

Part.02  
회귀분석

# | 회귀분석의 진단

FASTCAMPUS  
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택

# I 회귀분석의 진단

## ■ 회귀분석의 진단

- 적절한 변수를 통해 어느정도 성능지표가 잘 나오는 모델을 만들었다.
- 과연 이 회귀모델이 잘 만들어진 모델인것인가에 대한 진단이 필요.
- 회귀분석에서는 아래 잔차에 대한 세가지 가정이 존재 : 정규성, 독립성, 등분산성

$$e_i = y_i - \hat{y}_i$$

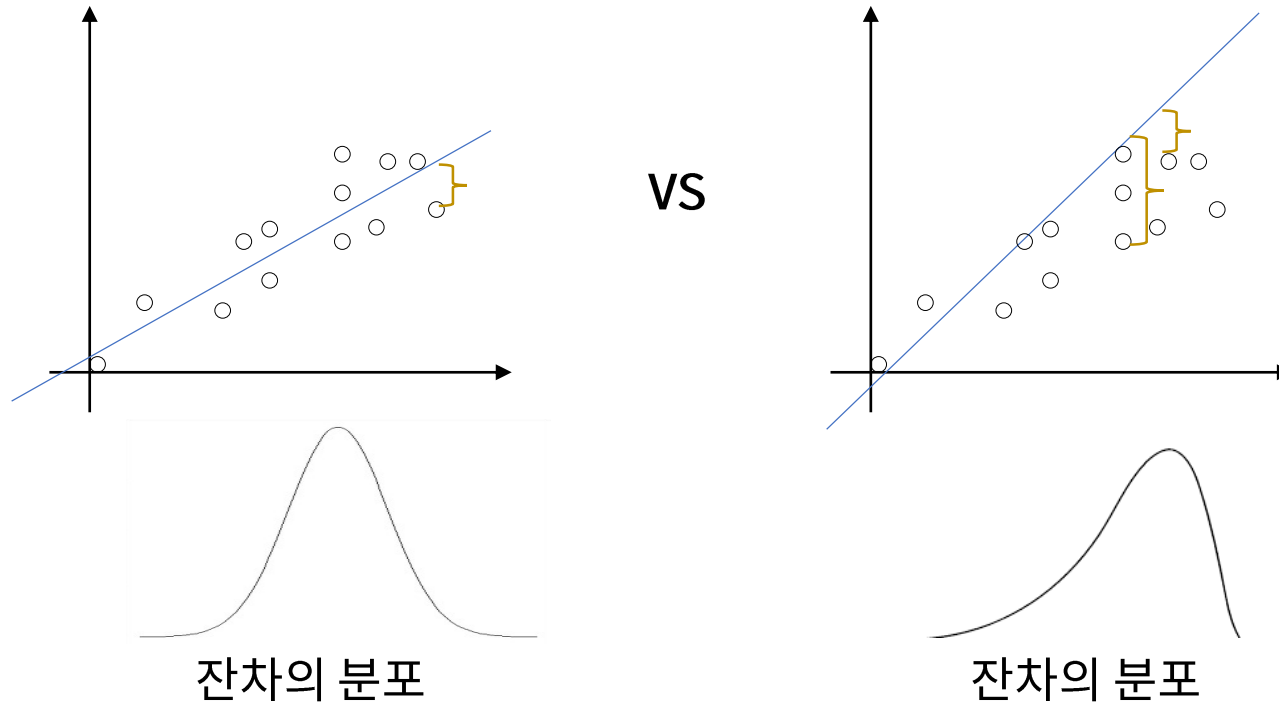
- 세가지 가정을 만족할 시 잘 만들어진 회귀모델이라 판단.

## ■ 회귀분석의 가정 :

$$\varepsilon_i \sim i.i.d N(0, \sigma^2), Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2), X \text{와 } Y \text{는 선형관계}$$

