

회귀분석팀

6팀

권남택
윤주희
진효주
한유진
황유나

INDEX

1. 회귀분석의 가정
2. 잔차 플랏
3. 선형성 진단과 처방
4. 등분산성 진단과 처방
5. 정규성 진단과 처방
6. 독립성 진단과 처방
7. 다중공선성

모델의 선형성

오차의 등분산성

오차의 정규성

오차의 독립성

모델의 선형성

오차의 등분산성

오차의 정규성

오차의 독립성

모델의 선형성

오차의 등분산성

오차의 정규성

오차의 독립성

- 가정이 깨질 경우 문제

개별 회귀계수의 분산 상승

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon, \quad \varepsilon \sim NID(0, \sigma^2)$$

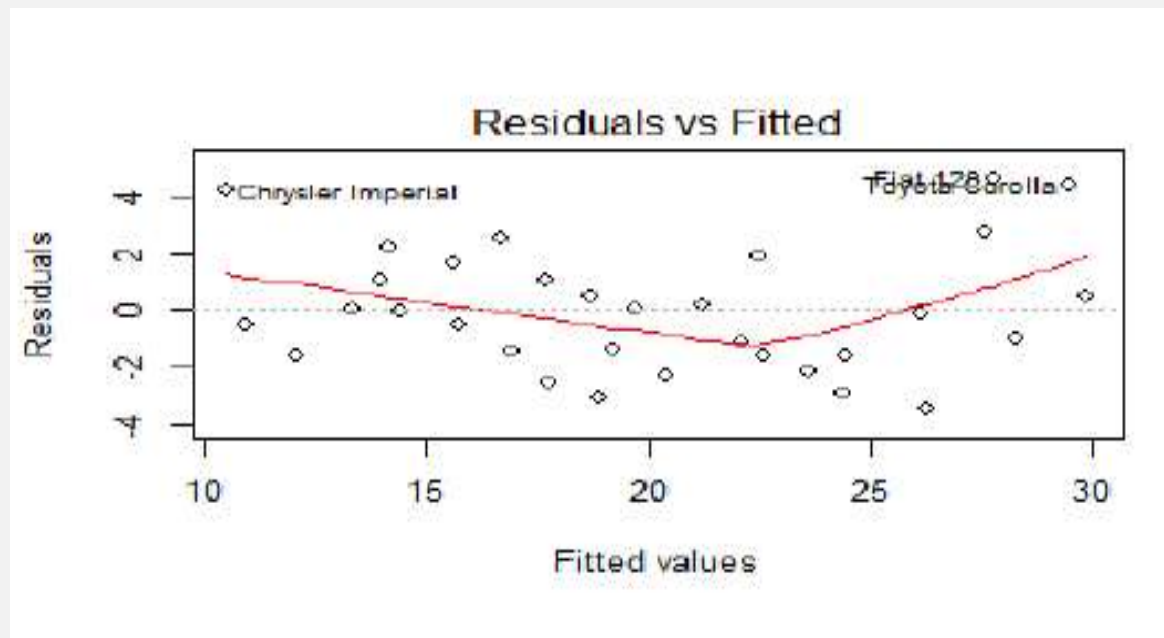
최소제곱추정량은 과소추정된 등분산을 가지고
t-value와 F-value를 산출

충분히 유의할 수 있는 귀무가설 기각

회귀계수 검정에 대한 신뢰성 저하

Residuals vs Fitted Normal QQ Plot Scale-Location Residuals vs Leverage

- 1번 : Residuals Vs Fitted



빨간선은 잔차들의 추세선

모델의 선형성, 등분산성을 확인할 수 있다.

