

# Time Series Analysis

-시계열분석-

5주차 정규세션 참고자료



투빅스 17기

신은아

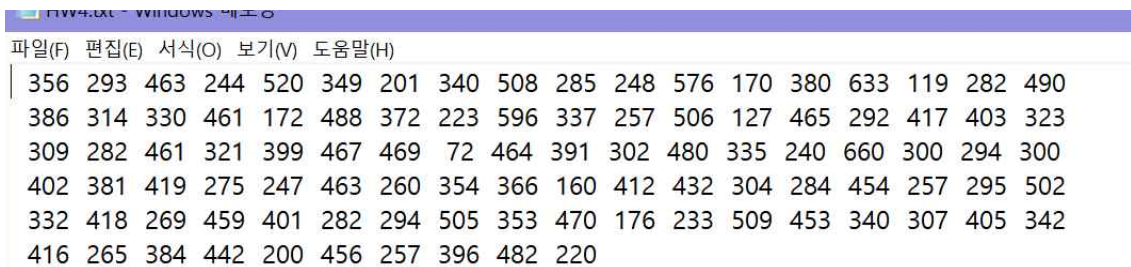
## 0. 차례

1. 서론
  - 1.1 데이터 소개
  - 1.2 분석 방법 소개
    - 1.2.1 모형의 식별
    - 1.2.2 모수의 추정
    - 1.2.3 모형의 진단
2. 본론
  - 2.1 모형의 적합
  - 2.2 과대적합 확인
3. 결론
4. 참고문헌

## 1. 서론

### 1.1 데이터 소개

본 분석에 사용할 데이터는 100개의 관측치를 가진 시계열 자료이다. 데이터는 txt 형태로 가로로 입력되어 있으며, 데이터 간의 구분자는 공백 기호다.



356	293	463	244	520	349	201	340	508	285	248	576	170	380	633	119	282	490
386	314	330	461	172	488	372	223	596	337	257	506	127	465	292	417	403	323
309	282	461	321	399	467	469	72	464	391	302	480	335	240	660	300	294	300
402	381	419	275	247	463	260	354	366	160	412	432	304	284	454	257	295	502
332	418	269	459	401	282	294	505	353	470	176	233	509	453	340	307	405	342
416	265	384	442	200	456	257	396	482	220								

### 1.2 분석 방법 소개

본 분석에서는 주어진 시계열 자료를 통해 적절한 모형을 적합함으로써 최적의 모형을 제안하고자 한다. 시계열 자료를 적합하기 위해서는 우선 시계열이 **정상적인지 여부**를 파악해야 한다. 시계열의 정상성 여부는 **시계열도**를 그려서 특별한 추세나 계절성이 있는지 확인함으로써 파악할 수 있다. 만약 시계열이 정상적이 아닌 경우에는 변환 혹은 차분 등을 통해서 정상화시켜야 한다. 정상적인 시계열자료에 모형을 적합시키는 방법은 모형의 차수를 결정하는 모형의 식별 단계, 식별된 모형의 모수들을 추정하는 모수의 추정 단계, 잠정모형의 타당성을 검토하여 최종 적합여부를 판단하는 모형의 진단 단계의 세 단계로 나뉜다.

### 1.2.1 모형의 식별

모형의 식별 단계에서는 ARMA(p,q) 모형에서 p,q를 결정한다. 이 때 모수의 절약과 추정의 효율성을 위해 간단한 모형을 선택하자는 간결의 원칙을 따른다. 모형의 차수는 시계열자료로부터 계산된 SACF와 SPACF를 통해 결정할 수 있다. AR(p) 과정은 p 이후의 시차에서 PACF 값이 0이 되며, MA(q) 과정은 q 이후의 시차에서 ACF 값이 0이 된다.

### 1.2.2 모수의 추정

모형의 식별을 통해 잠정적으로 모형이 선택되면 모수를 추정해야 한다. 모수의 추정방법은 적률추정법, 조건부 최소제곱추정법, 최대가능도추정법 등 다양하다.

