

# 잔차검진 - Residual Diagnostics

분석에 사용된 회귀모형의 적절성과 통계적 추론의 가정을 만족하는지를 확인하는 방법에 대해 알아보겠습니다.

## 1. 오차항의 가정

$$\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \sim iid N(0, \sigma^2)$$

오차항에서는 세 가지가 가정되어야 합니다.

### (1) 정규성

중심측량이 자유도가  $n-2$ 인  $t$ 분포를 따른다고 유도할 때 데이터가 정규분포를 만족한다고 가정하에  $t$ 분포를 유도합니다.

### (2) 등분산성

← 잔차의 등분산성

MSE는 모든 분산이 동일하다는 가정하에 유도합니다.

회귀분석과 분산분석에서 제일 중요한 것은 등분산성입니다.

### (3) 독립성

Y들이 선형 결합인 상태에서 분산을 유도하는데 이 때 Y가 독립이라고 가정합니다.

## 2. 잔차 - residual

잔차는 관측값과 예측값의 차이를 의미합니다.

이론 값을 도출할 때는 확률변수  $Y$ 를 이용합니다.

$$e_i = Y_i - \hat{Y}_i$$

실제 분석을 할 때는 관측값  $y$ 를 이용합니다.

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

★ 핵심!!!

잔차가 특정한 패턴을 가진다면 모형(설명되는 부분)에 추가해야 할 요소가 남아 있음을 의미합니다.

따라서 잔차는 “특정한 패턴이 없어야 합니다.”

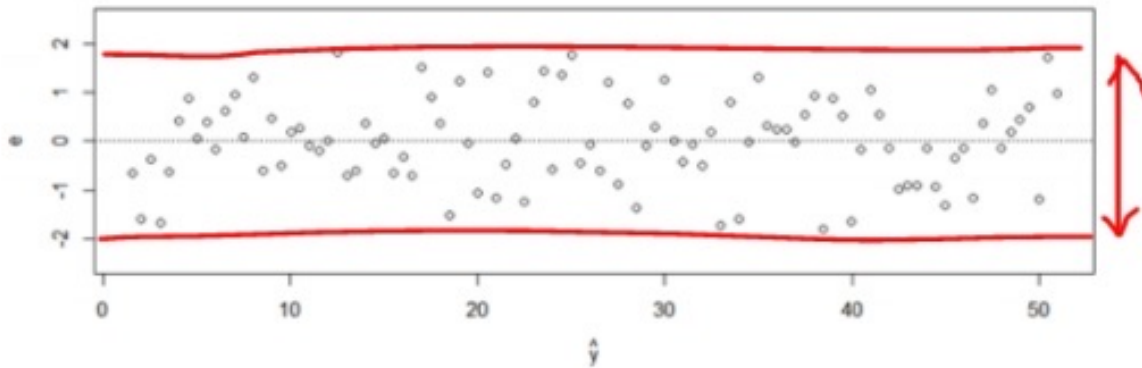
↖ 잔차는 random 해야 함.

잔차가 오차항의 가정을 심각하게 위반하면 통계적 추론에 문제가 발생합니다.

## (1) 정상적인 잔차그림

“등분산성”을 만족하는 잔차그림입니다.

← 시계열 정상성의 등분산성을 파악하는 방법과 동일

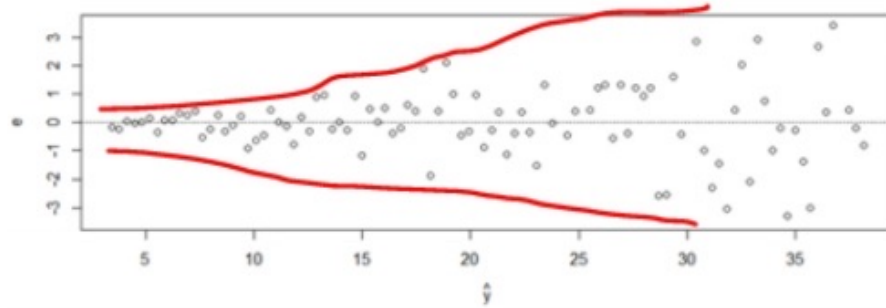


정상적인 잔차그림은 0을 중심으로  $\hat{y}$  값에 관계없이 일정 범위 내에서 **특정한 패턴을 가지지 않게 분포**됩니다.

## (2) 대표적인 비정상적 잔차 그림

잔차그림을 통해 어떤 문제가 있는지 시각적으로 확인하고 해결방법을 찾을 수 있습니다.

### a. $\hat{y}$ 가 커지면서 e의 폭이 커짐



잔차그림을 보았을 때 퍼져있는 정도가 다른 것을 알 수 있습니다.

이 경우에 등분산성을 만족하지 않습니다.

대안으로 <sup>등분산성 위해 시, '변환' 실시</sup> 반응변수의 변환이 있습니다.

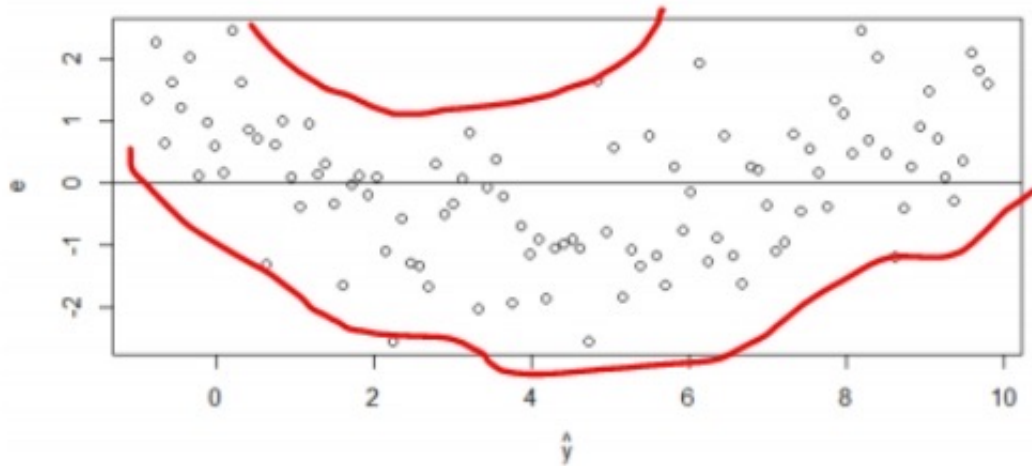
y값을 그대로 사용하지 않고 변환하는 것입니다.

★  $Y_i^* = \log(Y_i), \quad Y_i^* = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim iid N(0, \sigma^2)$

이처럼 log로 변환하면 어느정도 등분산성을 만족하게 됩니다.

b.  $\hat{y}$ 가 커지면서 e가 하강(상승)하다가 상승(하강)

함



이 경우에 설명변수의 제곱항이 생략되어 있을 가능성이 큼니다.

또한 등분산성을 만족하지 않습니다.

대안으로 제곱항을 추가하거나 변수변환방법이 있습니다.

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \quad \varepsilon_i \sim iid N(0, \sigma^2)$$

이처럼 제곱항을 추가하면 어느정도 등분산성을 만족하게 됩니다.