

Chapter. 02

시계열 데이터, 이것만은 꼭 분석하자

데이터 안정성 보장 필요성

FAST CAMPUS
ONLINE

직장인을 위한
파이썬 데이터분석
강사. 주세민

Chapter. 02

시계열 데이터, 이것만은 꼭 분석하자

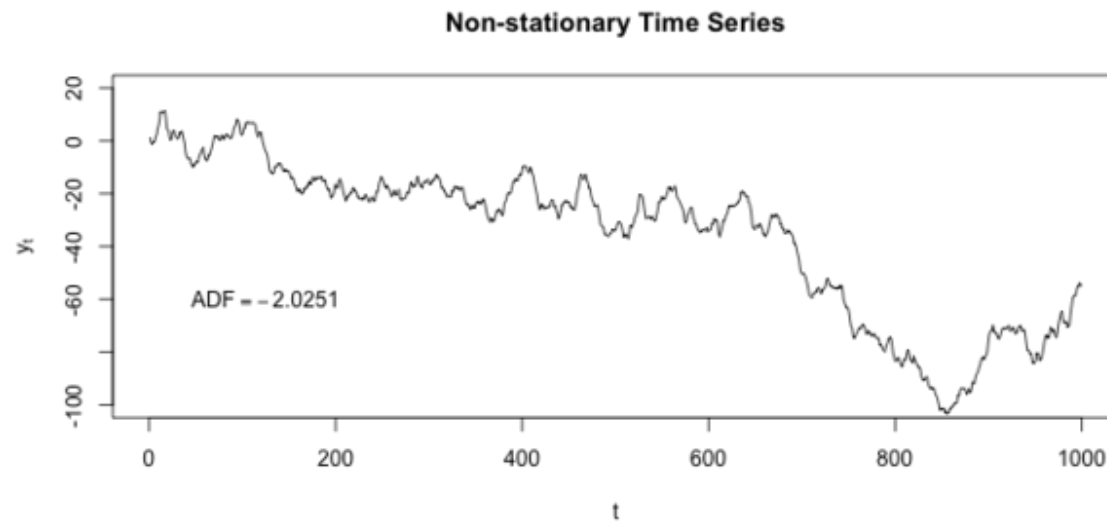
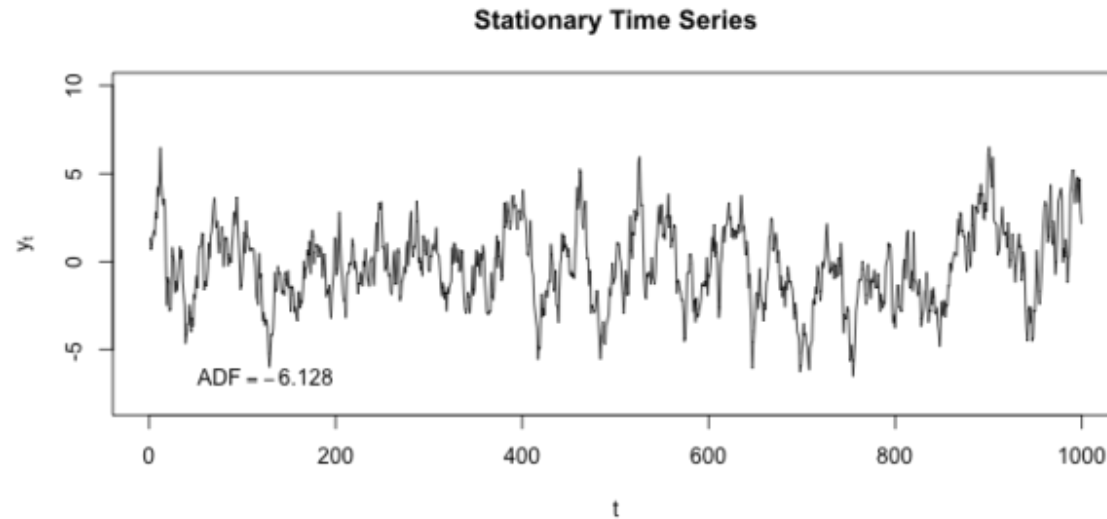
I 시계열 특성 > 안정성 (Stationarity)

