

Chapter06
선형회귀

I 잔차와 레버리지 분석

M T W T F S S

FASTCAMPUS
ONLINE

금융공학/퀀트 I

강사. 장순용

I 키포인트

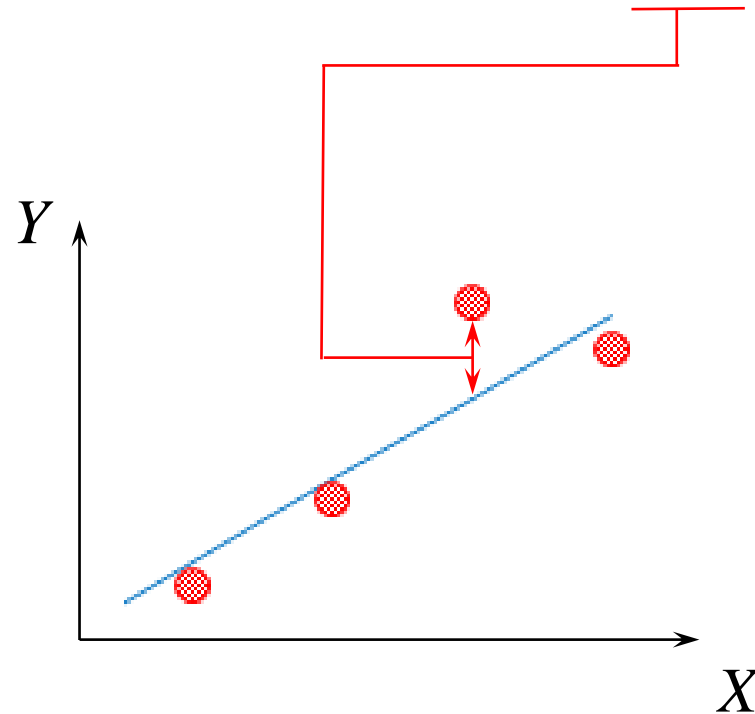
- 잔차와 레버리지 분석을 통한 영향력 분석.
- 쿡의 거리.
- 잔차 분석을 통한 선형회귀의 전제 조건 확인.

I 잔차와 레버리지 개요

- 잔차와 레버리지를 통한 영향력 분석을 하는 이유.
 - ⇒ 잔차 : 종속변수 Y 에서 특이값을 발견할 수 있음.
 - ⇒ 레버리지 : 설명변수 X 에서 특이값을 발견할 수 있음.
 - ⇒ 가장 임팩트가 큰 데이터 포인트 발견.

I 잔차

- 잔차는 모형이 예측하는 \hat{Y} 값과 실제 Y 값 사이의 차이이다.



⇒ 그러므로 Y 의 특이값을 쉽게 찾아낼 수 있다.

I 잔차

- 잔차분석을 통한 선형회귀 전제 조건의 확인:

⇒ 선형성: 종속변수는 설명변수의 선형조합으로 설명이 가능하다.

그러므로, 잔차에는 추세가 없다.

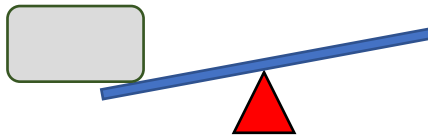
⇒ 독립성: 잔차는 순서와 상관없이 독립적이다.

⇒ 정상성: 잔차의 분포가 정규분포를 따른다.

⇒ 등분산성: 잔차의 분산이 순서와 무관하게 일정하다.

I 레버리지

- 레버리지는 X 값이 중앙에서 얼마나 멀리있는지 나타냄.
- 레버리지가 크다는 것은 회기계수에 영향이 크다는 의미.