### 1.2.3 모형의 진단

이 단계에서는 주로 잔차의 분석을 통해 잠정모형이 타당한지 여부를 판단한다. 잠정모형에 이상이 없다고 판단되면 이 모형을 예측모형으로 이용하고, 잠정모형이 타당하지 않으면 다시 모형의 식별 단계로 돌아가 새로운 잠정모형을 선택한다. 모형의 진단 단계에서는 주로 잔차분석과 과대적합 등의 분석을 수행한다. 잔차진단은 잔차들의 정규성검정이나 정규확률그림을 통해 잔차들이 정규분포를 따르는지 알아보고, 잔차의 시계열도를 그려보거나 잔차의 표본자기상관계수가 0인지 유의성검정을 통해 백색잡음 가정을 확인한다. 과대적합은 잠정모형에 모수를 추가하여 더 많은 개수의 모수를 포함하는 모형을 적합시키는 것이다. 잠정모형에 추가된 모수가 유의하다고 판정되거나 잠정모형에서의 모수추정값이 과대적합 후 얻어지는 모수추정값과 큰 차이가 있다면, 또는 과대적합된 모형에서 잔차들의 분산이 잠정모형에서 잔차들의 분산보다 작아졌다면 추가된 모수들에 설명력이 있다고 판단해 잠정모형을 새로운 모형으로 대체한다.

## 2. 본론

본 분석에서는 통계 프로그램 SAS를 이용해 시계열분석을 진행했으며, 분석에 사용한 SAS 코드는 아래와 같다.

```
data hw4;
infile 'c:/Users/Ss/Desktop/2021/시계열/HW4.txt';
input z @@; t+1; run;
```

```
/*시계열도*/
```

```
proc sgplot;
series x=t y=z; run;
```

```
proc arima;
identify var=z; run;
estimate q=1 method=cls; run; /*ma1 모형추정;
estimate q=3 method=cls; run; /*ma3 모형추정;
forecast lead=0 out=res; run; quit;
```

```
data res; set res; time=_n_; run;
```

```
/*잔차의 시계열도*/
```

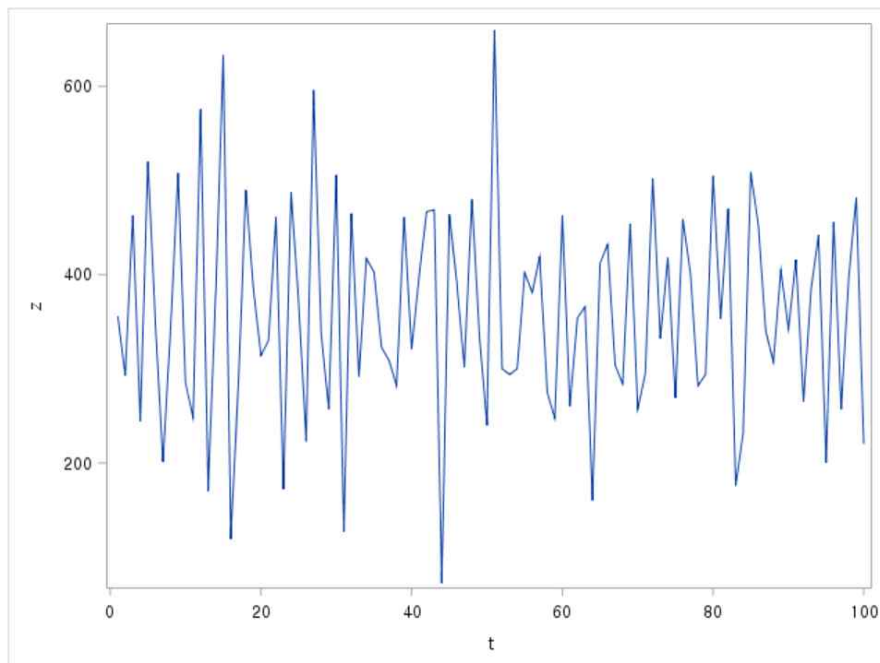
```
proc sgplot;
series x=time y=residual; refine 0; run;
```

```
/*과대적합*/
```

