시계열자료분석팀

5팀 김태훈 이소율 강희균 정희주 마채영

INDEX

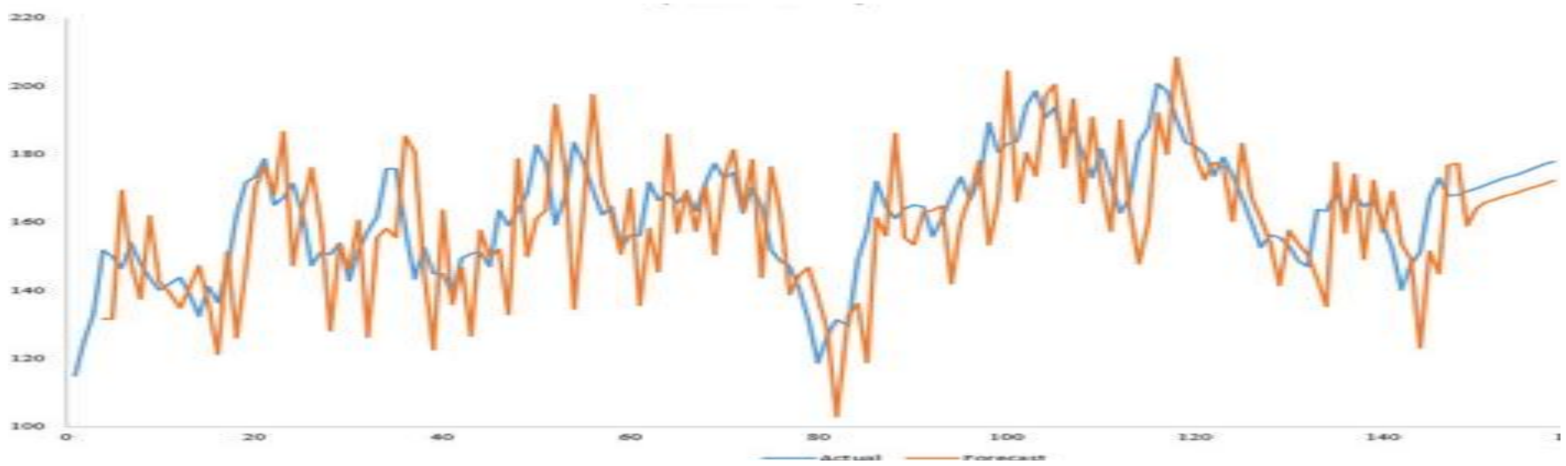
1

Review

Review

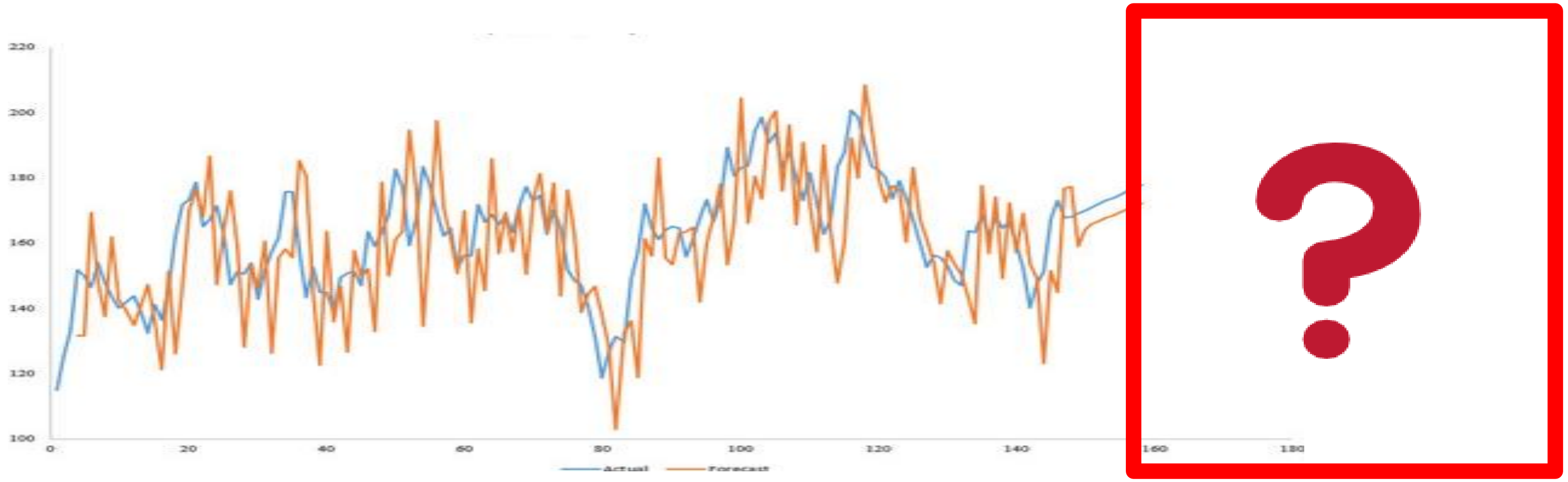
시계열 자료

시간의 흐름에 따라
관측되어 나타난 값



시계열 자료 분석

시계열 자료를 이용하여
미래를 예측하는 것



Review

시계열 자료

시간의 흐름에 따라
관측되어 나타난 값

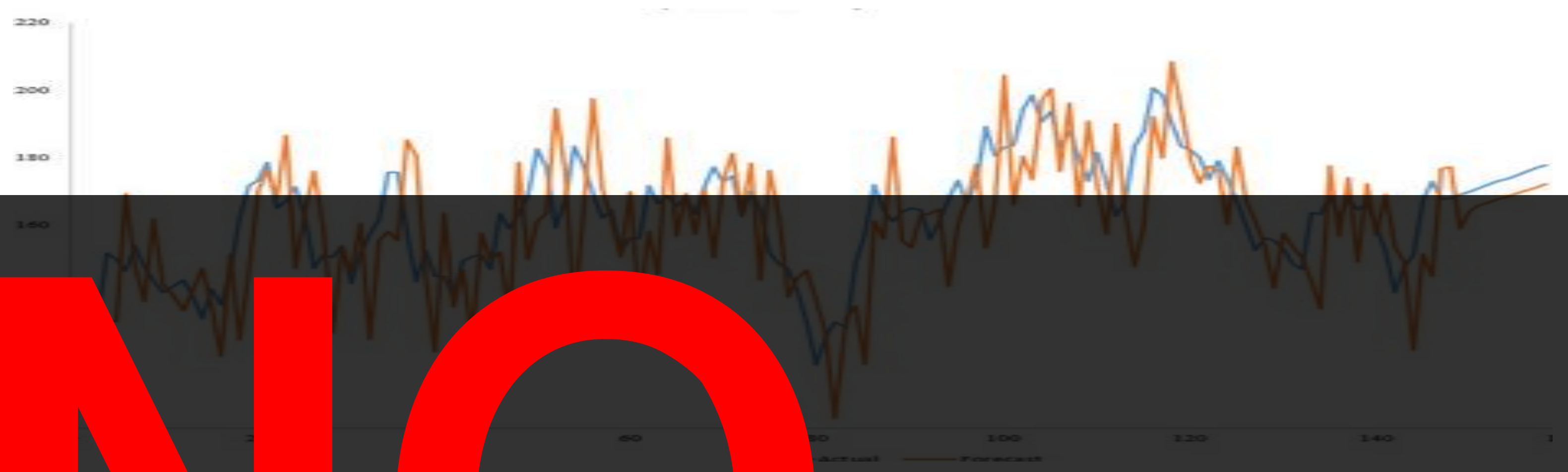
그런데, 아무런 조건 없이
미래를 예측하는 것이 가능한가요?

시계열 자료를 이용하여
미래를 예측하는 것

Review

시계열 자료

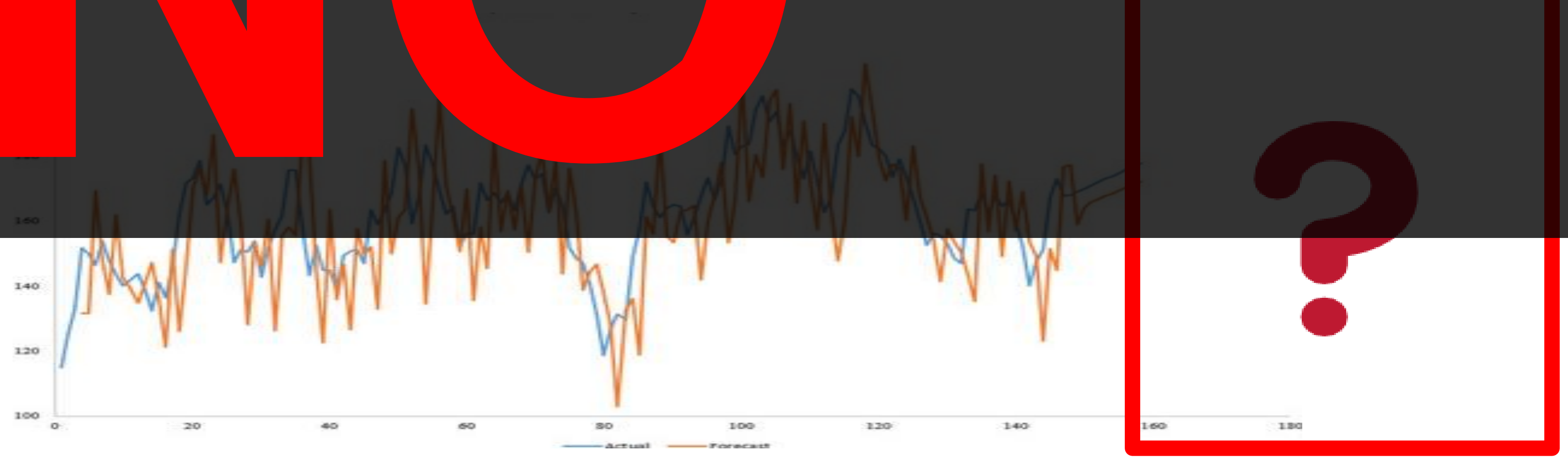
시간의 흐름에 따라
관측되어 나타난 값



NO

시계열 자료 분석

시계열 자료를 이용하여
미래를 예측하는 것



Review

STATIONARITY

시계열의 확률적인 성질들이 시간의 흐름에 따라 **변하지 않는다!**

분해(Decomposition)

STATIONARITY를 가정한 후 분해

$$X_t = \overbrace{m_t + s_t}^{Trend} + Y_t$$

Non Stationary part

Stationary Residuals

Review

TREND ESTIMATION

Non Stationary Part 를 제거해주면
Stationary Residuals 은 모델링이 가능!

$$X_t = \overbrace{m_t + s_t}^{Trend} + Y_t$$

Non Stationary part

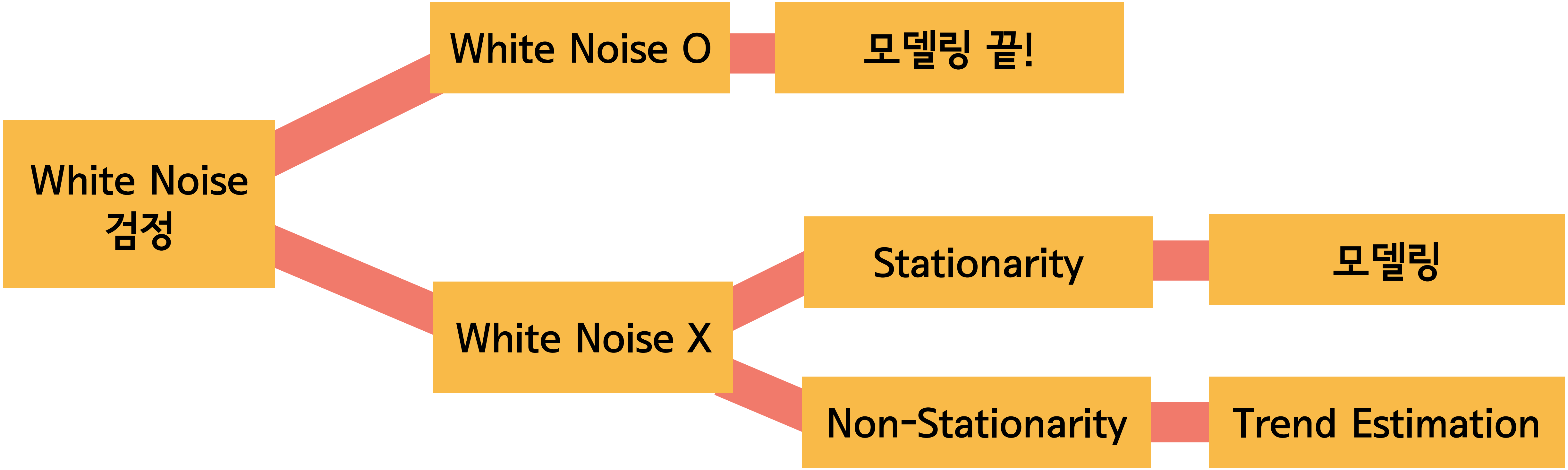
Stationary Residuals

추세 제거 방법

- OLS
- Moving Average Filter
- Exponential Smoothing
- Smoothing Splines
- Kernel Smoothing
- Differencing