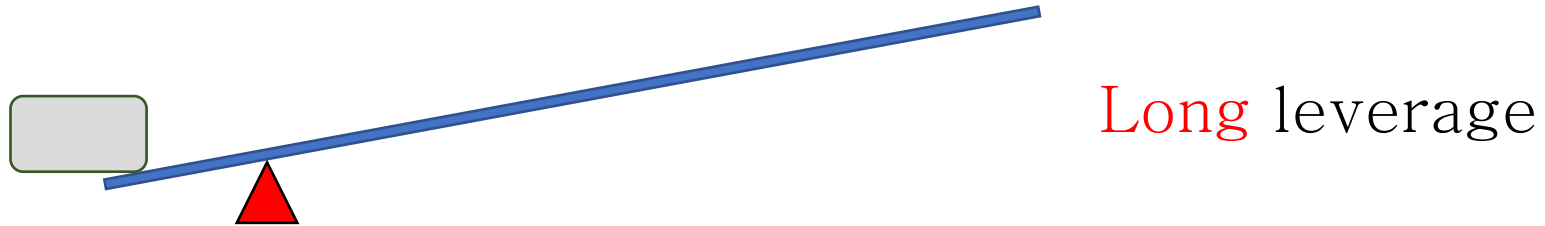


Short leverage

⇒ 그러므로 X 의 특이값을 쉽게 찾아낼 수 있다.

I 레버리지

- 레버리지는 X 값이 중앙에서 얼마나 멀리있는지 나타냄.
- 레버리지가 크다는 것은 회기계수에 영향이 크다는 의미.



⇒ 그러므로 X 의 특이값을 쉽게 찾아낼 수 있다.

I 레버리지

- i 번째 관측값의 레버리지:

$$\text{Leverage} = H_{ii} \longleftarrow \tilde{H} = \tilde{X}(\tilde{X}^t \tilde{X})^{-1} \tilde{X}^t$$

- “Sum rule”:

$$\sum_{i=1}^n H_{ii} = p \longleftarrow \text{파라미터의 개수}$$

I 레버리지

- 레버리지의 상대적 크기를 판단하는 기준:

$$\text{평균의 레버리지} \cong \frac{p}{n}$$

$$\text{큰 레버리지} > \frac{p}{n}$$

$$\text{작은 레버리지} < \frac{p}{n}$$

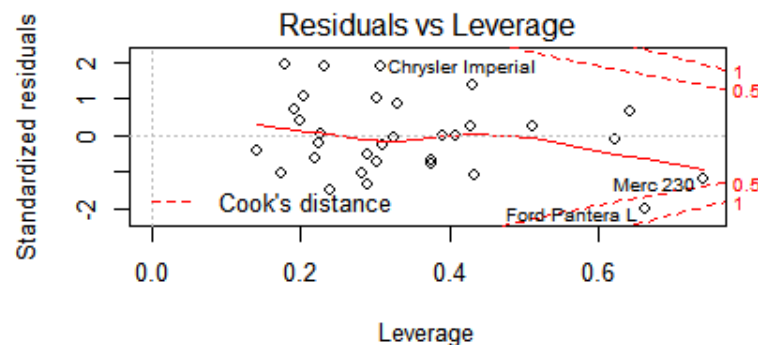
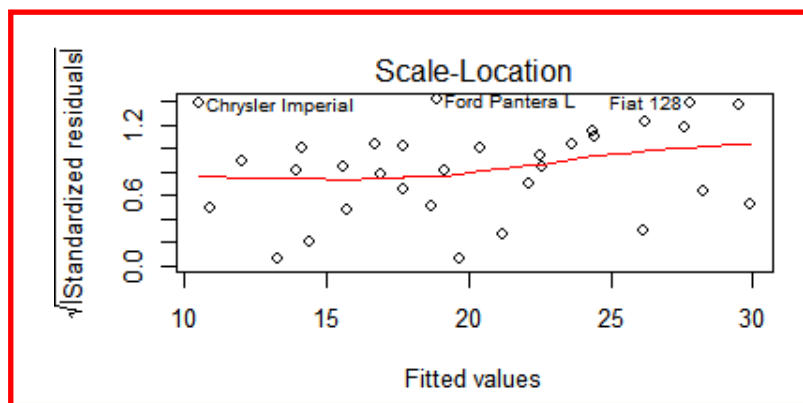
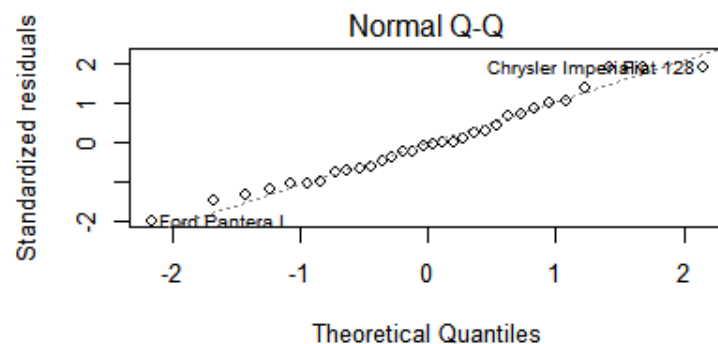
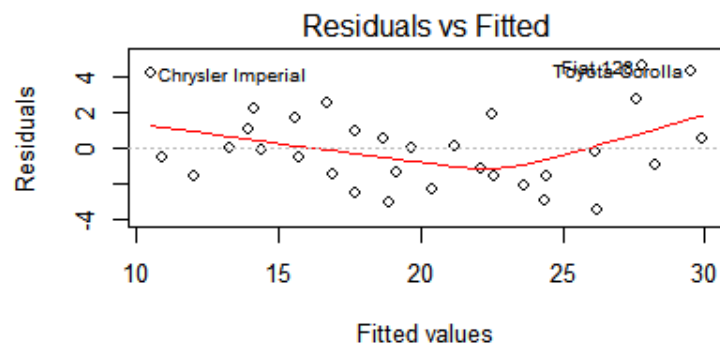
I 쿡의 거리