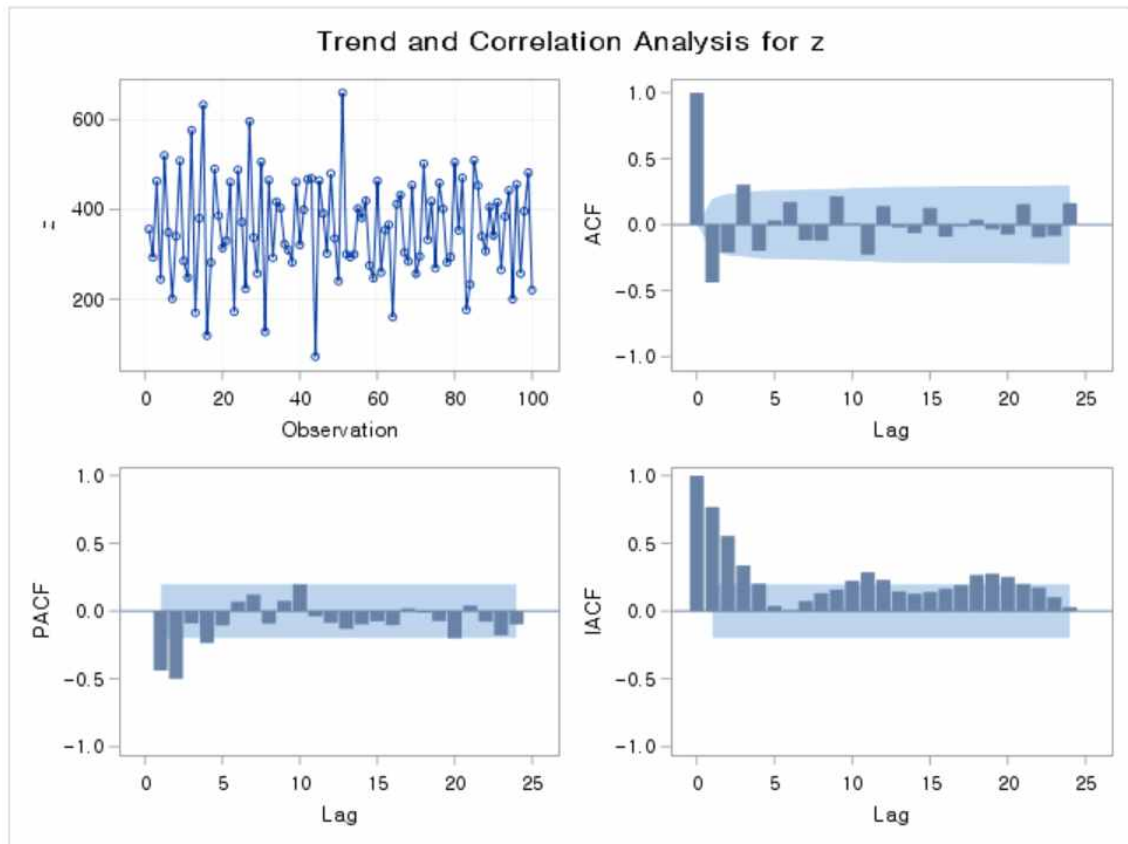
```
proc arima data=hw4;
identify var=z noprint; run;
estimate q=4; run;
estimate p=1 q=3; run;
```

## 2.1 모형의 적합

우선 주어진 자료의 정상성을 확인하기 위해 시계열도와 상관도표를 그려보면 다음과 같다.