Review

Stationary Residuals Y_t 에 대하여

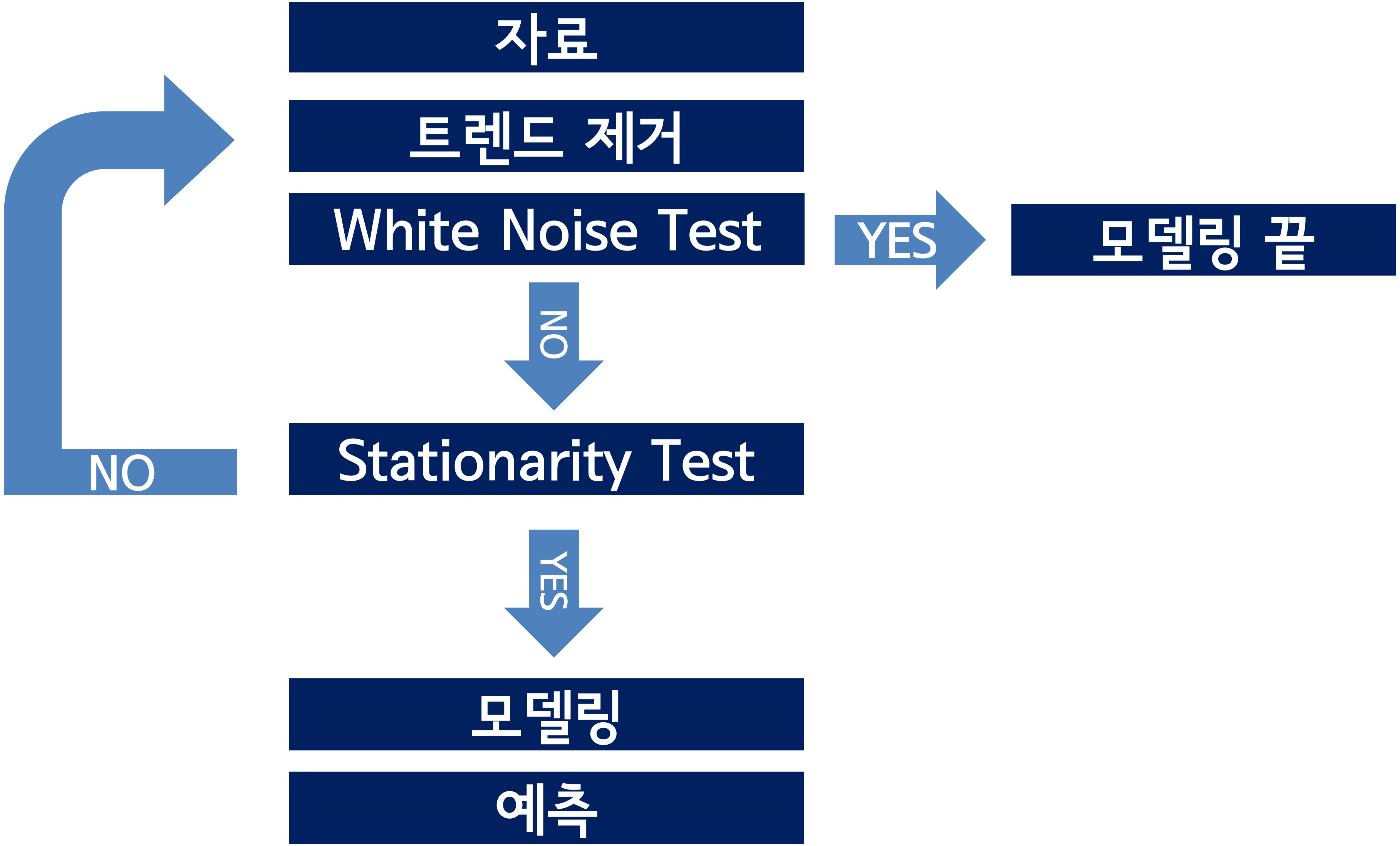


Review

Example!

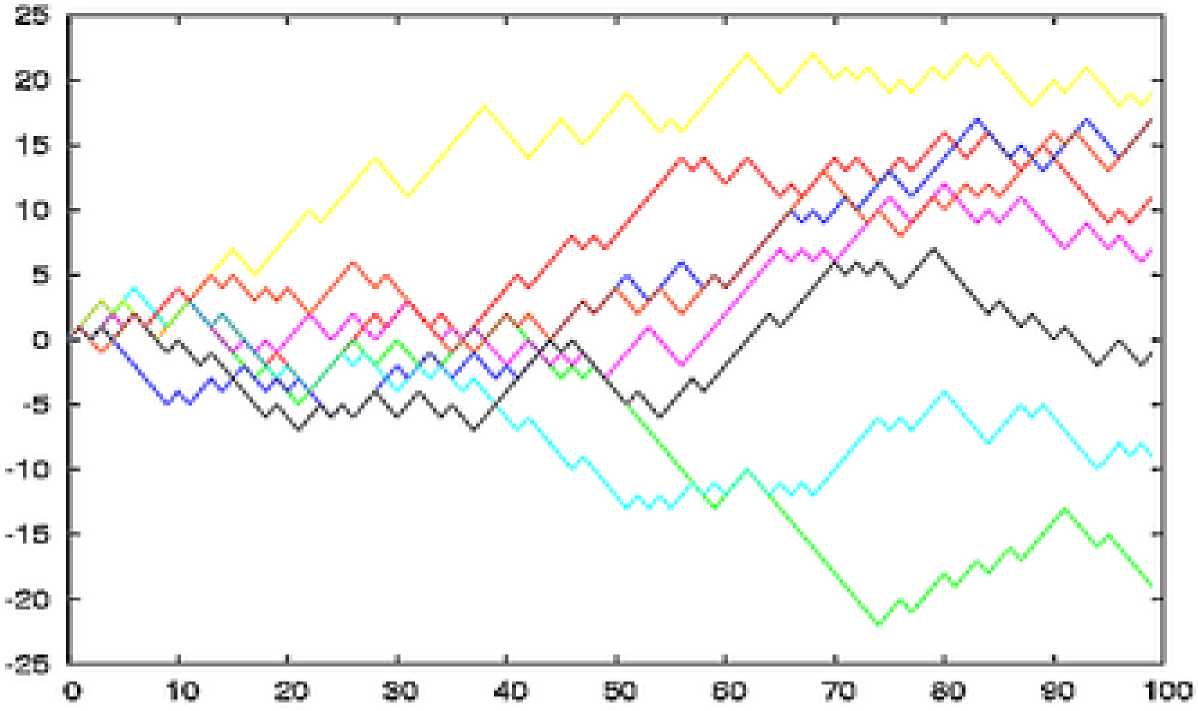
데이터 분석 단계별로 다시 복습해보자!

Review


자료
트렌드 제거
White Noise Test
YES
모델링 끝
NO
Stationarity Test
NO
YES
모델링
예측

Review

자료 1 예시 : Random Walk



Non-stationary process

$$X_t = a_t + X_{t-1}$$

차분하면 a_t **만 남게 된다!**

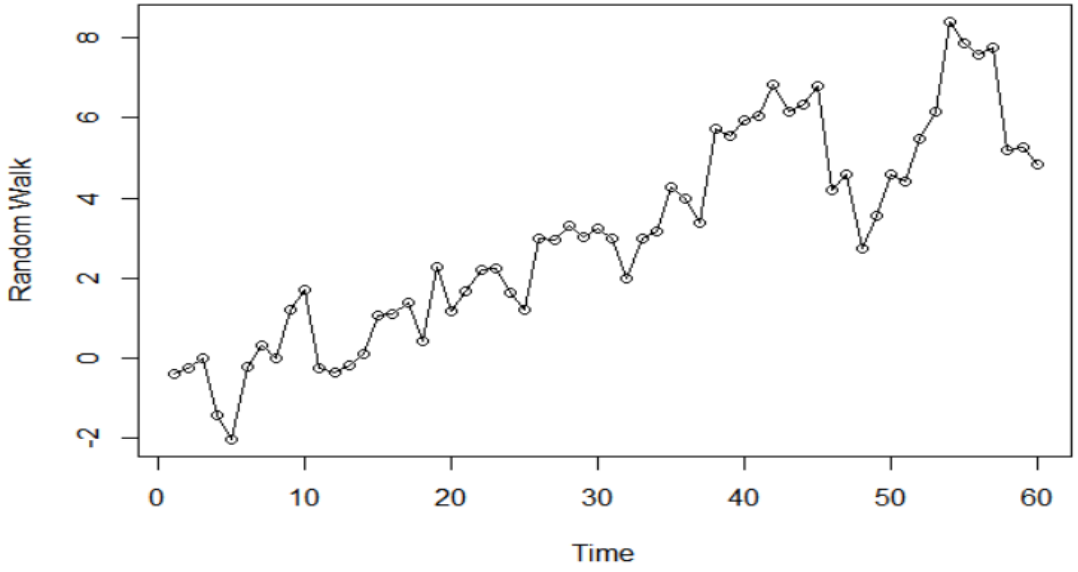
$$X_t - X_{t-1} = a_t$$

$$a_t \sim WN$$

Review

자료 1 예시 : Random Walk

```
## random walk process
data(rwalk)
random_walk <- rwalk
plot( rwalk, type ='o', ylab = 'Random Walk' )
decomprwalk <- diff( rwalk ) #make stationary by difference
plot.ts( decomprwalk )
acf( decomprwalk ) #acf plot of differenced randomwalk
test( decomprwalk ) # iid test for differenced randomwalk
```

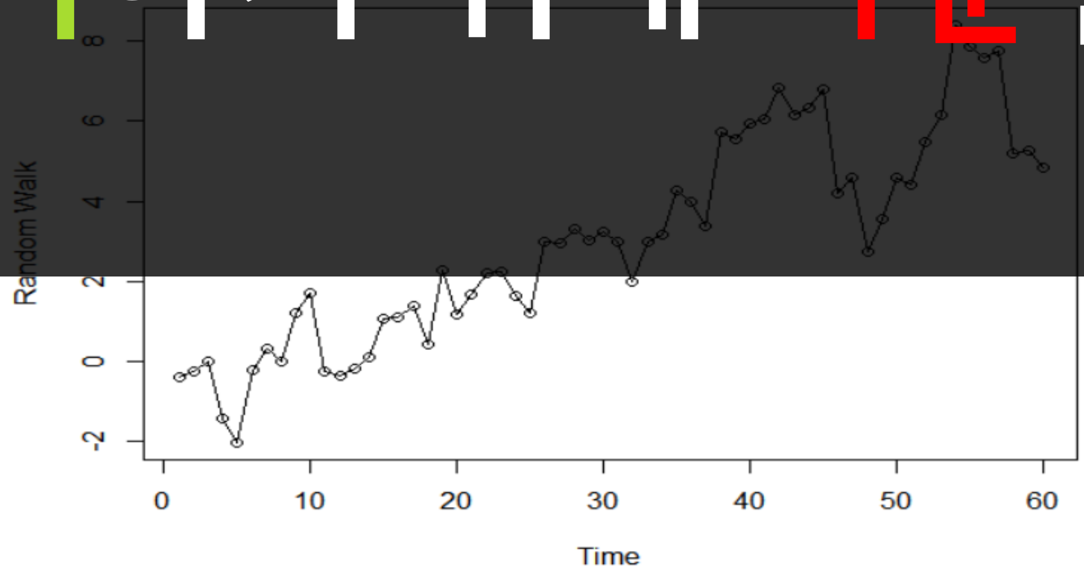


Review

자료 1 예시 : Random Walk

```
## random walk process
data(rwalk)
random_walk <- rwalk
plot( rwalk, type ='o', ylab = 'Random Walk' )
decomprwalk <- diff( rwalk ) #make stationary by difference
plot.ts( decomprwalk )
acf( decomprwalk ) #acf plot of differenced randomwalk
```

추세를 제거하기 위해 차분을 해보자!