- i 번째 관측값의 쿡의 거리 (Cook's Distance):

$$D_i = \frac{e_i^2}{P \times MSE} \left[\frac{H_{ii}}{(1 - H_{ii})^2} \right]$$

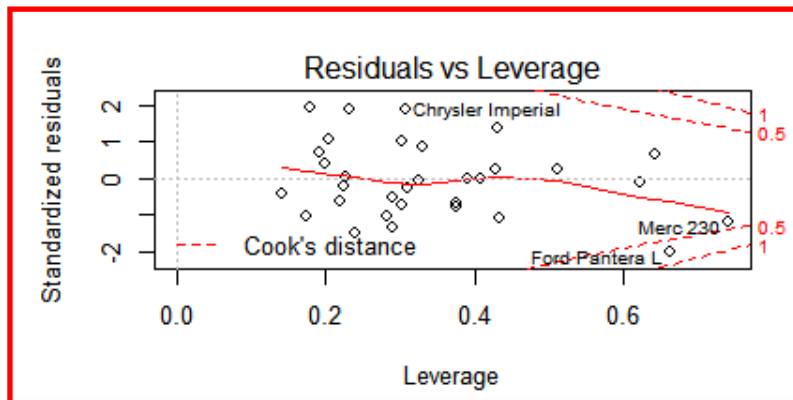
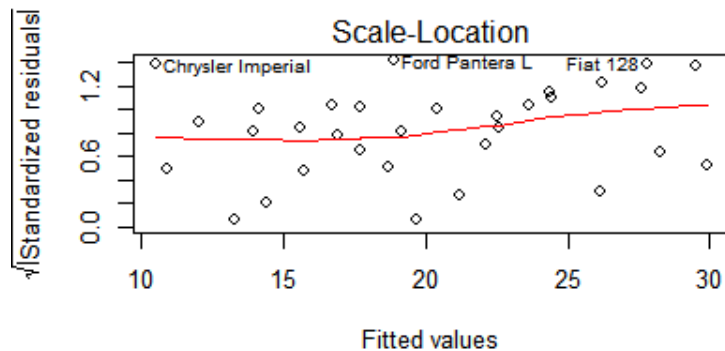
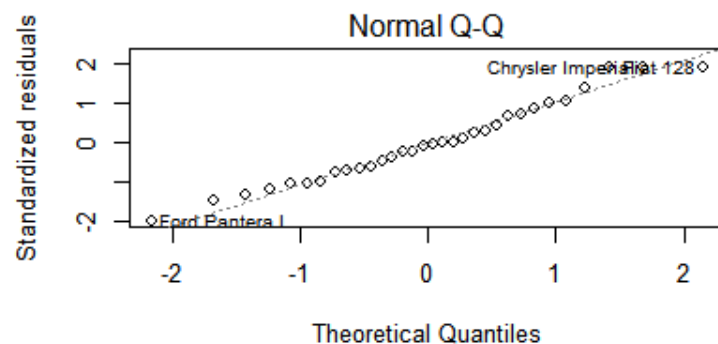
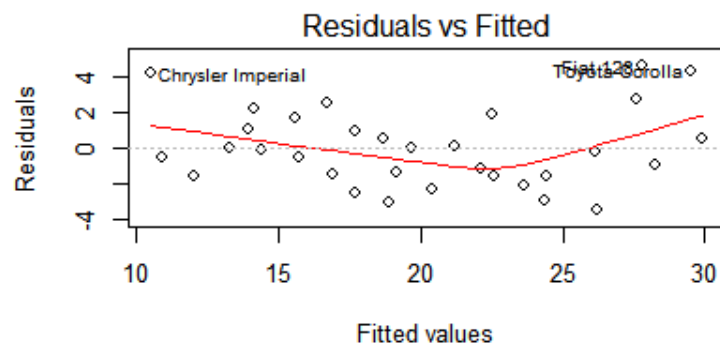
- 쿡의 거리는 전체적인 영향력을 나타내어 줌.
- 잔차와 레버리지의 개념의 “혼합”과도 유사함.

