

- OLS를 이용하면 회귀분석 가정이 맞지 않아도 회귀의 가중치는 계산할 수 있으나, 신뢰도는 파악할 수 없기에 추정오차가 있을 수 있다.  
즉, 가중치의 오차 범위 혹은 신뢰 구간을 계산할 수 없다면 OLS 결과도 부정  
신뢰적인 결론을 이끌어내기 어렵다.

## ○ 레버리지와 아웃라이어

- 개별적인 데이터 포인트 하나 하나가 회귀분석 결과에 미치는 영향력은 레버리지 분산이나 아웃라이어 분산은 함께 만족 받는다.
- 레버리지 (Leverage)는 실제 관측값  $y_i$ 가 예측치 (Prediction target)  $\hat{y}_i$ 에 미치는 영향  
력을 나타낸다. self-influence / self-sensitivity라고도 한다.
- 레버리지는 선형적으로 영향도 행렬의 대각성분  $h_{ii}$ 으로 정의 된다. 즉, 레버리지는 실제의 회귀값  
 $y_i$ 이 예측값  $\hat{y}_i$ 에 미치는 영향, 즉 예측값을 세 개 변수의 위치로 끌어당기는 정도는 나타내  
준다.  
$$\hat{y}_i = h_{i1}y_1 + h_{i2}y_2 + \dots + h_{ii}y_i + \dots + h_{iN}y_N$$
- 만약  $h_{ii}$  값이 (이 정도) 나이가 많을수록 0이 될 수만 있다면 모든 점 데이터에 대해  
실제 회귀값과 예측값이 일치하게 된다.

$$h_{ii} = 1, h_{ij} = 0 \text{ (for } i \neq j) \Rightarrow \hat{y}_i = y_i$$

- 모형에서 결정하고 있는 데이터와 동등한 값을 가지는 데이터. 즉, 잔차가 큰 데이터는  
아웃라이어라고 한다. 잔차의 크기는 독립변수의 영향 범위를 아웃라이머는 벗어나  
이 영향을 미치는 표준화된 잔차의 제곱에 따라 다르다.

\* 잔차의 크기가 크면 아웃라이머가 된다.

- 여기서, 잔차의 크기와 레버리지 크기가 모두 큰 경우, Cook's distance를 이용해  
판단할 수 있다.