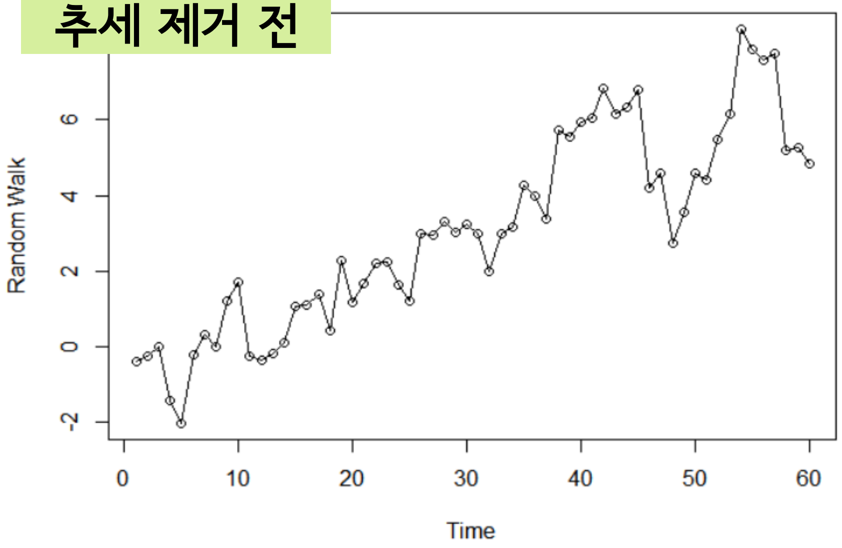
Review

분해 & 트렌드 제거 2 3 예시 : Random Walk

R

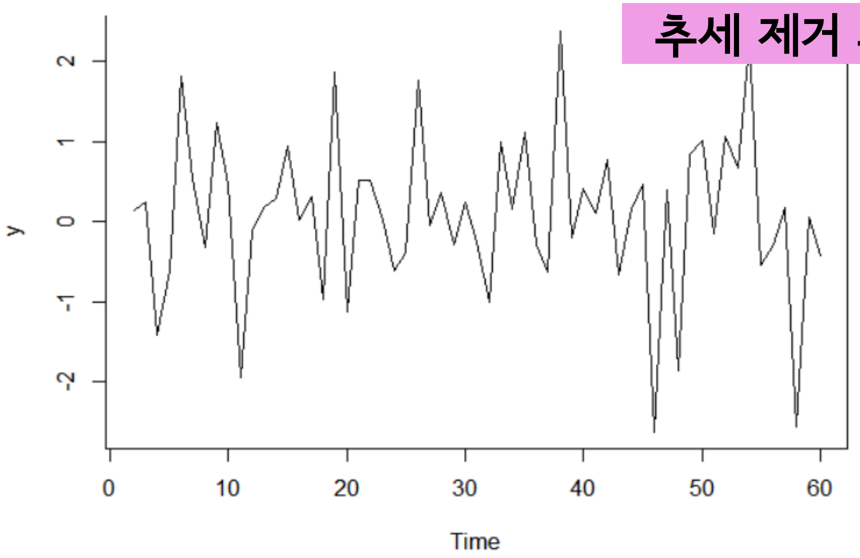
```
## random walk process
data(rwalk)
random_walk <- rwalk
plot( rwalk, type ='o', ylab = 'Random Walk' )
decomprwalk <- diff( rwalk ) #make stationary by difference
plot.ts( decomprwalk )
acf( decomprwalk ) #acf plot of differenced randomwalk
test( decomprwalk ) # iid test for differenced randomwalk
```

추세 제거 전



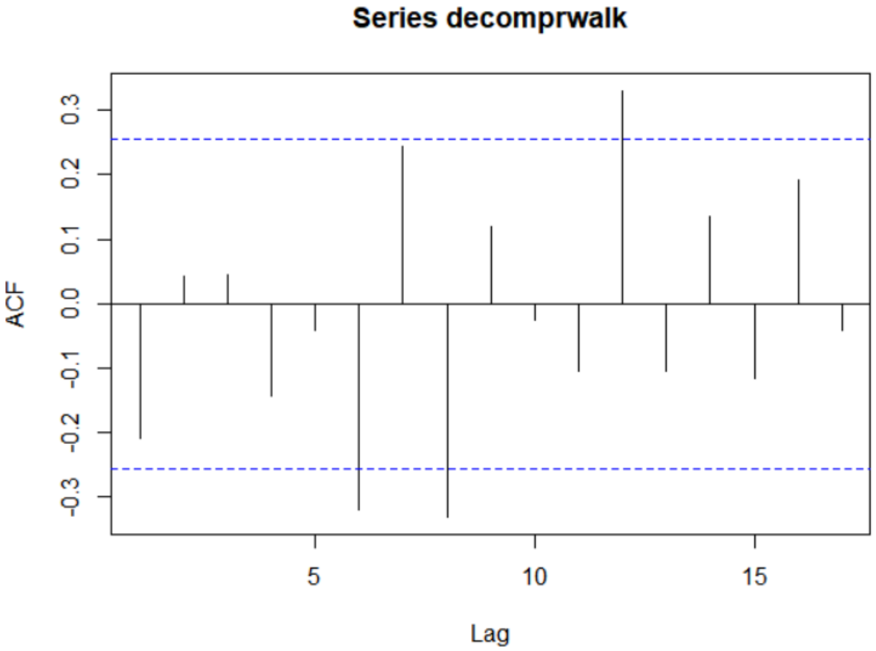
차분

추세 제거 후



Review

```
## random walk process
data(rwalk)
random_walk <- rwalk
plot( rwalk, type ='o', ylab = 'Random Walk' )
decomprwalk <- diff( rwalk ) #make stationary by difference
plot.ts( decomprwalk )
acf( decomprwalk ) #acf plot of differenced randomwalk
test( decomprwalk ) # iid test for differenced randomwalk
```



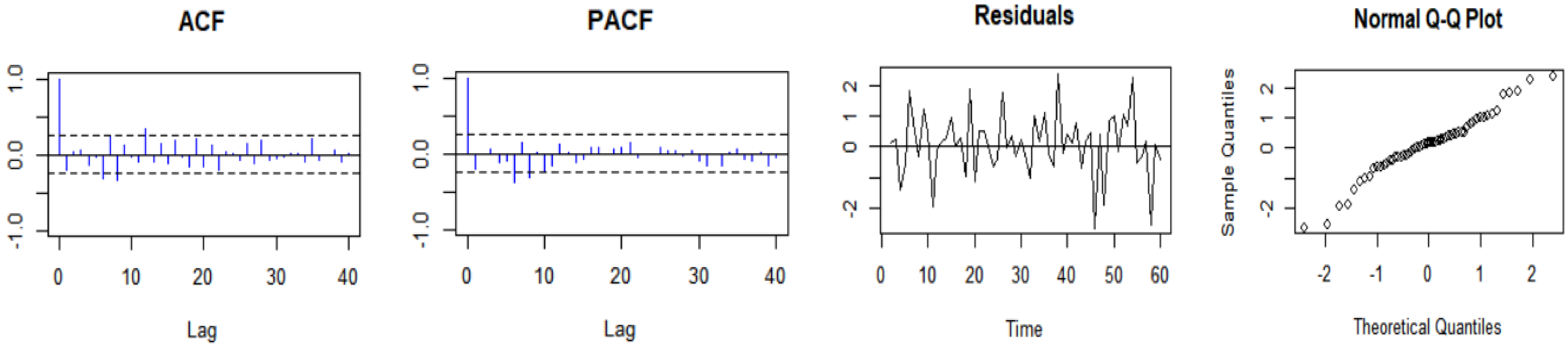
추세를 제거한 후 남은 White Noise에 대하여 ACF를 구했더니 0 주변에 그 값이 존재!

$$X_t \sim WN(0, \sigma_x{}^2)\ \gamma_x(r,s) = 0$$

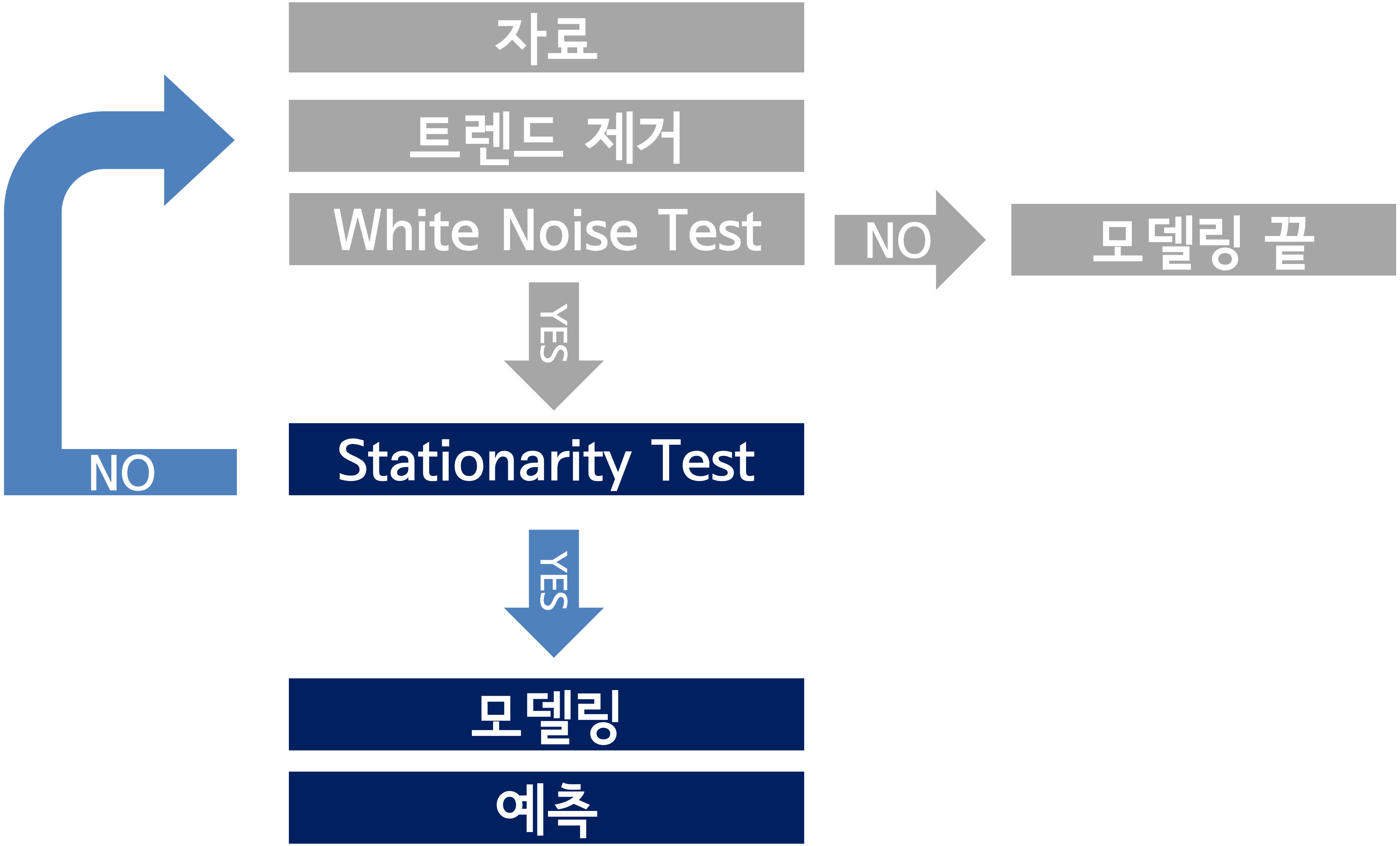
Review

White Noise 검정 4 예시 : Random Walk

```
> test( decomprwalk ) # iid test for differenced randomwalk
Null hypothesis: Residuals are iid noise.
Test                          Distribution Statistic   p-value
Ljung-Box Q                   Q ~ chisq(20)    49.07     3e-04 *
McLeod-Li Q                   Q ~ chisq(20)    28.49    0.0983
Turning points T       (T-38)/3.2 ~ N(0,1)        43    0.1169
Diff signs S           (S-29)/2.2 ~ N(0,1)        31    0.3711
Rank P             (P-855.5)/76.5 ~ N(0,1)       845    0.8908
```



Review


자료
트렌드 제거
White Noise Test
NO
모델링 끝
YES
NO
Stationarity Test
YES
모델링
예측

Review

모델링 방법?