<자료 1> 원계열  $Z_t$ 의 시계열도



<자료 2> 원계열  $Z_t$ 의 상관도표 (SACF, SPACF, SIACF)

시계열도를 보면 증가하는 추세나 계절요인이 존재하지 않음을 알 수 있다. 또한 자료의 SACF가 천천히 감소하는 형태를 보이지 않기 때문에 해당 데이터는 정상 시계열이라고 판단할 수 있다.

Autocorrelation Check for White Noise										
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations						
6	41.66	6	<.0001	-0.439	-0.211	0.304	-0.198	0.031	0.171	
12	58.39	12	<.0001	-0.119	-0.122	0.216	0.007	-0.229	0.142	
18	62.07	18	<.0001	-0.020	-0.064	0.127	-0.091	-0.015	0.037	
24	71.86	24	<.0001	-0.034	-0.075	0.156	-0.097	-0.085	0.164	

<자료 3> 원계열  $Z_t$ 의 백색잡음검정

또한, <자료 3>을 통해 주어진 자료가 백색잡음과정이라는 귀무가설을 기각할 수 있

으므로 ARMA 모델을 적합하기로 한다.

ARMA 모형의 차수를 식별하기 위해 <자료 2>의 상관도표를 확인하자. 자료의 ACF가 1-시차 이후로 대체적으로 0인 경향을 가지고 있으며, PACF와 IACF가 지수적으로 감소하는 형태를 가지고 있으므로 MA(1) 모형의 적합을 생각해볼 수 있다. 우선 MA(1)을 잠정모형으로 놓고 모수를 추정해보자.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
MU	358.55649	2.52261	142.14	<.0001	0
MA1,1	0.72724	0.06930	10.49	<.0001	1

Constant Estimate	358.5565
Variance Estimate	8182.097
Std Error Estimate	90.45494
AIC	1186.738
SBC	1191.948
Number of Residuals	100

<자료 4> MA(1) 적합 결과

모평균이 포함된 MA(1) 모형을 조건부 최소제곱법에 의해 추정한 결과는 <자료 4>와 같다. 모든 모수가 유의한 값을 가지기 때문에 적합된 모형은 다음과 같다.

$$Z_t - 358.56 = \varepsilon_t - 0.7272\varepsilon_{t-1}$$

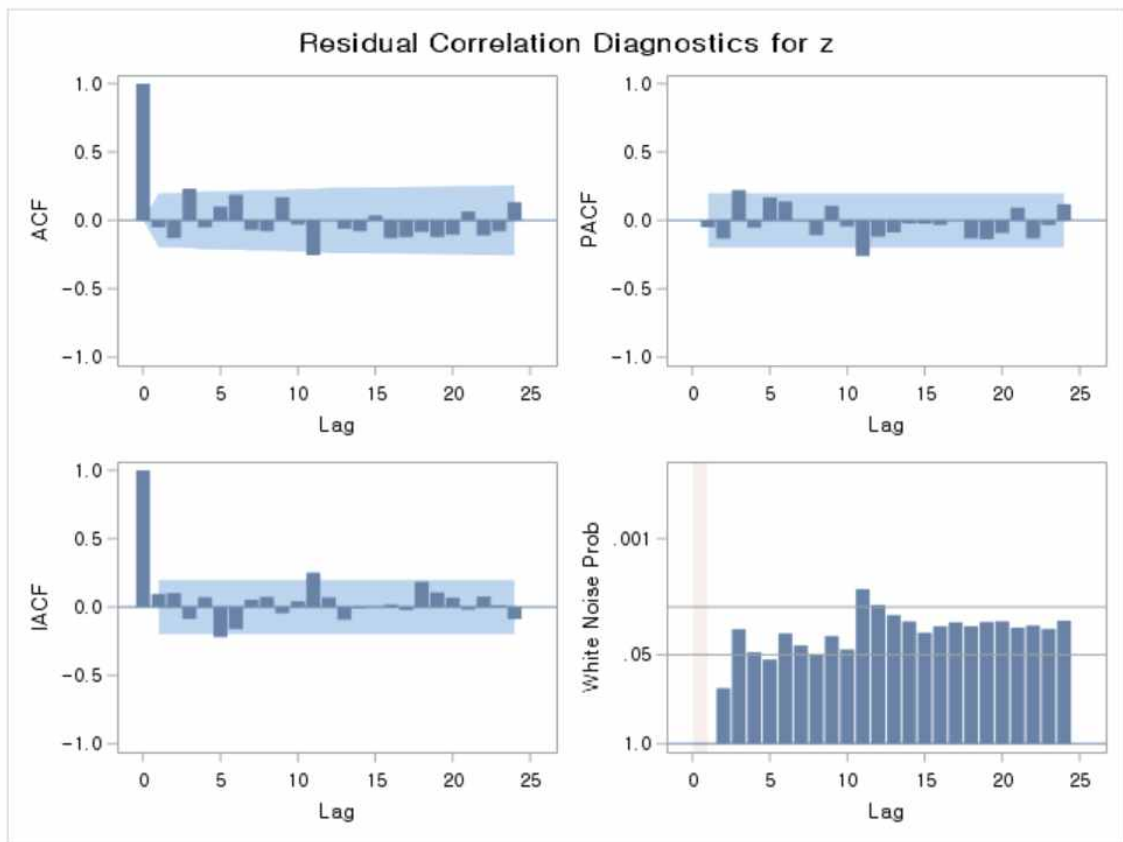
$$\hat{\sigma}_\varepsilon^2 = 8182.097$$

