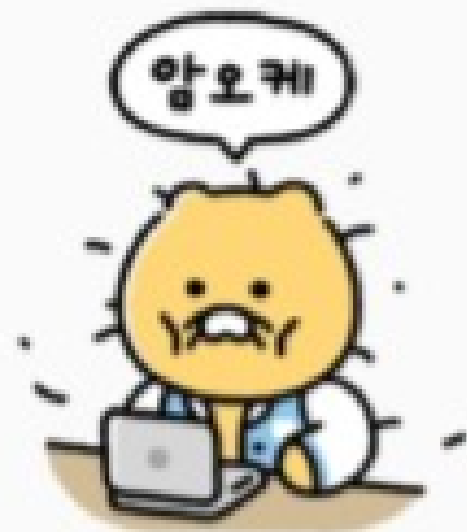
불규칙성이 너무 크므로 차트만 보고..

주식을 하지 않기..

이미지출처:

<https://vizibusy.tistory.com/31>





감사합니다!

