



SO5032 Lecture 6

Brendan Halpin

March 6, 2023

SO5032 Lecture 6

SO5032 Lecture 6

Residuals

Residuals

$$Y = b_0 + b_1 X_1 + \dots + b_k X_k + e$$

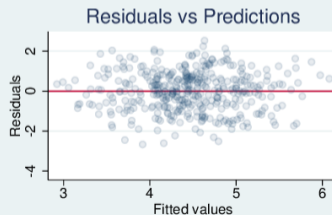
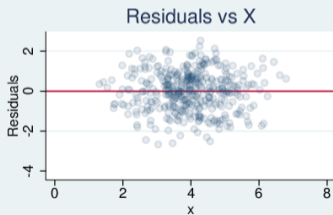
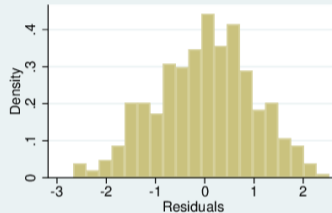
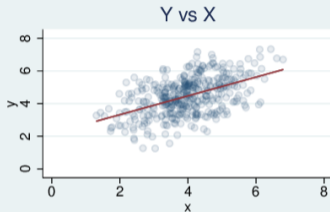
$$e \sim N(0, \sigma)$$

Characteristics

- Residuals will
 - have mean 0
 - be as small as possible
 - have no linear relationship to X variables
- Residuals should
 - be approximately normally distributed (symmetric is often enough)
 - not have a non-linear relationship to any X variable
 - have a constant spread, that is not related to X or Y values
- If correlated with variables not in the model, perhaps those variables should be included

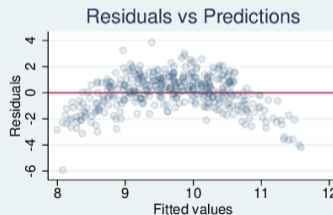
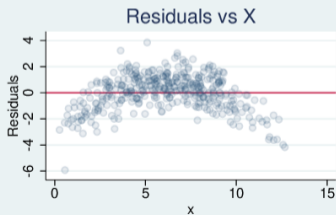
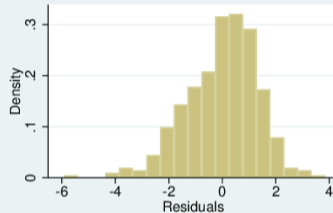
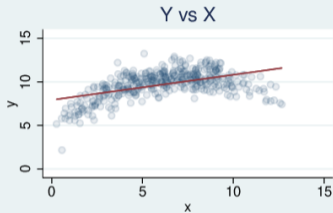
Examining residuals: ideal

Simple residuals



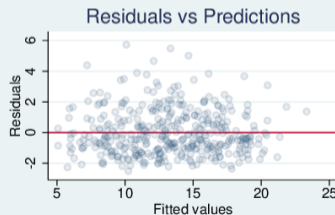
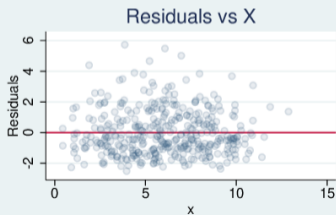
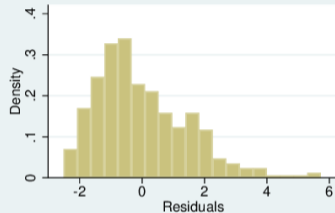
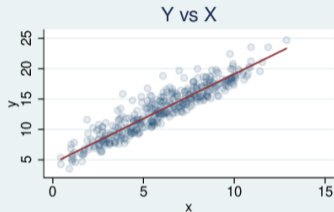
Examining residuals: Non-linear

Non-linear relationship



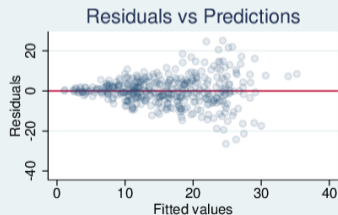
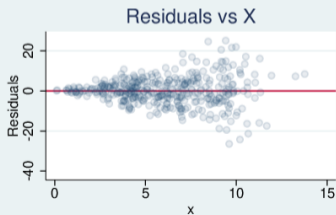
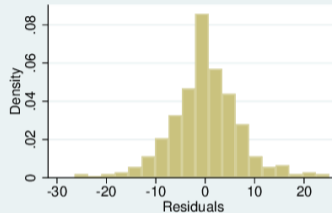
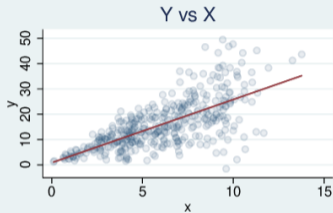
Examining residuals: asymmetric

Asymmetry of residuals



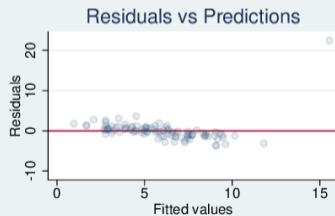
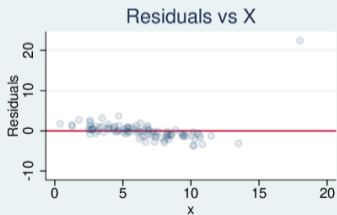
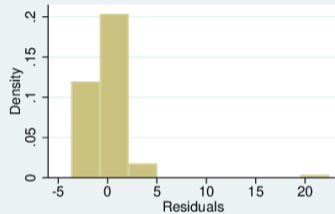
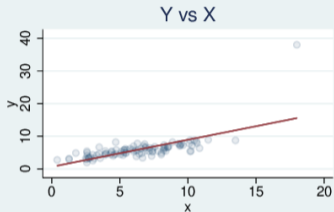
Examining residuals: heteroscedasticity

Heteroscedasticity: correlation between X and sigma

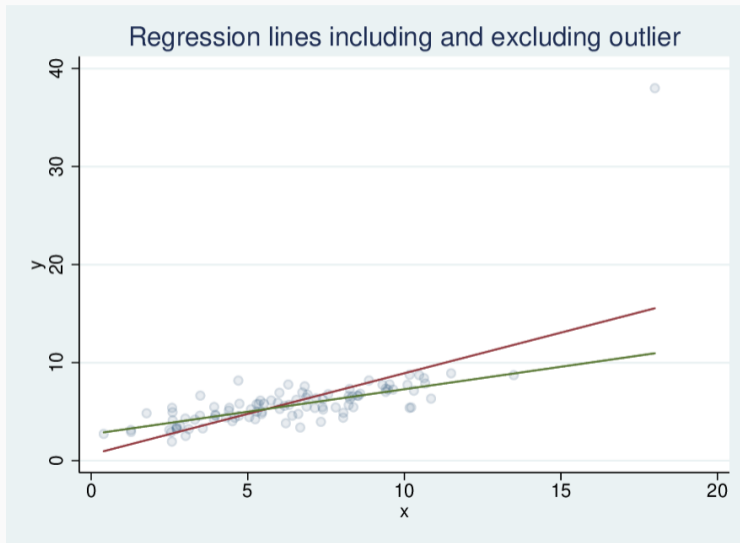


Examining residuals: Spotting outliers

Outliers



Examining residuals: Influence of outliers



SO5032 Lecture 6

Influence

Outliers may have undue influence

- $dfbeta$
- Cook's distance

Outlier interactive app

`http://teaching.sociology.ul.ie:3838/influence/`