그러면 해당 모형으로의 적합이 적절한지 알아보기 위해 모형의 진단 과정을 진행해보자.

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	12.90	5	0.0243	-0.051	-0.130	0.233	-0.053	0.100	0.187
12	24.95	11	0.0093	-0.071	-0.080	0.169	-0.031	-0.255	-0.001
18	31.16	17	0.0191	-0.062	-0.080	0.036	-0.132	-0.123	-0.085
24	39.89	23	0.0158	-0.122	-0.106	0.066	-0.111	-0.081	0.134

<자료 5> MA(1) 모형 적합 후의 포트맨토검정 결과

포트맨토검정 결과를 확인하면 잔차가 백색잡음과정을 따른다는 귀무가설을 기각한다는 결론을 내릴 수 있다. 따라서 MA(1) 모형을 적합한 결과 잔차가 백색잡음과정을 따르지 않기 때문에 적절한 모형 적합이라고 판단하기 어렵다.



<자료 6> MA(1) 잔차들의 자기상관계수

또한, 잔차의 자기상관계수들인 RSACF와 RSPACF를 확인하면 3-시차에서 모두 유의한 값을 나타내고 있음을 확인할 수 있다. 따라서, MA(1) 모형을 버리고 새로운 잠정모형인 MA(3) 모형을 설정하여 모수를 추정해보자.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
MU	358.51821	3.55413	100.87	<.0001	0
MA1,1	0.75689	0.09465	8.00	<.0001	1
MA1,2	0.21016	0.12052	1.74	0.0844	2
MA1,3	-0.37694	0.09547	-3.95	0.0002	3

Constant Estimate	358.5182
Variance Estimate	7547.264
Std Error Estimate	86.87499
AIC	1180.6
SBC	1191.02
Number of Residuals	100

<자료 7> MA(3) 적합 결과

모평균이 포함된 MA(3) 모형을 조건부 최소제곱법에 의해 추정한 결과는 <자료 7>와 같다.  $\mu$ 와  $\theta_1, \theta_3$ 은 유의한 값을 보이고 있으며,  $\theta_2$ 는 제한적으로 유의미한 값을 나타낸다. 적합된 모형은 다음과 같다.