잔차 플랏

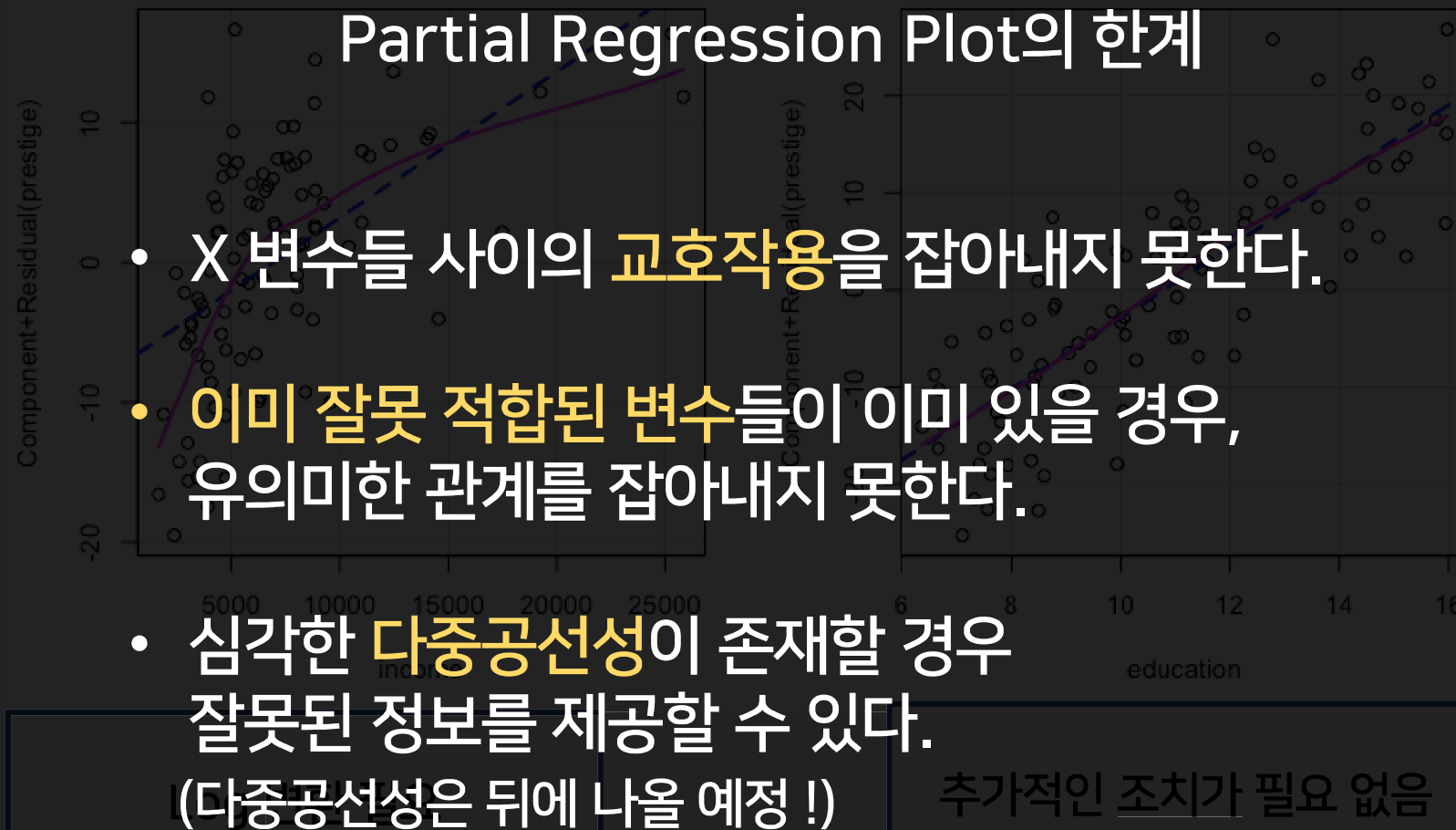
crPlots

변수 변환

Polynomial Reg.