“시계열이 Stationary하다”

**= 시계열 데이터가 미래에 똑같은 모양일
확률이 매우 높다**

**즉, 시계열이 안정적이지 않으면 현재의 패턴이 미래에 똑같이
재현되지 않으므로, 그대로 예측 기법 적용하면 안된다.**

I 안정한 시계열과 안정하지 않는 시계열은 대략 눈으로도 파악 가능합니다.



I 불안정한 시계열을 그대로 예측에 활용하는 경우...

설명력 (R^2) 90% 이상, 정확도 90% 이상 나옴.
그러나, Spurious regression (가성적 회귀), Overfitting

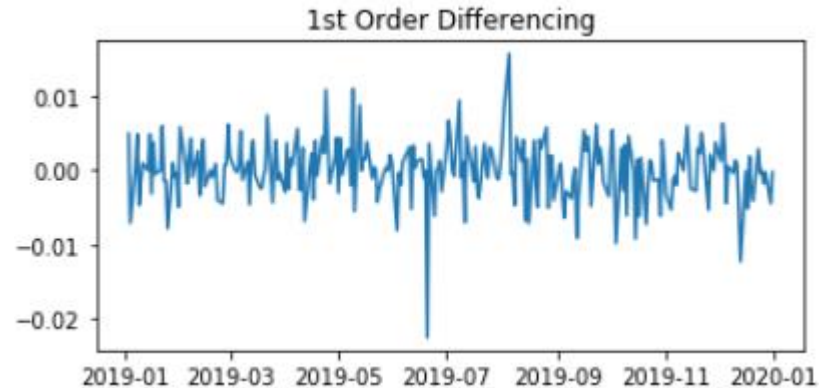
→ Granger and Newbold (1974) 이후 시계열 데이터 분석의 discipline으로 자리잡고 있음

따라서,

원계열('환율')보다는



차분 데이터 ('환율 증감') 사용



I 시계열 안정성 판별법

“(Augmented) Dickey Fuller Test”

- 귀무가설 (Null Hypothesis) : 원계열은 안정적이지 않다.