I 시각화를 통한 외상치와 영향력 확인



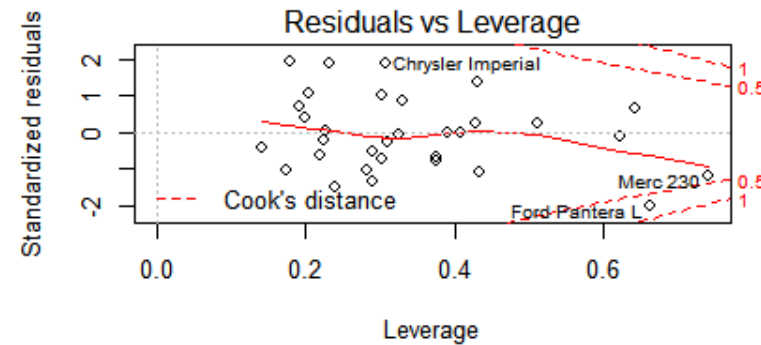
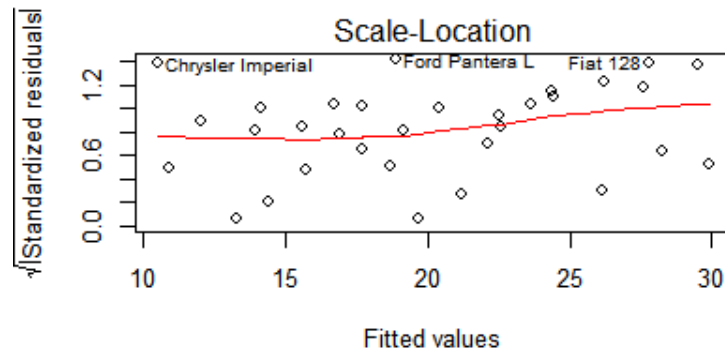
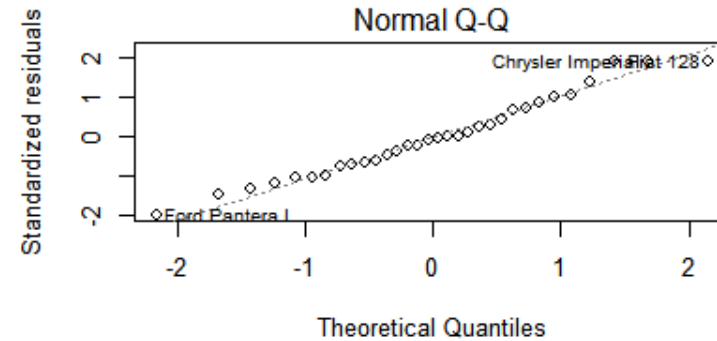
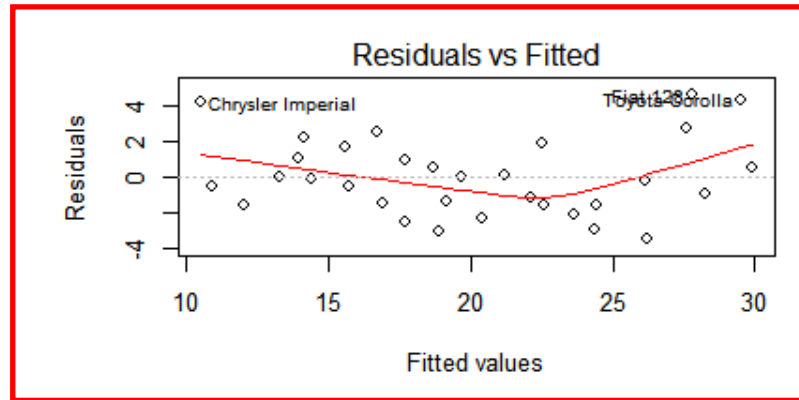
⇒ 시각적으로 외상치 (표준화된 잔차) 확인.