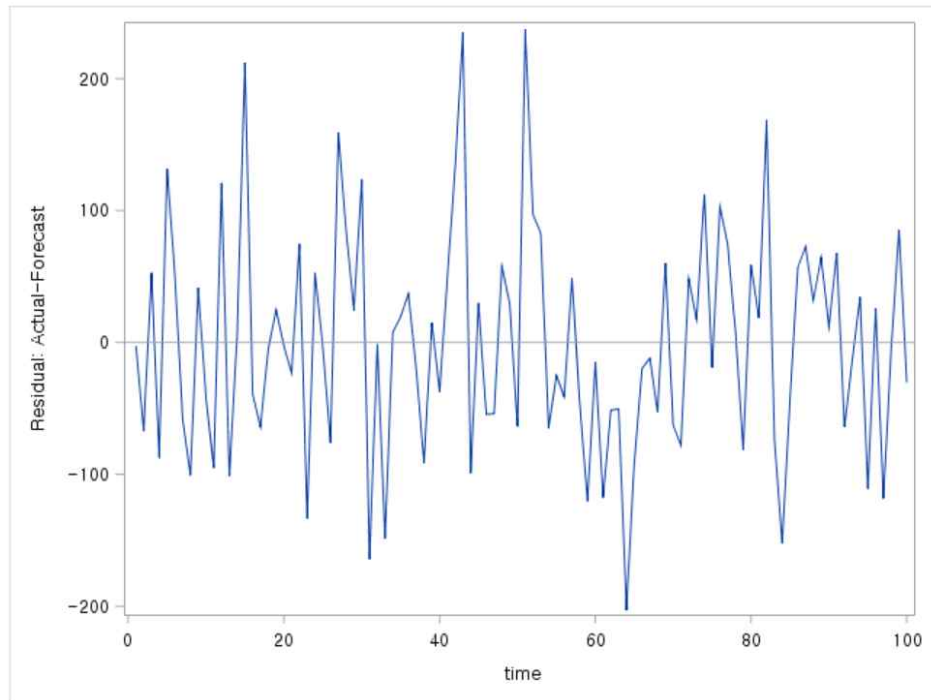
$$Z_t - 358.52 = \varepsilon_t - 0.7569\varepsilon_{t-1} - 0.2102\varepsilon_{t-2} + 0.3769\varepsilon_{t-3}$$

$$\hat{\sigma}_\varepsilon^2 = 7547.264$$

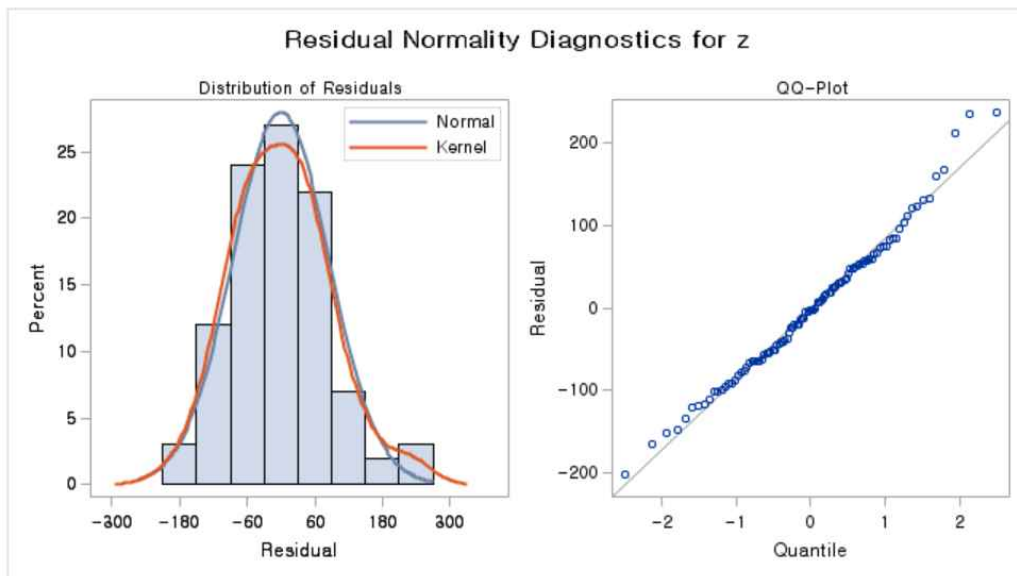
이전에 적합했던 MA(1) 모형보다 분산 추정 결과도 작고, AIC, SBC값도 작음을 확인할 수 있다.