Trend를 제거한 후
ARMA 모델을 사용

VS

ARIMA 모델을 사용

그렇지 않을 경우

추세가 선형적이라면
ARIMA 모델을 사용하여
차분할 경우 편리함

2

Properties

Properties

AR모델 이란?

자기회귀모형(Auto Regression Model, AR Model)

시계열의 현재 관측값이 과거 p기간 동안 관측값에 직접적으로 의존한다!

오차항(노이즈), WN를 따른다.

AR(1) $$X_t = \phi X_{t-1} + a_t$$

AR모델의 계수 Phi

AR(p) $$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t$$

Properties

MA모델 이란?

이동평균모형(Moving Average Model, MA Model)

시계열의 현재 관측값이 현재와 과거 q기간 동안 오차항에만 의존한다!

MA모델의 계수 Theta

MA(1) $$X_t = a_t + \theta a_{t-1}$$

오차항(노이즈), WN를 따른다.

MA(q) $$X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q}$$

Properties

ARMA모델 이란?

자기회귀이동평균모형(Auto Regressive Moving Average, ARMA Model)

→ AR모형과 MA모형의 선형 결합

ARMA(1,1) $\boxed{X_t - \phi X_{t-1} = a_t + \theta a_{t-1}}$

ARMA(p,q) $\boxed{X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q}}$

ARIMA모델 이란?

먼저 차분을 하고 ARMA 적용하는 모델

d : 차분하는 횟수

ARIMA(p,d,q)

$$\phi(p)\Delta^{d}X_t = \theta(q)a_t$$

Properties

AR, MA, ARMA는 각각 다음의 조건들을 만족해야 한다!

Stationarity (정상성)	시계열 자료분석의 기본 전제
Causality (인과성)	현재 데이터가 과거 데이터에만 의존한다
Invertibility (가역성)	현재 에러가 과거 데이터에만 의존한다

Properties

AR, MA, ARMA는 각각 다음의 조건들을 만족해야 한다!

Stationarity (정상성) — 시계열 자료분석의 기본 전제

앞으로 다루는 AR, MA, ARMA가 3가지 properties를 만족한다고 가정!

Causality (인과성) — 현재 데이터가 과거 데이터에만 의존한다

Invertibility (가역성) — 현재 에러가 과거 데이터에만 의존한다

Properties

Stationarity of AR(1)

$$\begin{aligned} X_t &= \phi X_{t-1} + a_t \\ &= \phi^2 X_{t-2} + \phi a_{t-1} + a_t \\ &= \phi^3 X_{t-3} + \phi^2 a_{t-2} + \phi a_{t-1} + a_t \\ &\vdots \\ &= \boxed{\phi^n X_{t-n} + \phi^{n-1} a_{t-n+1} + \phi^{n-2} a_{t-n+2} + \cdots + \phi a_{t-1} + a_t} \end{aligned}$$

$X_{t-1} = \phi X_{t-2} + a_{t-1}$ 대입

$X_{t-2} = \phi X_{t-3} + a_{t-2}$ 대입

AR(1)

$$X_t = \phi X_{t-1} + a_t$$
$$X_{t-1} = \phi X_{t-2} + a_{t-1}$$
$$X_{t-2} = \phi X_{t-3} + a_{t-2}$$

Properties

Stationarity of AR(1)

- $|\phi| < 1$ 일 때만 stationary!

$|\phi| < 1$ 인 경우 무한등비급수에 의해
X_{t-n} 이 0으로 수렴!

$$X_t = \phi^n X_{t-n} + \phi^{n-1} a_{t-n+1} + \phi^{n-2} a_{t-n+2} + \cdots + \phi a_{t-1} + a_t$$

$$X_t = \phi^{n-1} a_{t-n+1} + \phi^{n-2} a_{t-n+2} + \cdots + \phi a_{t-1} + a_t$$

a_t끼리의 관계식만 남는다!
다시 말해, WN을 따르는 에러(a_t)들 간 선형결합이므로 stationarity

- $|\phi| > 1$ 의 경우는 현재 값이 미래 값에 의존하게 된다.
  = 시계열에서 불가능함

Properties

Stationarity of AR(1)

- $|\phi| = 1$ 의 경우 앞에서 예로 들었던 Random Walk

Example. $\phi = 1$ 일 때

$$
\begin{aligned}
X_t &= a_t + X_{t-1} \\
&= a_t + a_{t-1} + X_{t-2} \\
&= a_t + a_{t-1} + a_{t-2} + X_{t-3} \\
&\vdots \\
&= a_t + a_{t-1} + \cdots + a_{t-n} + X_{t-n-1}
\end{aligned}
$$

X_t는 노이즈들의 합

$$
\begin{aligned}
V(X_t) &= V(a_t + a_{t-1} + \cdots) \\
&= V(a_t) + V(a_{t-1}) + \cdots
\end{aligned}
$$

$V(X_t) = t\sigma^2$ 으로
t값에 따라 증가

Properties

Stationarity of AR(1)

- $|\phi| = 1$ 의 경우 앞에서 예로 들었던 Random Walk

Example. $\phi = 1$ 일 때

$$
\begin{aligned}
X_t &= a_t + X_{t-1} \\
&= a_t + a_{t-1} + X_{t-2} \\
&= a_t + a_{t-1} + a_{t-2} + X_{t-3} \\
&\vdots \\
&= a_t + a_{t-1} + \cdots + a_{t-n} + X_{t-n-1}
\end{aligned}
$$

X_t는 노이즈들의 합

Non stationarity!

$$
\begin{aligned}
V(X_t) &= V(a_t + a_{t-1} + \cdots) \\
&= V(a_t) + V(a_{t-1}) + \cdots
\end{aligned}
$$

$\mathrm{V}(X_t) = t\sigma^2$ 으로 t값에 따라 증가

Properties

Stationarity of AR(1)

- $|\phi| = 1$ 의 경우 앞에서 예로 들었던 Random Walk

Example. $\phi = 1$ 일 때

$$
\begin{aligned}
X_t &= a_t + X_{t-1} \\
&= a_t + a_{t-1} + X_{t-2} \\
&= a_t + a_{t-1} + a_{t-2} + X_{t-3} \\
&\vdots \\
&= a_t + a_{t-1} + \cdots + a_{t-n} + X_{t-n-1}
\end{aligned}
$$

X_t는 노이즈들의 합

AR의 경우

$|\phi|<1$ 일 경우만 stationary!

Non stationarity!

$$
\begin{aligned}
V(X_t) &= V(a_t + a_{t-1} + \cdots) \\
&= V(a_t) + V(a_{t-1}) + \cdots
\end{aligned}
$$

$V(X_t) = t\sigma^2$ 으로 t값에 따라 증가

Properties

Stationarity of MA(1)

$$X_t = a_t + \theta a_{t-1}$$

MA(1)

WN을 따르는 에러(a_t)들 사이의
선형 결합이므로 stationary!

Properties

Stationarity of ARMA(p,q)

$$\boxed{X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q}}$$

Backshift Operator 'B'를 사용하여 식을 다시 전개한다

$$\boxed{\begin{aligned} BX_t &= X_{t-1} \\ B^2 X_t &= X_{t-2} \\ &\vdots \\ B^k X_t &= X_{t-k} \end{aligned}}$$

$$\underbrace{(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)}_{\phi(B)\text{라는 함수로 정의}} X_t = \underbrace{(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q)}_{\theta(B)\text{라는 함수로 정의}} a_t$$

$$\boxed{\phi(B) X_t = \theta(B) a_t}$$

Properties

Stationarity of ARMA(p,q)

$$\phi(B)X_t = \theta(B)a_t$$

$$\phi(Z)X_t = \theta(Z)a_t$$

B=Z로 치환

$$X_t = \frac{\theta(Z)}{\phi(Z)}a_t = \frac{1}{\phi(Z)}\theta(Z)a_t = \frac{1}{\phi(Z)}(a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q})$$

ARMA(p,q) has unique stationary solution if and only if

$$\boxed{\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \neq 0 \text{ for all } |z| = 1}$$

Properties

Causality of AR(1)

$$X_t = \phi^n X_{t-n} + \phi^{n-1} a_{t-n+1} + \phi^{n-2} a_{t-n+2} + \cdots + \phi a_{t-1} + a_t$$

마찬가지로 $|\phi| < 1$일 때만 causality!

$$\boxed{X_t = \phi^{n-1} a_{t-n+1} + \phi^{n-2} a_{t-n+2} + \cdots + \phi a_{t-1} + a_t}$$

현재 데이터(X_t)가 과거의 에러(a_t)들에 의존하므로 causality 만족

Properties

Causality of MA(1)

$$X_t = a_t + \theta a_{t-1}$$

MA(1)

현재 데이터(X_t)가 과거의 에러(a_t)들에 의존하므로 causality 만족

Properties

Causality of ARMA(p,q)

$$X_t = \frac{\theta(Z)}{\phi(Z)} a_t = \frac{1}{\phi(Z)} \theta(Z) a_t = \frac{1}{\phi(Z)} (a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q})$$

현재 데이터(X_t)가 과거의 에러(a_t)들에 의존하여 causality 만족하려면

ARMA(p, q) is causal if

$$\boxed{\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \neq 0 \text{ for all } |z| \leq 1}$$

Invertibility of AR(1)

$$X_t = \Phi X_{t-1} + a_t$$

$$a_t = X_t - \Phi X_{t-1}$$

현재 에러(a_t)가 과거의 데이터(X_t)들에
의존하므로 invertibility 만족

Invertibility of MA(1)

$$X_t = a_t + \theta a_{t-1}$$

$$\begin{aligned} a_t &= X_t - \theta a_{t-1} \\ &= X_t - \theta X_{t-1} + \theta^2 a_{t-2} \\ &= X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \theta^3 a_{t-3} \\ &\;\vdots \\ &= \boxed{X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \cdots - \theta^{n-1} X_{t-n+1} + \theta^n a_{t-n}} \end{aligned}$$

$a_{t-1} = X_{t-1} - \theta a_{t-2}$ 대입

$a_{t-2} = X_{t-2} - \theta a_{t-3}$ 대입

MA(1)

$$a_t = X_t - \theta a_{t-1}$$
$$a_{t-1} = X_{t-1} - \theta a_{t-2}$$
$$a_{t-2} = X_{t-2} - \theta a_{t-3}$$

Invertibility of MA(1)

- $|\theta| < 1$일 때만 invertible!