– 테스트 방법

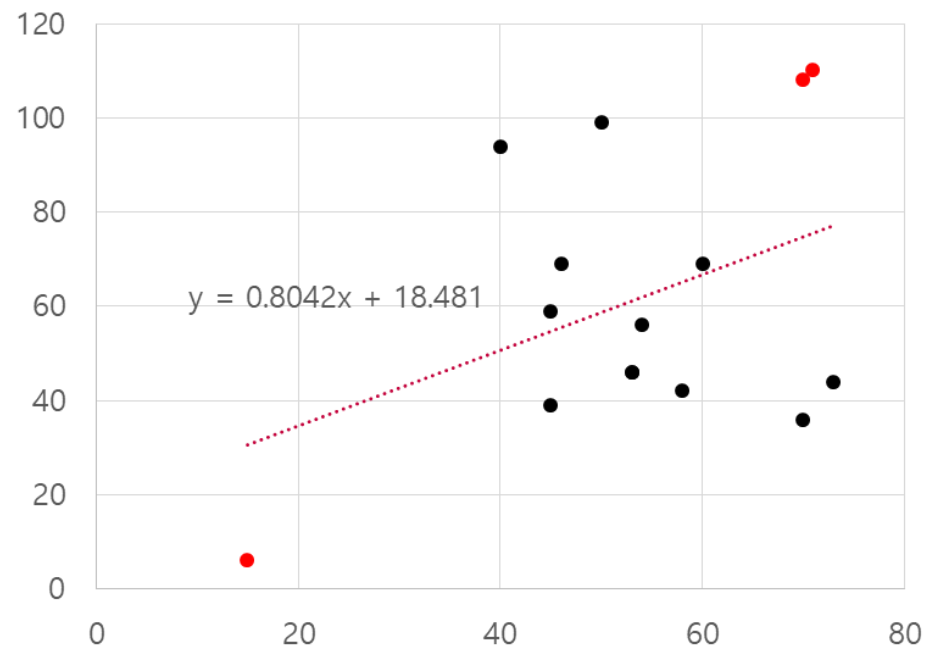
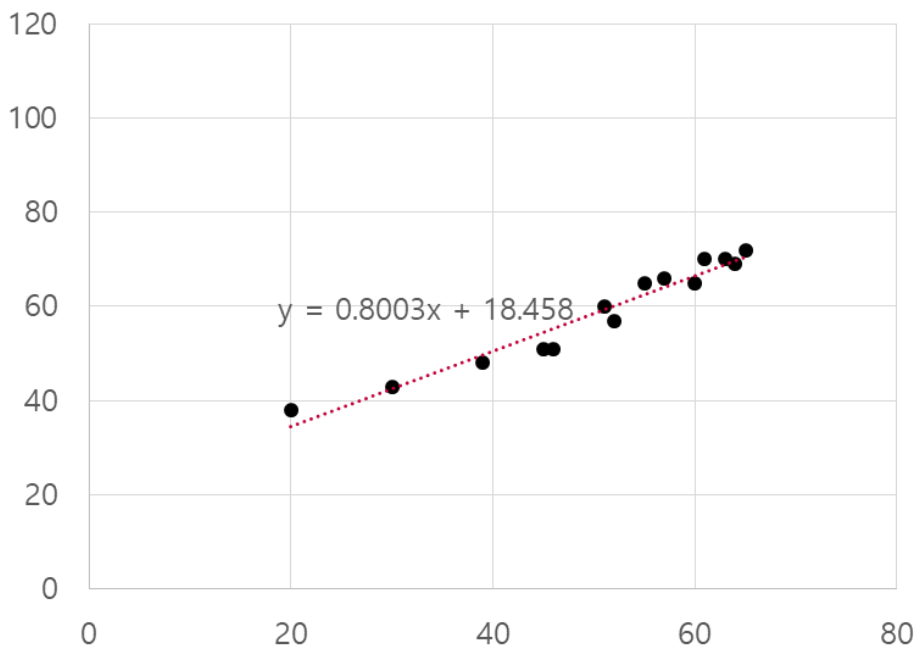
$$\Delta y_t = (\rho - 1)y_{t-1} + u_t = \delta y_{t-1} + u_t \quad \text{testing } \delta = 0$$