# I 회귀분석의 진단

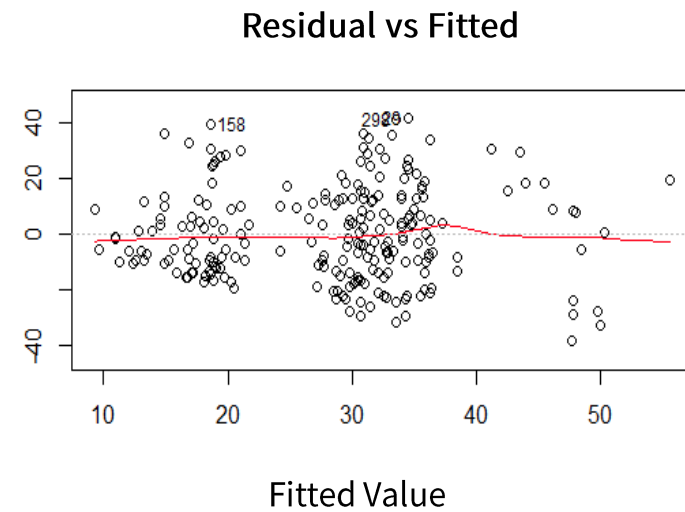
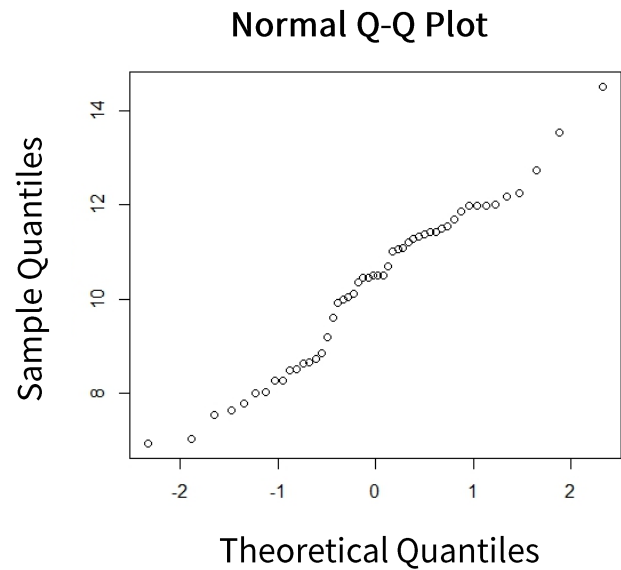
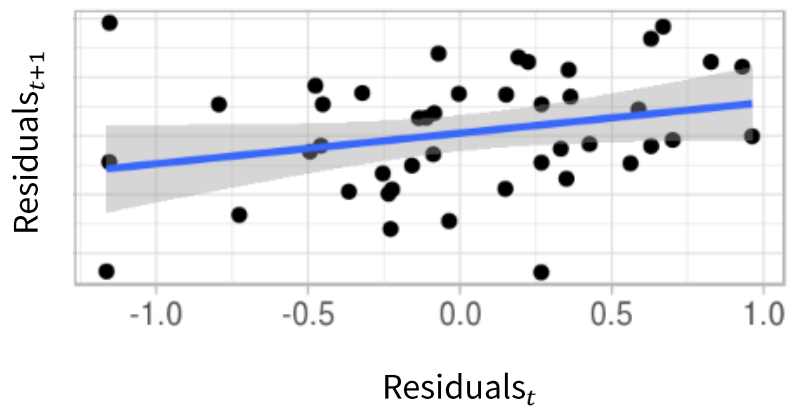
## 회귀분석의 진단



- 회귀모델을 잘 만들었을 경우, 잔차는 정규분포를 따른다.

# I 회귀분석의 진단

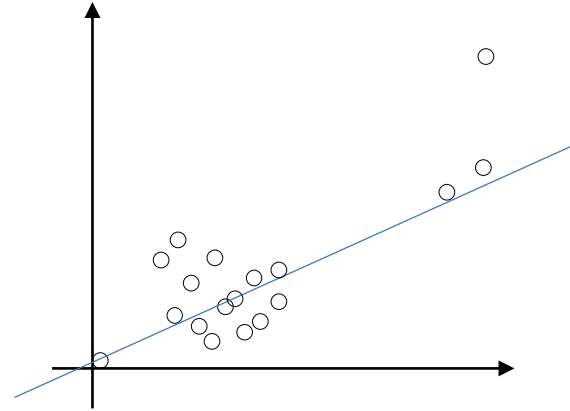
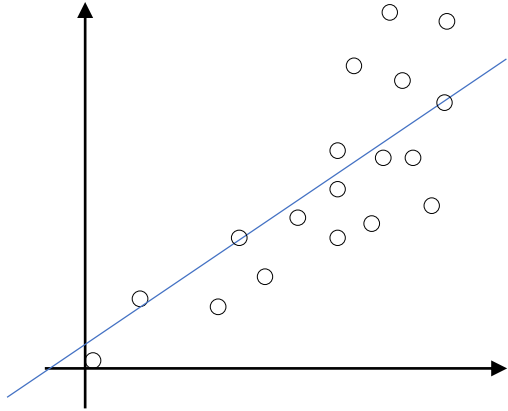
## ■ 회귀분석의 진단



- 일반적으로 Residuals산점도, Normal Q-Q Plot과 Residual vs fitted plot으로 진단

# I 회귀분석의 진단

## ■ 잔차가 가정에 위배된 경우



1) Y에 대하여 log또는 root를 씌워 줌

2) 이상치 제거

3) 다항회귀분석

Part.02  
회귀분석

# I 다항회귀분석

FASTCAMPUS  
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택