$|\theta| < 1$ 인 경우 무한등비급수에 의해 a_{t-n}이 0으로 수렴!

$$a_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \cdots - \theta^{n-1} X_{t-n+1} + \theta^n a_{t-n}$$

현재 에러(a_t)가 과거의 데이터(X_t)들에 의존하므로 invertibility 만족

Properties

Invertibility of ARMA(p,q)

$$X_t = \frac{\theta(Z)}{\phi(Z)} a_t \,,\; a_t = \frac{\phi(Z)}{\theta(Z)} X_t$$

$$\boxed{a_t = \frac{\phi(Z)}{\theta(Z)} X_t = \frac{1}{\theta(Z)} \phi(Z) X_t = \frac{1}{\theta(Z)} (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p})}$$

현재 에러(a_t)가 과거의 데이터(X_t)들에 의존하여 invertibility 만족하려면

ARMA(p,q) is invertible, that is,

$$Z_t = \theta(B)^{-1}\phi(B) X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad \sum_{j=0}^{\infty} |\pi_j| < \infty$$

if

$$\boxed{\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q \neq 0 \text{ for all } |z| \leq 1}$$

Properties

Summary

	Stationarity	Causality	Invertibility
AR(1)	$\lvert\phi\rvert < 1$	$\lvert\phi\rvert<1$	
MA(1)			$\lvert\theta\rvert<1$
ARMA(p,q)	$\phi(p) \neq 0$ for all $\lvert p\rvert = 1$	$\phi(p) \neq 0$ for all $\lvert p\rvert \leq 1$	$\theta(q) \neq 0$ for all $\lvert q\rvert \leq 1$

3

ACF/PACF

ACF / PACF

ACF (by Yule-Walker Equations)

$$AR(1),\ X_t = \varphi X_{t-1} + a_t$$

양 변에 X_{t-h} 을 곱하자

$$X_t X_{t-h} = \varphi X_{t-1} X_{t-h} + a_t X_{t-h}$$

$$\gamma_x(h) = E(X_t X_{t-h}) = E(\varphi X_{t-1} X_{t-h}) + E(a_t X_{t-h})$$

$$\gamma_x(h) = \varphi \gamma_x(h-1) \quad (\text{h>1})$$

$\gamma_x(h) = \varphi^n$, **지수적 감소**

ACF / PACF

ACF (by Yule-Walker Equations)

$$MA(1), X_t = a_t + \theta a_{t-1}$$

$$X_t X_{t-h} = a_t X_{t-h} + \theta a_{t-1} X_{t-h}$$

양 변에 X_{t-h} 을 곱하자

$$\gamma_x(h) = E(X_t X_{t-h}) = E(a_t X_{t-h}) + E(\theta a_{t-1} X_{t-h}) = 0$$

ACF of MA(q) ; γ_x $(h)=0$

q 시점 이후 모든 값이 0

ACF / PACF

PACF (by Yule-Walker Equations)

$$AR(1),\ X_t = \varphi X_{t-1} + a_t$$

$$\varphi_{22} = \mathrm{corr}(\ X_t - \varphi X_{t-1},\ X_{t-1} - \varphi X_{t-2}) = \mathrm{corr}(\ a_t\ , a_{t-1}\ ) = 0$$

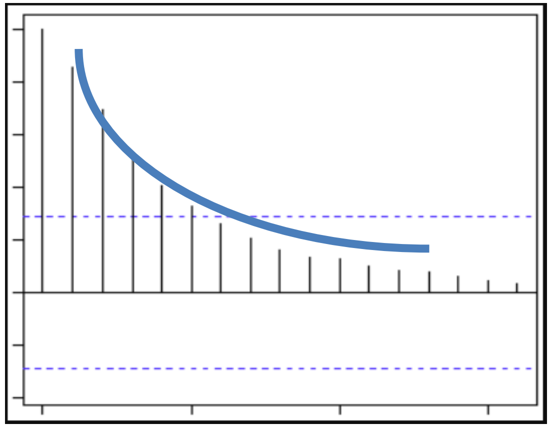
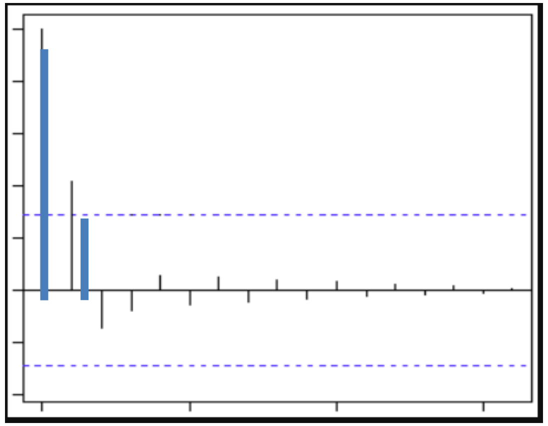
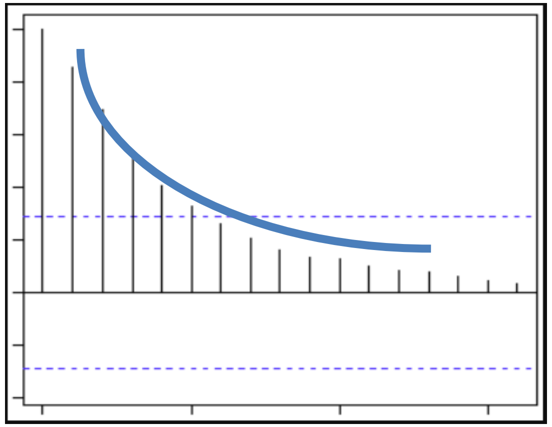
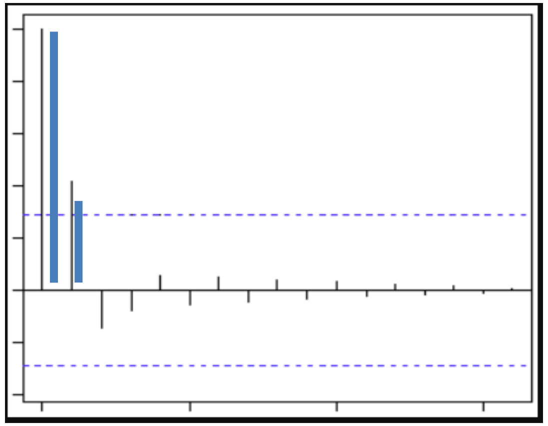
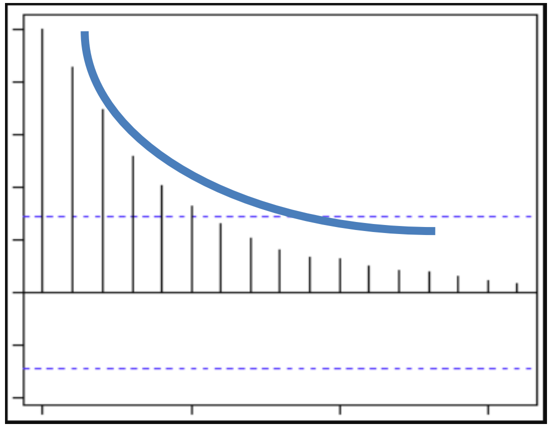
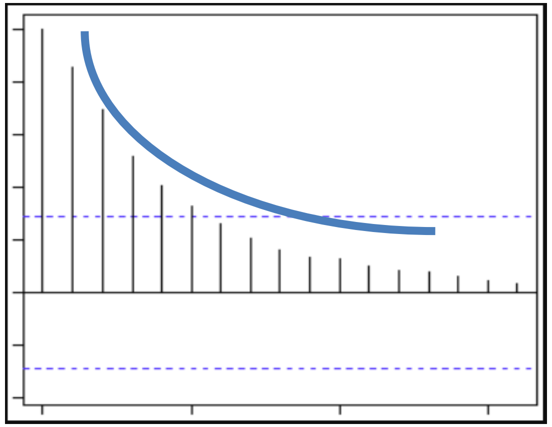
PACF of AR(q) ; φ_hh=0

p 시점 이후 모든 값이 0

PACF of MA(q) ; φ_hh= θ^h

지수적 감소

ACF / PACF

	AR	MA	ARMA
ACF			
PACF			

Prediction

Example!

실제 시계열 자료를 분석해보자

Prediction

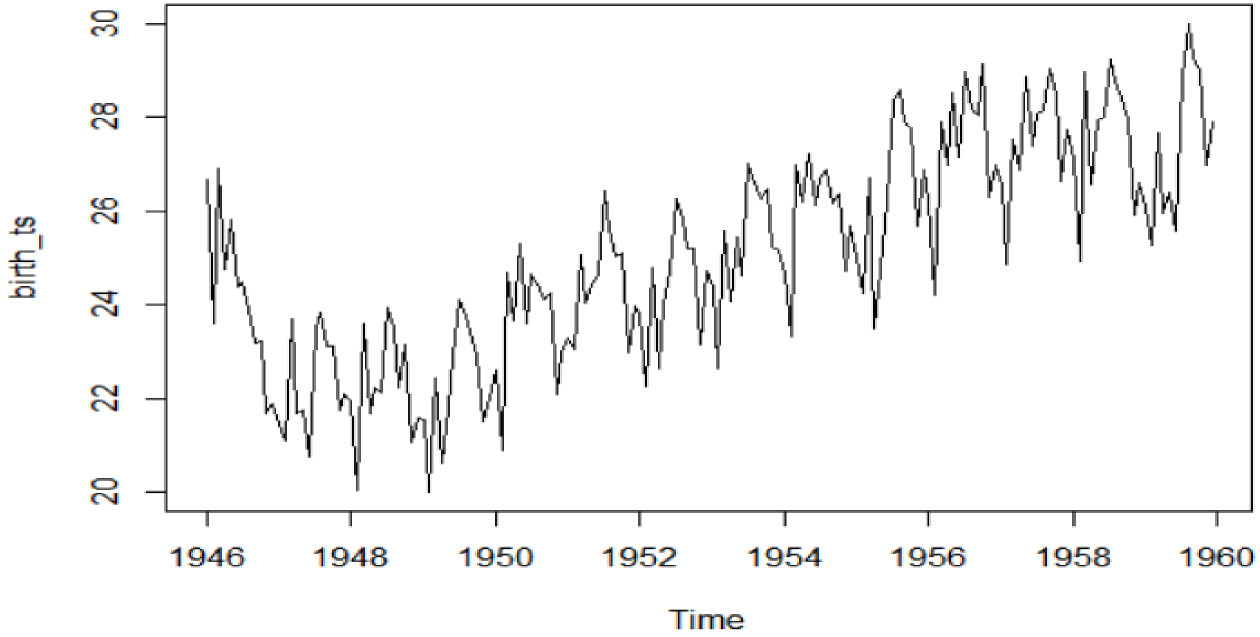
예시 : Birth Data

```
births <- scan( "http://robjhyndman.com/tsdldata/data/nybirths.dat" )
birth_ts <- ts( births, frequency = 12, start = c( 1946, 1 ) )
ts.plot( birth_ts )

test( birth_ts ) ## it is not iid...

arima_birth <- auto.arima( birth_ts, ic = c( 'aicc' ) )
arima_birth ## Sarima(2,1,2) * (1,1,1) is the best model
            ## you don't have to understand sarima process at now,
            ## It is same as arima process considering seasonality.

tsdisplay( residuals( arima_birth ) ) ## residual plot. there is no correlation between residuals
test( residuals( arima_birth ) ) ## test completed. Q) why do residuals should satisfy iid?

birth_forecast <- forecast( arima_birth, h = 30 )
plot( birth_forecast ) #you can see predict interval getting large as time goes on. (property of ARIMA)
                       # h means predict time. do it again with h = 100 !
```

Prediction

예시 : Birth Data

```r
births <- scan( "http://robjhyndman.com/tsdldata/data/nybirths.dat" )
birth_ts <- ts( births, frequency = 12, start = c( 1946, 1 ) )
ts.plot( birth_ts )

test( birth_ts ) ## it is not iid...
```



Prediction

예시 : Birth Data

```r
births <- scan( "http://robjhyndman.com/tsdldata/data/nybirths.dat" )
birth_ts <- ts( births, frequency = 12, start = c( 1946, 1 ) )
ts.plot( birth_ts )

test( birth_ts ) ## it is not iid...
```

```
> test( birth_ts ) ## it is not iid...
Null hypothesis: Residuals are iid noise.
Test                          Distribution Statistic   p-value
Ljung-Box Q                  Q ~ chisq(20)   1366.31         0 *
McLeod-Li Q                  Q ~ chisq(20)   1358.38         0 *
Turning points T  (T-110.7)/5.4 ~ N(0,1)       115    0.4253
Diff signs S       (S-83.5)/3.8 ~ N(0,1)        76    0.0457 *
Rank P          (P-7014)/364.5 ~ N(0,1)      11097         0 *
```

Prediction

예시 : Birth Data

```
births <- scan( "http://robjhyndman.com/tsdldata/data/nybirths.dat" )
birth_ts <- ts( births, frequency = 12, start = c( 1946, 1 ) )
ts.plot( birth_ts )

test( birth_ts ) ## it is not iid...
```

auto.arima를 이용하여 모델을 구해보자!

```
> test( birth_ts ) # it is not iid...
Null hypothesis: Residuals are iid noise.
Test                          Distribution Statistic   p-value
Ljung-Box Q                   Q ~ chisq(20)   1366.31        0 *
McLeod-Li Q                   Q ~ chisq(20)   1358.38        0 *
Turning points T  (T-110.7)/5.4 ~ N(0,1)          115   0.4253
Diff signs S        (S-83.5)/3.8 ~ N(0,1)           76   0.0457 *
Rank P           (P-7014)/364.5 ~ N(0,1)        11097        0 *
```