그러면 해당 모형으로의 적합이 적절한지 알아보기 위해 모형의 진단 과정을 진행해 보자.





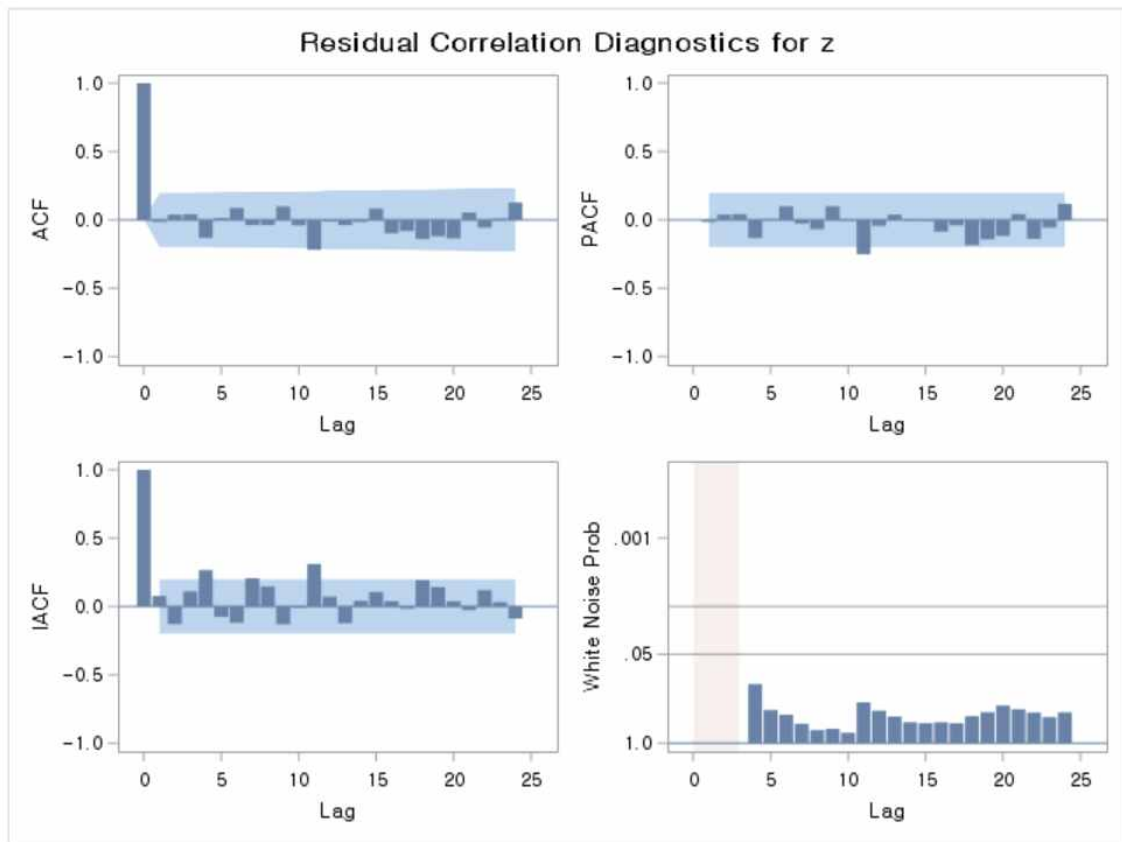
<자료 8> MA(3) 모형의 잔차의 시계열도



Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.06	3	0.3829	-0.016	0.038	0.041	-0.133	0.011	0.087
12	10.21	9	0.3337	-0.039	-0.039	0.097	-0.041	-0.220	-0.013
18	15.73	15	0.4003	-0.039	-0.016	0.083	-0.098	-0.081	-0.142
24	22.85	21	0.3520	-0.120	-0.136	0.054	-0.056	0.008	0.127

### <자료 9> MA(3) 모형 적합 후 잔차분석 결과

잔차의 QQ Plot이 대체적으로 직선 형태를 띠는 것으로 보아, 잔차들이 정규성을 만족한다는 사실을 확인할 수 있다. 또한 포트맨토검정 결과를 확인하면 잔차들이 백색 잡음과정이라는 귀무가설을 기각할 수 없어 MA(3) 모형으로의 적합이 적절하다고 판단된다.