– 테스트 결과 해석

: p-value가 0.05보다 작으면, 귀무가설 기각. 즉, 안정적인 시계열

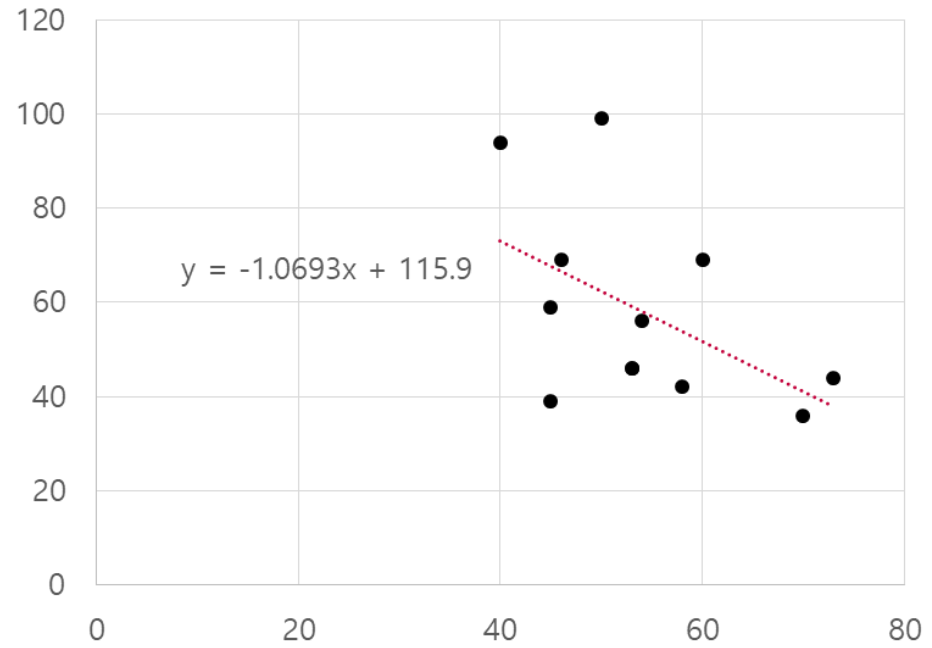
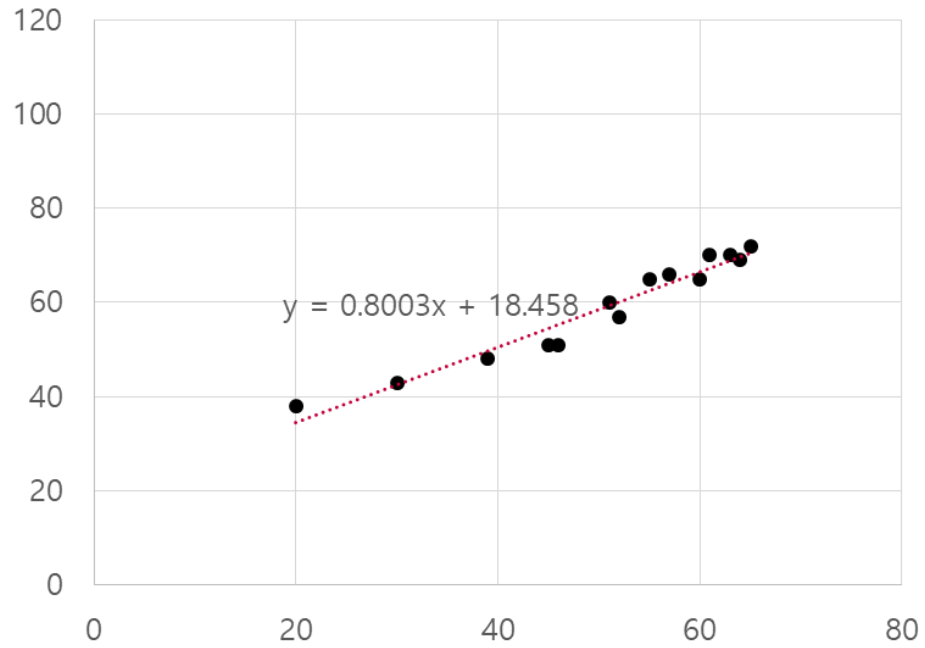
: P-value가 0.05보다 크면, 귀무가설 채택. 즉, 불안정한 시계열