Prediction

예시 : Birth Data

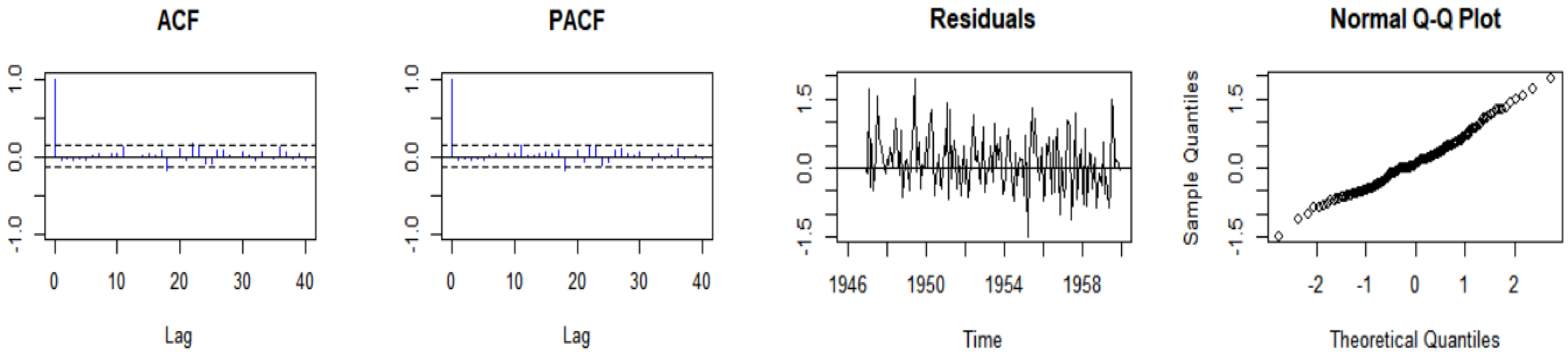
```
arima_birth <- auto.arima( birth_ts, ic = c( 'aicc' ) )
arima_birth ## Sarima(2,1,2) * (1,1,1) is the best model
            ## you don't have to understand sarima process at now,
            ## It is same as arima process considering seasonality.
```

```
> arima_birth ## Sarima(2,1,2) * (1,1,1) is the best mode
Series: birth_ts
ARIMA(2,1,2)(1,1,1)[12]

Coefficients:
         ar1      ar2      ma1     ma2     sar1     sma1
      0.6539  -0.4540  -0.7255  0.2532  -0.2427  -0.8451
s.e.  0.3004   0.2429   0.3228  0.2879   0.0985   0.0995

sigma^2 estimated as 0.4076:  log likelihood=-157.45
AIC=328.91   AICc=329.67   BIC=350.21
```

Prediction

예시 : Birth Data

```
tsdisplay( residuals( arima_birth ) ) ## residual plot. there is no correlation between residuals
test( residuals( arima_birth ) ) ## test completed. Q) why do residuals should satisfy iid?
```



4

Parameter Estimation

Parameter Estimation

계수 추정 방법

OLS

R ar.ols

단순회귀방법의 가정이
시계열 자료를 분석하는 데
필요한 가정과 **충돌**!

4 Parameter Estimation

계수 추정 방법

OLS

ar.ols

Conditional Least Square Estimation

단순회귀방법의 가정이
시계열 자료를 분석하는 데
필요한 가정과 **충돌**!

Parameter Estimation

CLSE : AR(1)

$$X_t = \phi X_{t-1} + a_t, a_t \sim WN$$

Let $a_0 = \mathbf{0}$

계수 ϕ의 값은?

Parameter Estimation

CLSE : AR(1)

$$X_t = \phi X_{t-1} + a_t,\ a_t \sim WN$$

Let $a_0 = 0$

a_t 제곱의 합을 최소로 하는

ϕ의 값이 목표

계수 ϕ의 값은?

Parameter Estimation

CLSE : AR(1)

$$X_t = \phi X_{t-1} + a_t, a_t \sim WN$$

Let $a_0 = \mathbf{0}$

$$a_t = X_t - \phi X_{t-1}$$
$$\sum a_t^2 = \sum (X_t - \phi X_{t-1})^2$$

$$\Rightarrow \quad \phi = (\sum X_t X_{t-1}) / (\sum X_t^2)$$

CLSE : MA(1)

$$X_t = a_t - \theta a_{t-1} \; , \; a_t \sim WN$$

Let $a_0 = 0$

계수 θ 의 값은?

CLSE : MA(1)

$$X_t = a_t - \theta a_{t-1}, \quad a_t \sim WN$$

Let $a_0 = 0$

$\tilde{a}_t$ 제곱의 합을 최소로 하는 θ의 값이 목표

계수 θ 의 값은?

Parameter Estimation

CLSE : MA(1)

$$X_t = a_t - \theta a_{t-1}, \quad a_t \sim WN$$

Let $a_0 = 0$

$$\tilde{a}_1 = X_1 + \theta a_0 = X_1$$

$$\tilde{a}_2 = X_2 + \theta \tilde{a}_1 = X_2 + \theta X_1$$

$$\sum_{t=1}^{n} \tilde{a}_t^2 \Rightarrow \underset{\theta}{\operatorname{argmin}} \sum_t \tilde{a}_t^2$$

Parameter Estimation

CLSE : ARMA(1,1)

$$X_t = \phi X_{t-1} + a_t - \theta a_{t-1}, \quad a_t \sim WN$$

Let $a_0 = 0$

계수 ϕ와 θ의 값은?

CLSE : ARMA(1,1)

$$X_t = \phi X_{t-1} + a_t - \theta a_{t-1}, \quad a_t \sim WN$$

$$\text{Let } a_0 = 0$$

$\tilde{a}_t$ 제곱의 합을 최소로 하는
ϕ와 θ의 값이 목표

계수 ϕ와 θ의 값은?

Parameter Estimation

CLSE : ARMA(1,1)

$$X_t = \phi X_{t-1} + a_t - \theta a_{t-1}, \; a_t \sim WN$$

Let $a_0 = 0$

$$\widetilde{a_1} = X_1 - \phi X_0 + \theta * a_0$$
$$\widetilde{a_2} = X_2 - \phi X_1 + \theta * \tilde{a}_1$$

$$\sum_{t=1}^{n} \widetilde{a_t^2} \Rightarrow \underset{\theta\,\phi}{\operatorname{argmin}} \sum_t \widetilde{a_t^2}$$

Parameter Estimation

CLSE : ARMA(1,1)

$$X_t = \phi X_{t-1} + a_t - \theta a_{t-1}, \quad a_t \sim WN$$

Let $a_0 = 0$

ARMA(1,1)을
ARMA(p,q)로 확장해본다면..?

$$\widetilde{a_1} = X_1 - \phi X_0 + \theta * a_0$$
$$\widetilde{a_2} = X_2 - \phi X_1 + \theta * \tilde{a}_1$$

$$\sum_{t=1}^{n} \widetilde{a_t^2} \Rightarrow \underset{\theta\ \phi}{\operatorname{argmin}} \sum_t \tilde{a}_t^2$$

Parameter Estimation

CLSE : ARMA(p,q)

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + a_t - \theta_1 a_{t-1} \ldots - \theta_q a_{t-q}$$

$$\phi = (\phi_1, \ldots, \phi_p)'$$
$$\theta = (\theta_1, \ldots, \theta_q)'$$

ARMA(1,1)과 같이 정리 후 최소를 만족하는 argmin식 유도!

5

Prediction

Prediction

AR(1)

$$\boxed{X_t = \phi X_{t-1} + a_t, a_t \sim WN}$$

이때, $\widehat{\boldsymbol{X}}_{t+1}$는 어떻게 구하는 걸까?

Prediction

AR(1)

$$\boxed{X_t = \phi X_{t-1} + a_t, a_t \sim WN}$$

$$\begin{aligned}
\widehat{X_{t+1}} &= E[\phi X_t + a_{t+1} / X_t, X_{t-1} \cdots] \\
&= \phi X_t \\
\widehat{X_{t+2}} &= E[\phi X_{t+1} + a_{t+2} / X_t, X_{t-1} \cdots] \\
&= \phi E[X_{t+1}] \\
&= \phi^2 X_t
\end{aligned}$$

평균 = 0

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} a_{n+t}$$

$t \to \infty,$

$$X_t = \phi X_t + a_t$$

$$\hat{X}_t = \frac{a_t}{1-\phi}$$

Prediction

MA(1)

$$\boxed{X_t = a_t - \theta a_{t-1},\ a_t \sim WN}$$

X_t 의 평균값을 구할 수 없으므로 **AR(1)의 형태**로 바꾸어서 풀기!

$$X_t = (1-\theta B)a_t$$

$$a_t = (1-\theta B)^{-1}X_t$$

$$= X_t + \theta X_{t-1} + \theta^2 X_{t-2} + \theta^3 X_{t-3} \ldots$$

$$X_t = a_t - \theta X_{t-1} - \theta^2 X_{t-2} - \theta^3 X_{t-3} \ldots$$

Prediction

ARMA(2,2)

$$\boxed{X_t = X_{t-1} - 0.24X_{t-2} + a_t + 0.4a_{t-1} + 0.2a_{t-2}}$$

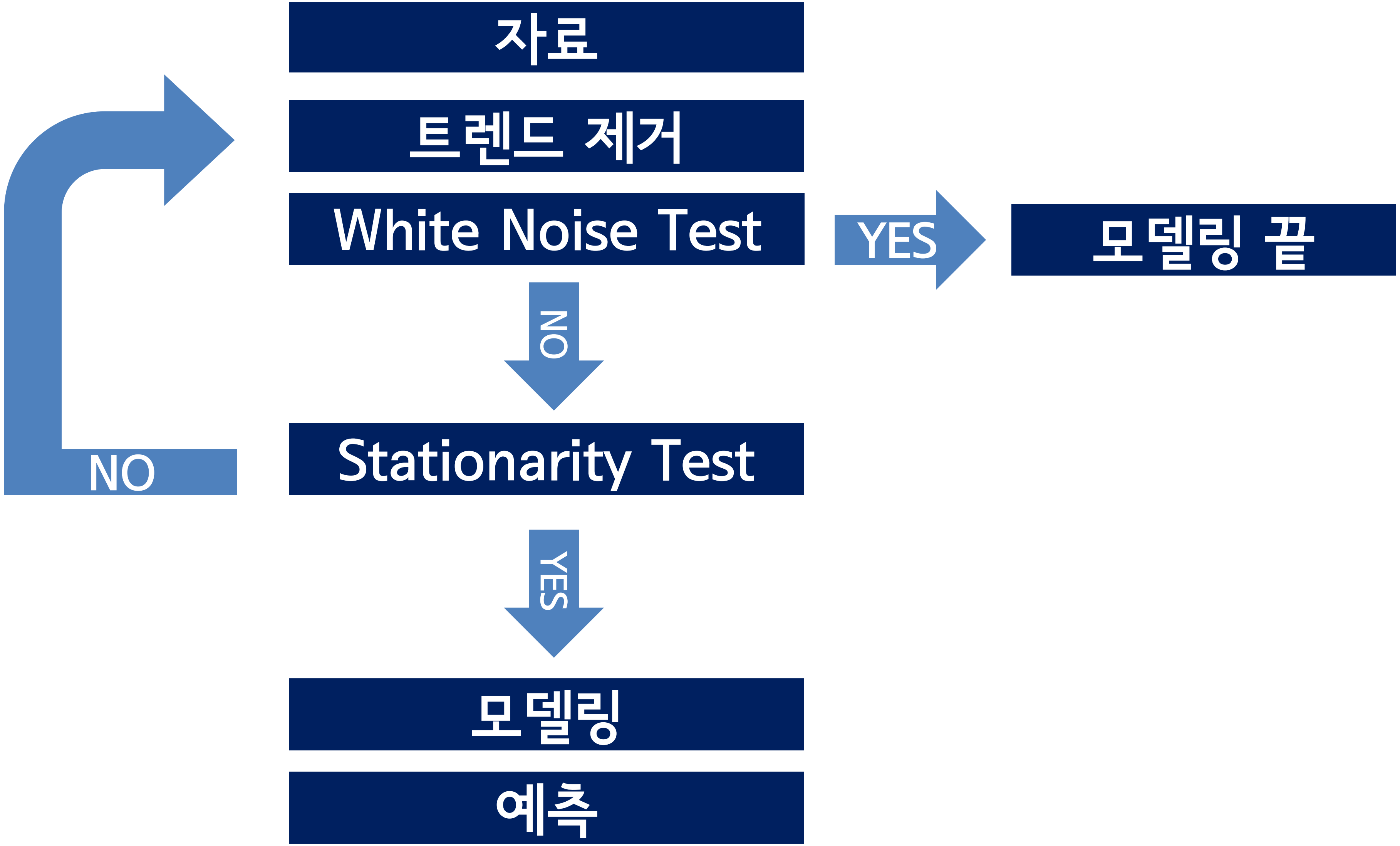
$$\begin{aligned}\widehat{X_{t+1}} &= E[X_t - 0.24X_{t-1} + a_{t+1} + 0.4a_t + 0.2a_{t-1} / X_t \ldots] \\ &= \underbrace{X_t - 0.24X_{t-1}}_{\text{알고 있는 자료}} + 0 + \underbrace{0.4\tilde{a}_t + 0.2\widetilde{a_{t-1}}}_{\text{CLSE 과정에서 구했던 값을 대입!}}\end{aligned}$$

Prediction

Example!