Component + Residual Plots

Partial Regression Plot의 한계

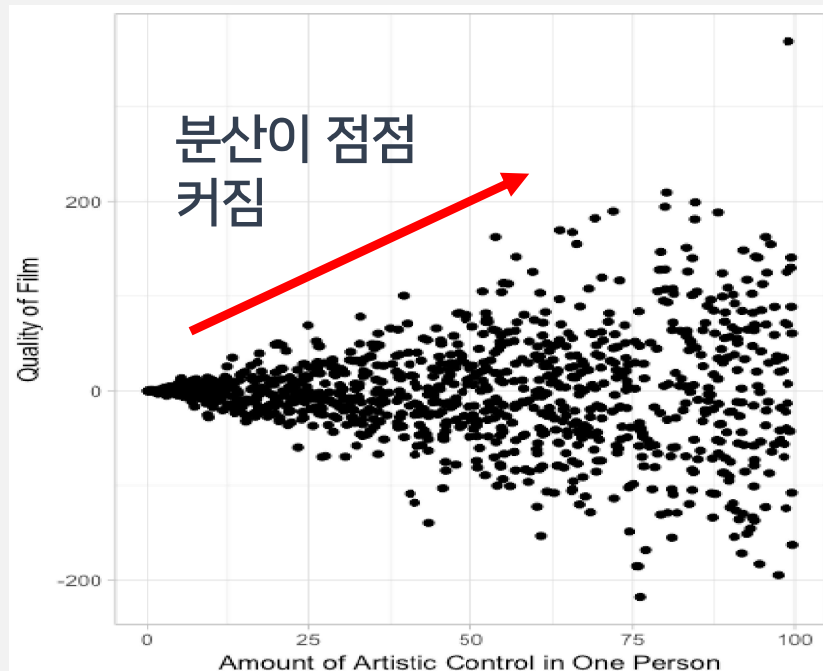


잔차 플랏

BP Test

WLS

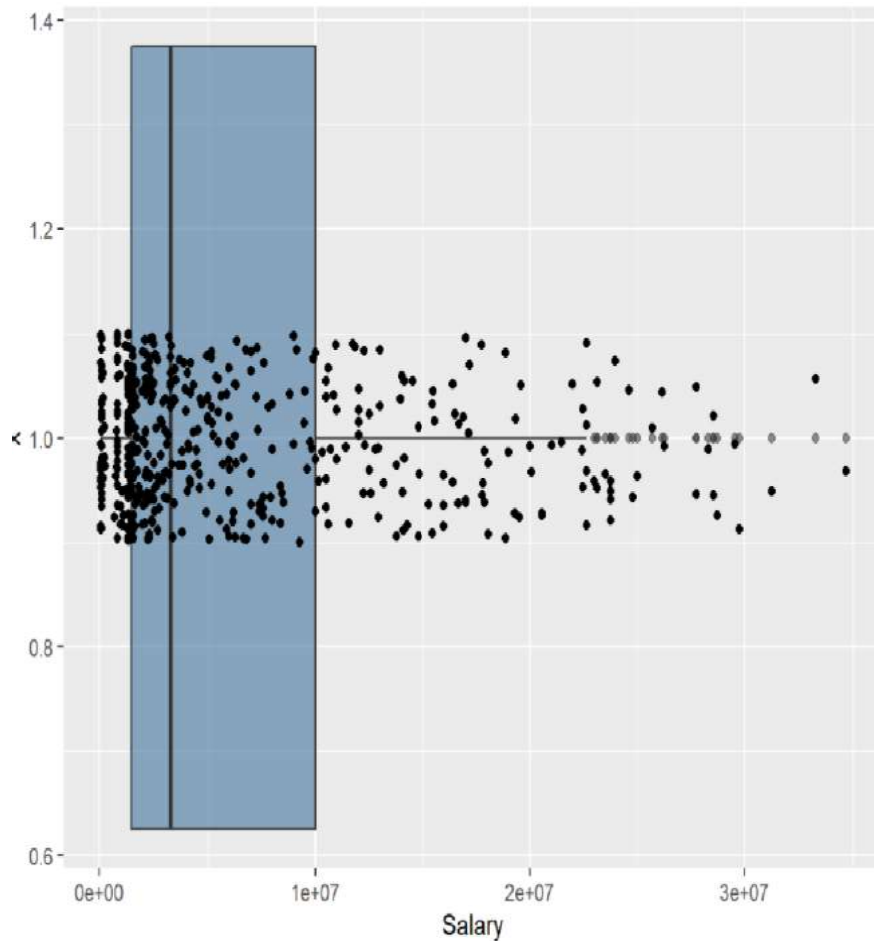
가중치는 사전 지식을 통해 정하거나 잔차 플랏을 보고 경험적으로 선정



분산이 점점 커지는 경우,
가중치를 $w_i \propto \frac{1}{x_i^2}$ 와 같은 방식으로
설정

WLS를 통해 구한 추정량은 다시 BLUE가 된다!