I 참고 : p-value



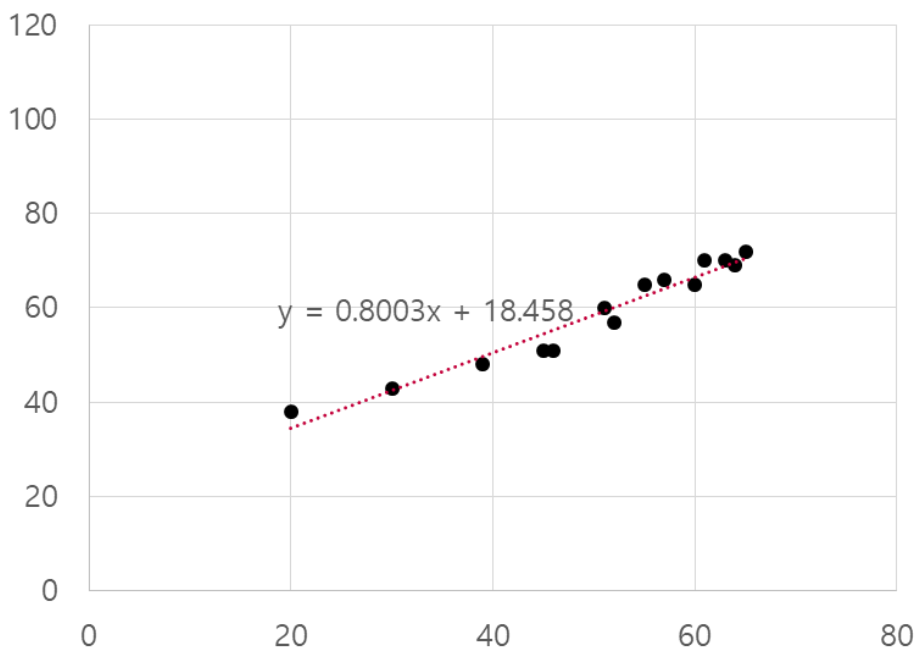
1) 데이터 출처 : 빅데이터를 지배하는 통계의 힘 (2013)

I 참고 : p-value

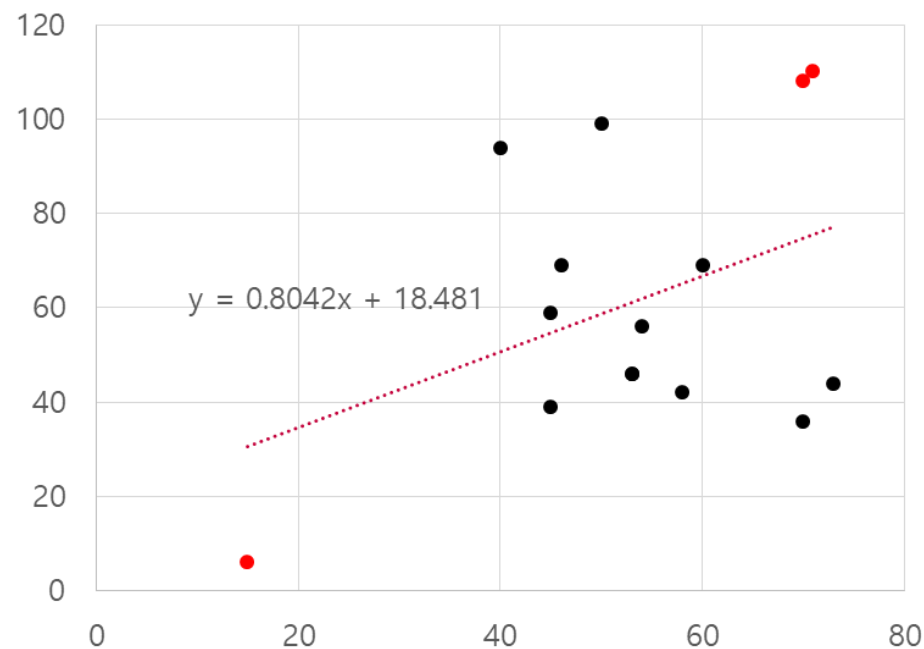


I 참고 : p-value

즉, 결과값에 대한 p-value는 정말 특별한 법칙을 따르는 것인지 (p-value 작음)
우연히 이런 결과가 나왔는지 (p-value 큼) 척도



변수	추정치	표준오차	95%신뢰구간	p-Value
절편	18	1.5	14.9~21.2	<0.001
x 계수	0.8	0.03	0.7~0.9	<0.001



변수	추정치	표준오차	95%신뢰구간	p-Value
절편	18	35	-55.5~91.5	0.61
x 계수	0.8	0.6	-0.5~2.1	0.23

I 시계열 안정성 판별법의 p-value 해석을 이해하셨나요?