### <자료 10> MA(3) 잔차들의 자기상관계수

또한 MA(3) 모형 적합 시 잔차들의 자기상관계수를 살펴보면 RSACF와 RSPACF가 거의 0의 값을 가지고 있다.

## 2.2 과대적합 확인

이제 과대적합을 통하여 모형진단을 해보자. MA(3) 모형보다 모수의 개수가 하나 더 많은 모형들인 MA(4)와 ARMA(1,3) 모형을 적합시켜 보기로 하자.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
MU	358.52686	3.41807	104.89	<.0001	0
MA1,1	0.76489	0.10260	7.46	<.0001	1
MA1,2	0.20525	0.12290	1.67	0.0982	2
MA1,3	-0.39945	0.12269	-3.26	0.0016	3
MA1,4	0.03784	0.10402	0.36	0.7169	4

Constant Estimate	358.5269
Variance Estimate	7614.679
Std Error Estimate	87.26213
AIC	1182.442
SBC	1195.468
Number of Residuals	100

<자료 11> MA(4) 적합 결과

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
MU	358.53532	3.31156	108.27	<.0001	0
MA1,1	0.56991	0.25616	2.22	0.0285	1
MA1,2	0.34390	0.19975	1.72	0.0884	2
MA1,3	-0.36925	0.09584	-3.85	0.0002	3
AR1,1	-0.20193	0.27483	-0.73	0.4643	1

Constant Estimate	430.9328
Variance Estimate	7602.471
Std Error Estimate	87.19215
AIC	1182.281
SBC	1195.307
Number of Residuals	100

<자료 12> ARMA(1,3) 적합 결과

MA(4) 모형의 적합 결과 추가된 모수  $\theta_4$ 의 값이 유의하지 않다고 도출되었으며, 추정된 분산도  $\hat{\sigma}_\varepsilon = 7614.679$ 로 MA(3) 모형보다 높다. ARMA(1,3) 모형도 역시 추가된 모수  $\phi_1$ 의 값이 유의하지 않다고 도출되었으며, 추정된 분산도  $\hat{\sigma}_\varepsilon = 7602.471$ 로 MA(3) 모형보다 높다. 또한, MA(3) 모형의 AIC, SBC값은 각각 1180.6, 1191.02으로, MA(4) 모형과 ARMA(1,3) 모형의 값들보다 작으므로 최종적인 예측모형으로 MA(3) 모형을 선택했다.

### 3. 결론

결론적으로, 주어진 시계열 자료의 최종적인 예측모형은 MA(3) 모형이며, 조건부 최소제곱법에 의해 적합된 모형은 다음과 같다.

$$Z_t - 358.52 = \varepsilon_t - 0.7569\varepsilon_{t-1} - 0.2102\varepsilon_{t-2} + 0.3769\varepsilon_{t-3}$$

$$\hat{\sigma}_\varepsilon^2 = 7547.264$$

해당 모형으로 적합했을 시, 잔차가 정규성을 만족함과 동시에 백색잡음과정이기 때문에 모형의 적합이 적절하다고 판단할 수 있다. 또한 모형의 과대적합을 확인해보아도, 모수가 추가된 모형들보다 잠정모형의 선택이 더욱 적절함을 확인할 수 있었다.

### 4. 참고문헌

- ◆ 「SAS/ETS와 R을 이용한 시계열분석 5판(조신섭, 2019)」, 율곡출판사.