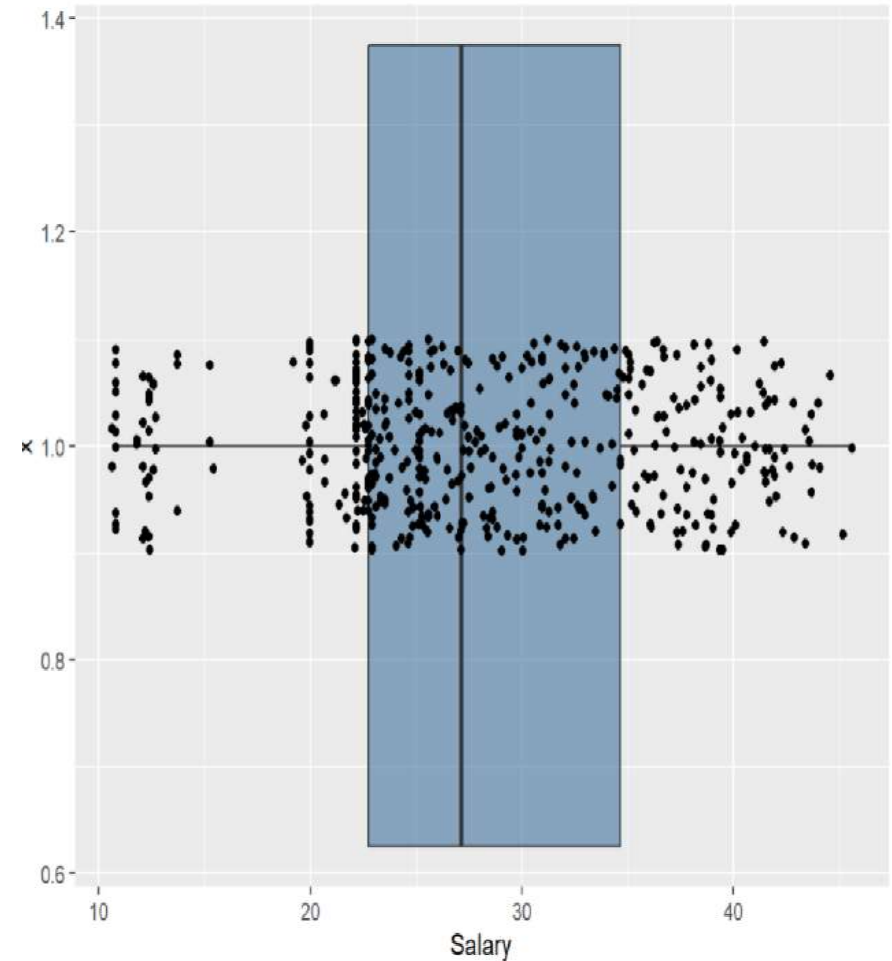
진단(QQ plot)



진단(test)



Box-Cox



y의 분포 달라짐, test결과 오차의 정규성 문제, 등분산성 문제 해결!

진단(test)

처방

가변수 만들기

뚜렷한 계절성이 있다고 판단되면,
이를 위해 가변수를 만든다.

시계열 분석으로
넘어가기!

다른 보정 방법도 가능하지만,
시계열 모델링을 하는 것이 적절하다.

다중공선성

다중공선성의 문제

다중공선성의 판별

다중공선성의 해결

- 다중공선성이란?

: 예측변수 X들간의 선형관계가 있는 경우

$Det(X) \neq 0, but Det(X) \approx 0$ 인 상황

$$\hat{\beta} = (X^t X)^{-1} X^t y, \quad Var(\hat{\beta}) = \sigma^2 (X^t X)^{-1},$$
$$(X^t X)^{-1} = \frac{1}{Det(X^t X)} adj(X^t X)$$

➡ $Det(X) \approx 0$ 임에 따라 분산이 커지고, 그에 따라 회귀계수 추정이 매우 불안정해짐

다중공선성

다중공선성의 문제

다중공선성의 판별

다중공선성의 해결

변수선택
(Variable
Selection)



정규화
(Regularization)

차원축소
(Dimension
Reduction)



Regularization + convex optimization