“(Augmented) Dickey Fuller Test”

- 귀무가설 (Null Hypothesis) : 원계열은 안정적이지 않다.

– 테스트 방법

$$\Delta y_t = (\rho - 1)y_{t-1} + u_t = \delta y_t$$

p-value가 0.05보다 작다는 말은 특이하게도
(?) 안정적이라는 말임

– 테스트 결과 해석

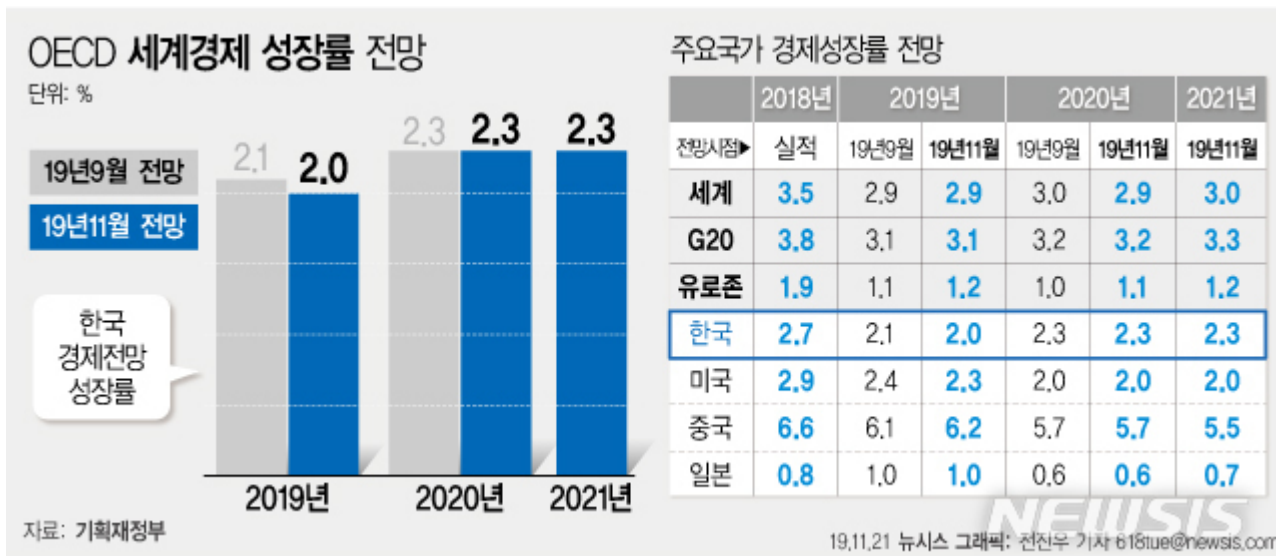
: p-value가 0.05보다 작으면, 귀무가설 기각. 즉, 안정적인 시계열

: P-value가 0.05보다 크면, 귀무가설 채택. 즉, 불안정한 시계열

I 불안정한 시계열을 안정적인 시계열로 변경하는 가장 보편적인 방법은 성장률로 변환해서 예측하는 것입니다.

- 대부분의 시계열 예측은 ‘변화’를 예측합니다.

- Log difference도 성장율은 나타냅니다.



$$\text{성장률} = \frac{y_{t+1} - y_t}{y_t}$$

$$= \log y_{t+1} - \log y_t$$

[참고:증명]

$$\frac{d \log y_t}{dt} = \frac{d \log y_t}{dy_t} \frac{dy_t}{dt} = \frac{dy_t / dt}{y_t}$$

I 다음 강의에서 다룰 내용

- 시계열 안정성 테스트 및 해석
 - Augmented Dickey Fuller Test
- 불안정한 데이터를 안정적 데이터로 변환
 - 성장률, log difference