실제 시계열 자료를 분석하고 예측해보자!

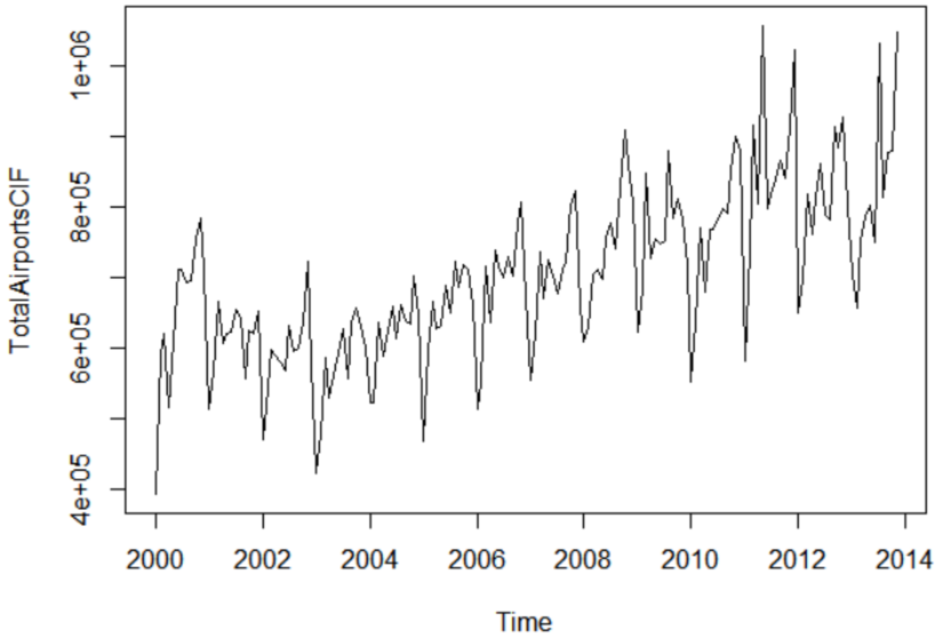
Review

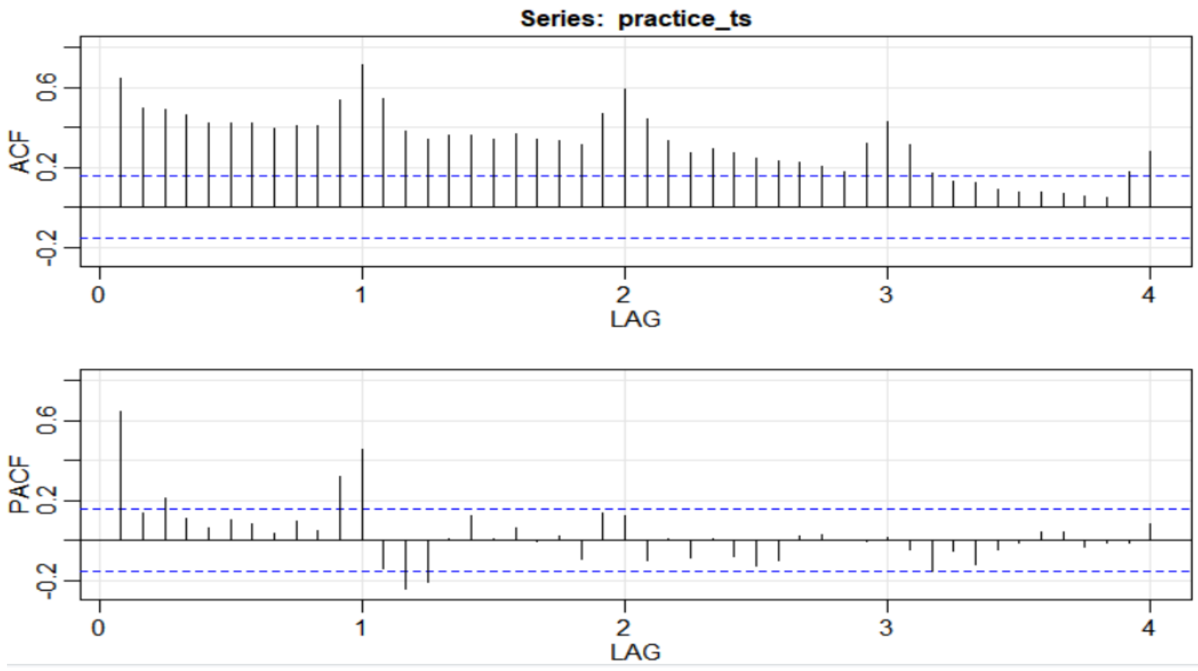

자료
트렌드 제거
White Noise Test
YES
모델링 끝
NO
Stationarity Test
NO
YES
모델링
예측

Prediction

예시 : Airport Data

```r
### Load data and convert dataframe to timeseries data
practice <- read.csv(file='_imports.csv')
practice <- practice %>% select( TotalAirportsCIF ) ## choose 1 variable
practice_ts <- ts( practice, frequency = 12, start = c( 2000, 1 ) ) ## dataframe to timeseries data
plot.ts( practice_ts )
acf2( practice_ts) # seasonality..?
```



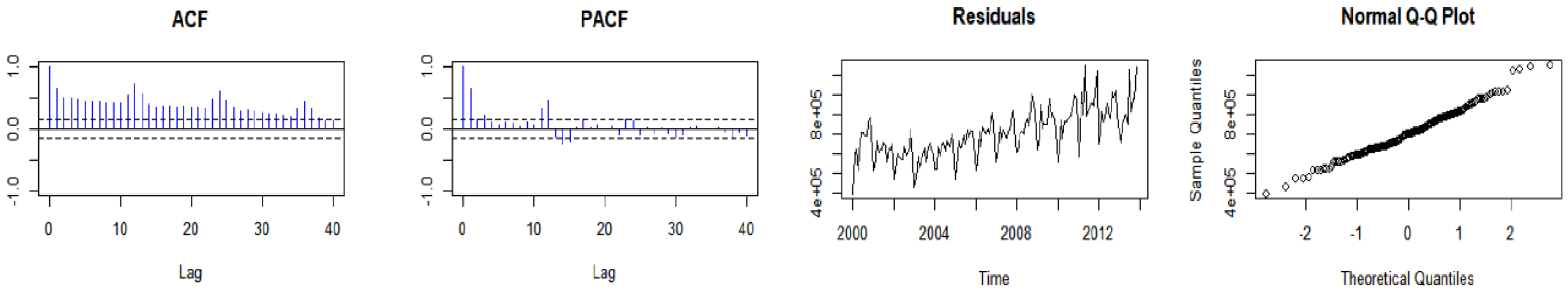


Prediction

예시 : Airport Data

```r
### IID / WN test for data
test( practice_ts ) # data is not iid / WN
```

```
> test( practice_ts ) # data is not iid / WN
Null hypothesis: Residuals are iid noise.
Test                        Distribution Statistic   p-value
Ljung-Box Q                Q ~ chisq(20)    739.15         0 *
McLeod-Li Q                Q ~ chisq(20)    704.14         0 *
Turning points T    (T-110)/5.4 ~ N(0,1)        96    0.0098 *
Diff signs S         (S-83)/3.7 ~ N(0,1)        94    0.0033 *
Rank P          (P-6930.5)/361.3 ~ N(0,1)    10625         0 *
```



Prediction

예시 : Airport Data

```
### Stationarity test for data
adf.test( practice_ts ) # data is stationary
kpss.test( practice_ts ) # data is non-stationary

adf.test( diff( practice_ts ) ) #data is stationary
kpss.test( diff( practice_ts ) ) #data is stationary
```

adf.test의 H0 = Data is not stationary

kpss.test의 H0 = Data is stationary

```
> kpss.test( practice_ts ) # data is non-stationary

        KPSS Test for Level Stationarity

data:  practice_ts
KPSS Level = 4.0018, Truncation lag parameter = 2, p-value = 0.01
```

```
> kpss.test( diff( practice_ts ) ) #data is stationary

        KPSS Test for Level Stationarity

data:  diff(practice_ts)
KPSS Level = 0.024508, Truncation lag parameter = 2, p-value = 0.1
```

Prediction

예시 : Airport Data

R

```
### Stationarity test for data
adf.test( practice_ts ) # data is stationary
kpss.test( practice_ts ) # data is non-stationary

adf.test( diff( practice_ts ) ) #data is stationary
kpss.test( diff( practice_ts ) ) #data is stationary
```

```
> adf.test( practice_ts ) # data is stationary

        Augmented Dickey-Fuller Test

data:  practice_ts
Dickey-Fuller = -4.4567, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

```
> kpss.test( practice_ts ) # data is non-stationary

        KPSS Test for Level Stationarity

data:  practice_ts
KPSS Level = 4.0018, Truncation lag parameter = 2, p-value = 0.01
```

```
> adf.test( diff( practice_ts ) ) #data is stationary

        Augmented Dickey-Fuller Test

data:  diff(practice_ts)
Dickey-Fuller = -7.9675, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

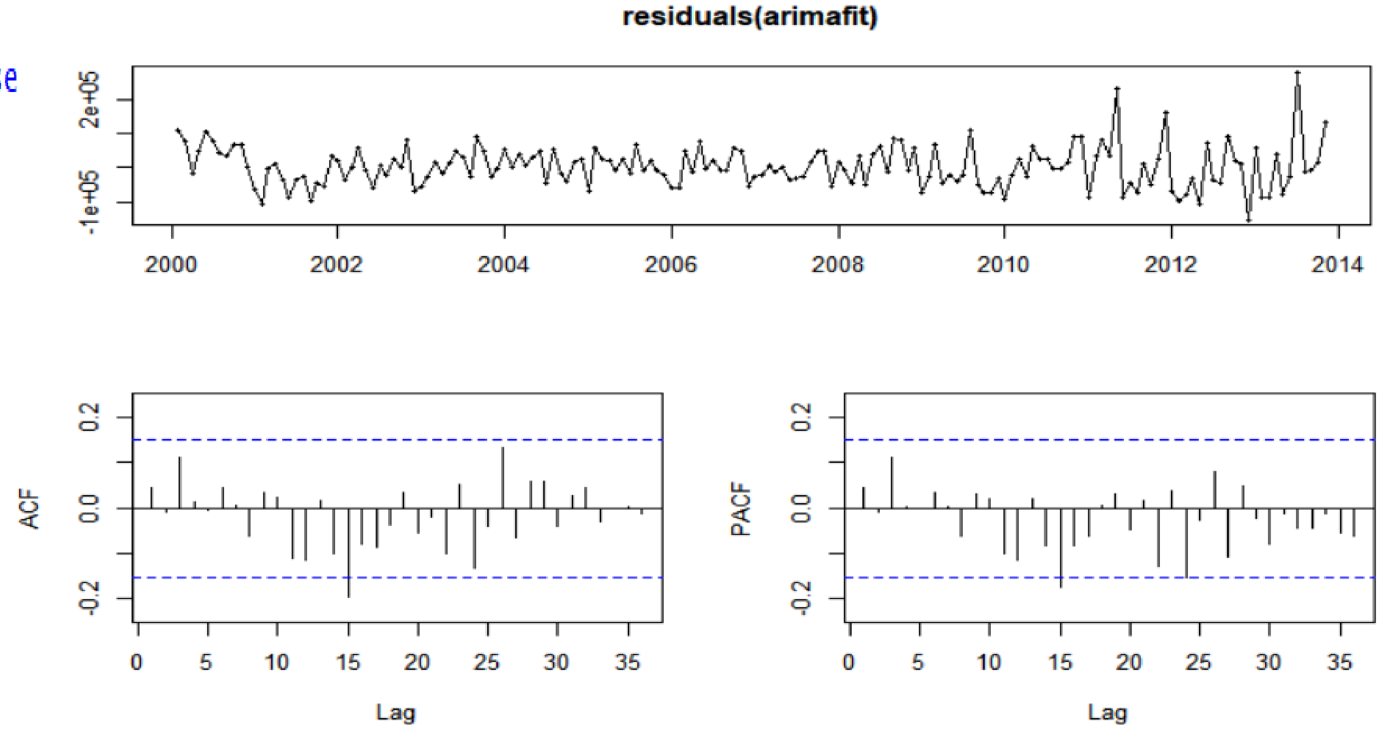
```
> kpss.test( diff( practice_ts ) ) #data is stationary

        KPSS Test for Level Stationarity

data:  diff(practice_ts)
KPSS Level = 0.024508, Truncation lag parameter = 2, p-value = 0.1
```