I 시각화를 통한 외상치와 영향력 확인



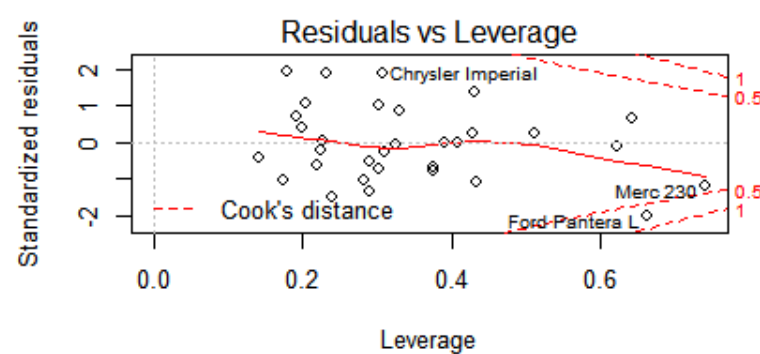
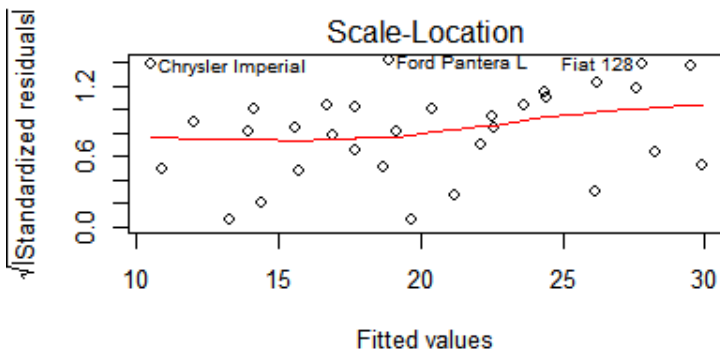
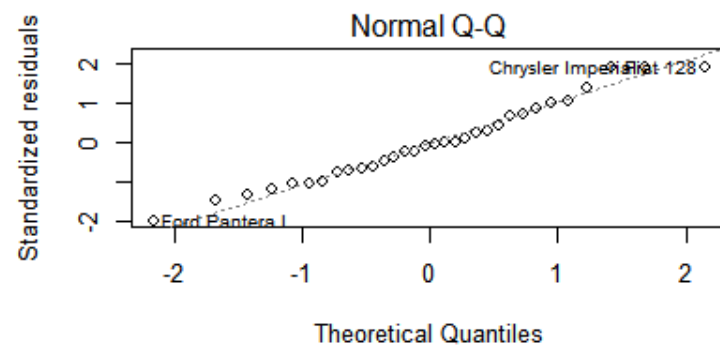
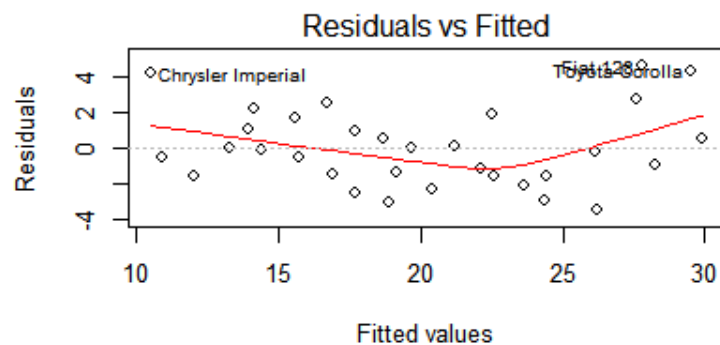
⇒ 시각적으로 **영향력** (레버리지, 쿡의 거리) 확인.

I 시각화를 통한 선형회귀 전제조건 확인



⇒ 시각적으로 선형성, 독립성, 등분산성 확인.

I 시각화를 통한 선형회귀 전제조건 확인



⇒ QQ plot을 사용해서 시각적으로 정상성 확인.

I 끝.

감사합니다.