Prediction

예시 : Airport Data

```r
arimafit <- auto.arima( diff( practice_ts ), ic = 'aicc' )
arimafit ## aic is lower than raw data model, so we will use this model.
tsdisplay( residuals( arimafit ) )
test( residuals( arimafit ) ) ##residuals are IID.
```

```
> arimafit ## aic is lower than raw data model, so we will use
Series: diff(practice_ts)
ARIMA(0,0,1)(2,0,0)[12] with zero mean

Coefficients:
          ma1     sar1    sar2
      -0.7403   0.4047  0.3933
s.e.   0.0665   0.0804  0.0856

sigma^2 estimated as 3.791e+09:  log likelihood=-2070.61
AIC=4149.22   AICc=4149.47   BIC=4161.67
```



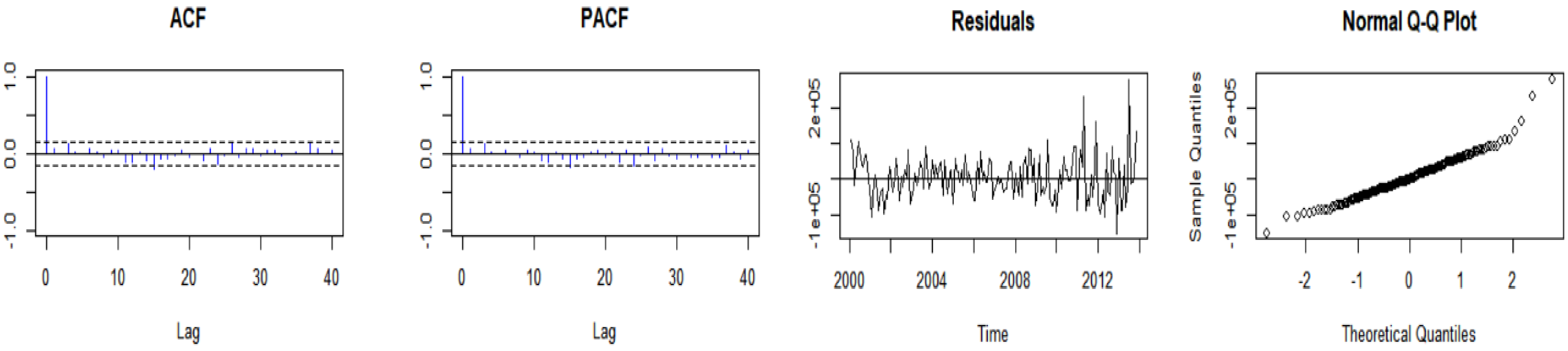
Prediction

예시 : Airport Data

R

```
arimafit <- auto.arima( diff( practice_ts ), ic = 'aicc' )
arimafit ## aic is lower than raw data model, so we will use this
tsdisplay( residuals( arimafit ) )
test( residuals( arimafit ) ) ##residuals are IID.
```

```
> test( residuals( arimafit ) ) ##residuals are IID.
Null hypothesis: Residuals are iid noise.
Test                        Distribution Statistic   p-value
Ljung-Box Q                 Q ~ chisq(20)     20.87   0.4051
McLeod-Li Q                 Q ~ chisq(20)     38.94   0.0068 *
Turning points T  (T-109.3)/5.4 ~ N(0,1)        105   0.4225
Diff signs S       (S-82.5)/3.7 ~ N(0,1)         85   0.5028
Rank P        (P-6847.5)/358.1 ~ N(0,1)       6416   0.2282
```

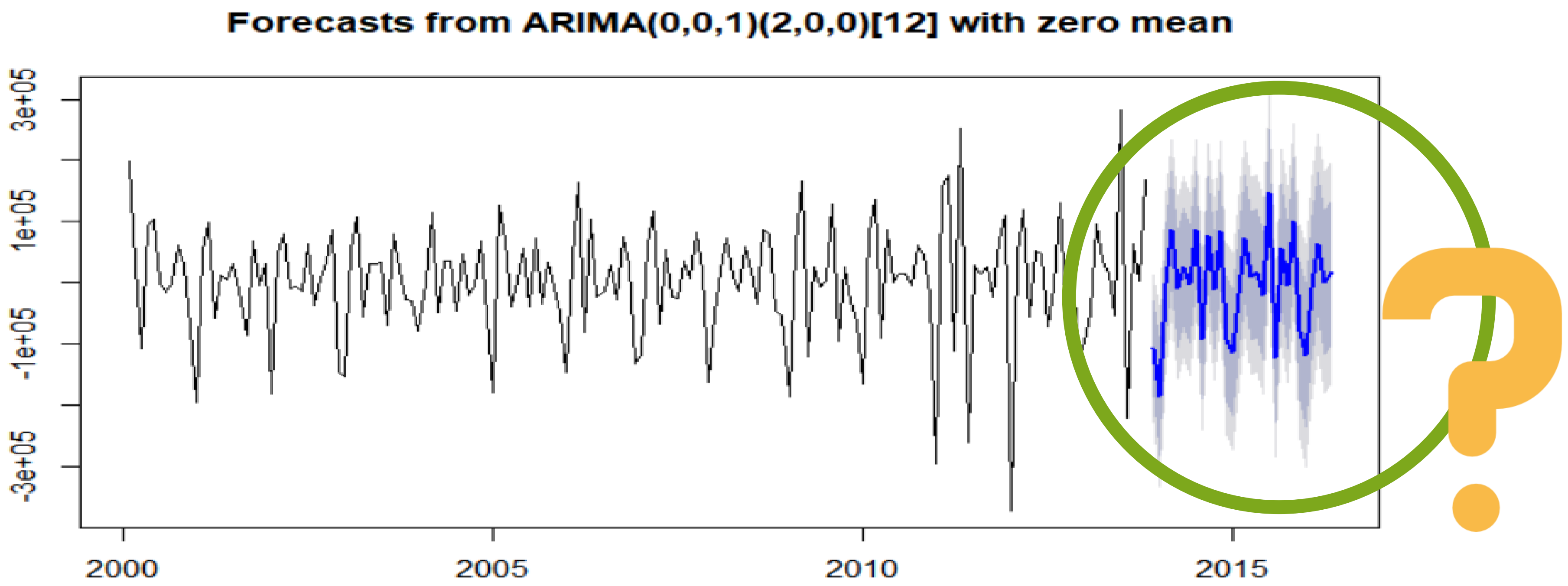


Prediction

예시 : Airport Data

R

```
practice_forecast <- forecast( arimafit, h = 30 )
plot(practice_forecast)
```



Prediction

Predict Interval

ex) AR(2)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t$$

미래값의 예측(prediction)에는 오차가(predict interval) 존재!

$$\widehat{X_{t+1}} = \phi_1 X_t + \phi_2 X_{t-1}$$

$$X_{t+1} - \widehat{X_{t+1}} = a_{t+1} \sim \ WN(0, \sigma_a^2)$$

$$\widehat{X_{t+1}} \pm z_{\alpha/2} \sigma_a^2$$

Prediction

Predict Interval

ex) AR(2)

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + a_t$$

σ_a^2 은 어떻게 구하는 거지?

미래값의 예측(prediction)에는 오차가(predict interval) 존재!

$$\widehat{X_{t+1}} = \phi_1 X_t + \phi_2 X_{t-1}$$

$$X_{t+1} - \widehat{X_{t+1}} = a_{t+1} \sim \quad WN(0, \sigma_a^2)$$

$$\widehat{X_{t+1}} \pm z_{\alpha/2} \sigma_a^2$$

Prediction

Predict Interval

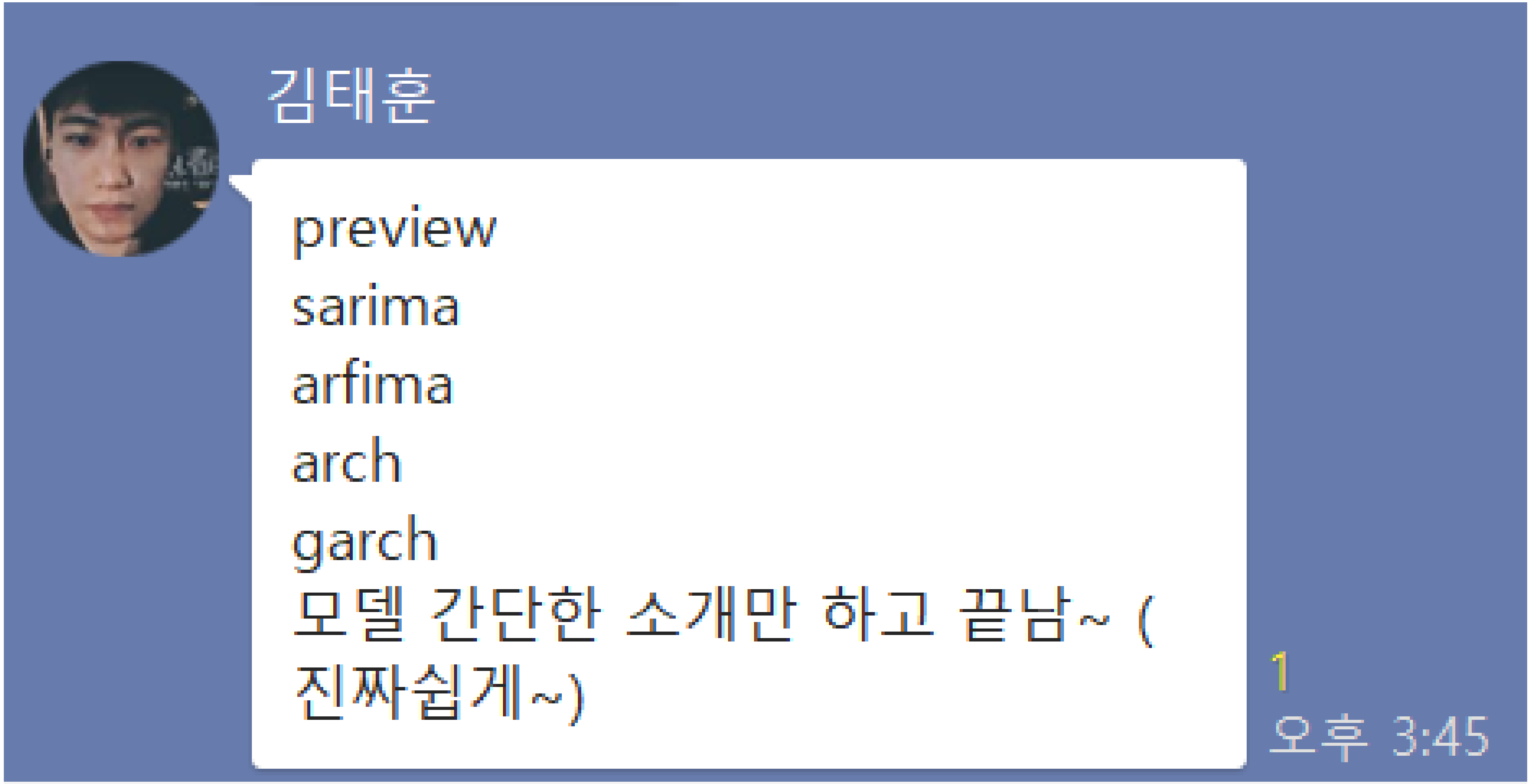
$$\sigma_a^2$$

Yule Walker Equation

$$\sqrt{\frac{SSR}{n-(p+q)}}$$

ARMA(p,q)

Preview


김태훈
preview
sarima
arfima
arch
garch
모델 간단한 소개만 하고 끝남~ (
진짜쉽게~)
1
오후 3:45