SO5032 Lecture 6

Brendan Halpin

March 3, 2024

Outline

SO5032 Lecture 6

SO5032 Lecture 6

Residuals

Residuals

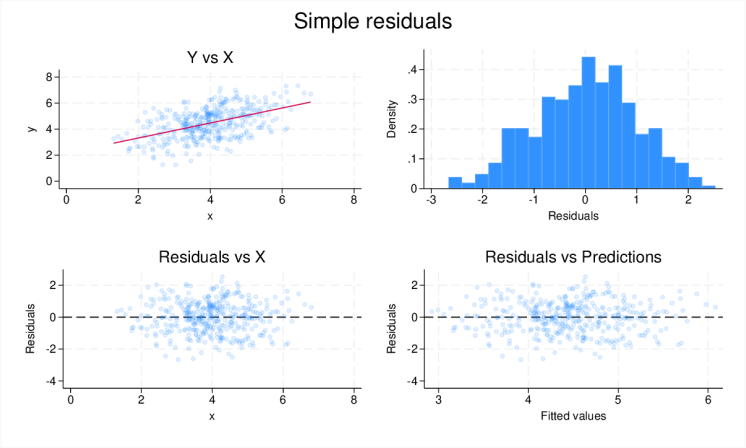
$$Y = b_0 + b_1 X_1 + ... + b_k X_k + e$$

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

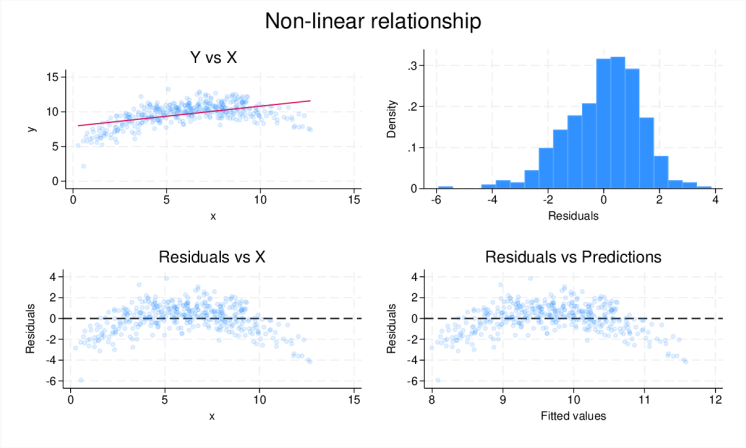
Characteristics

- 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

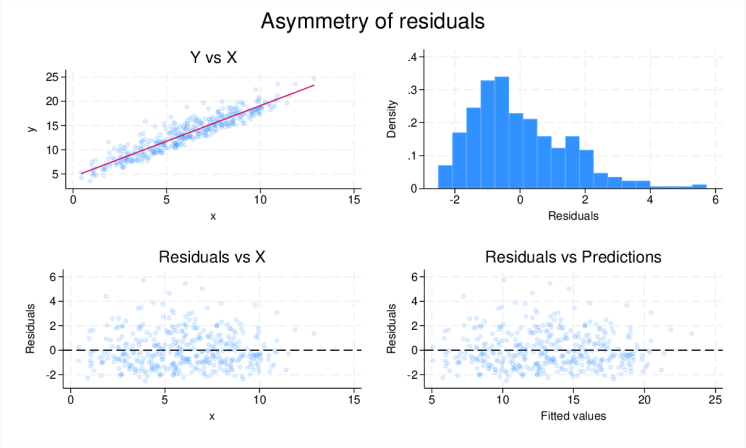
Examining residuals: ideal



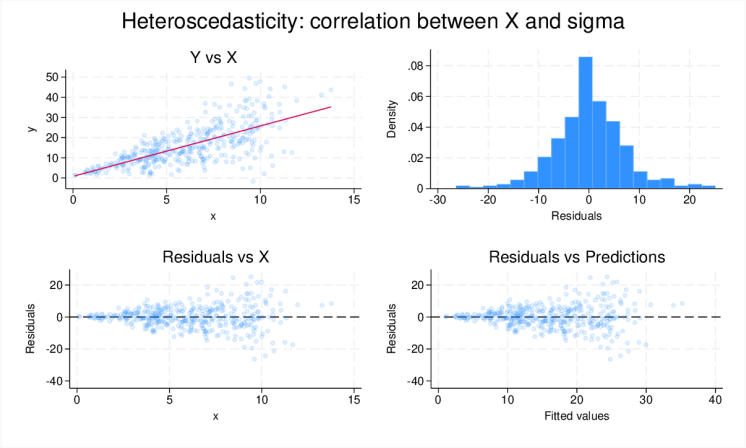
Examining residuals: Non-linear



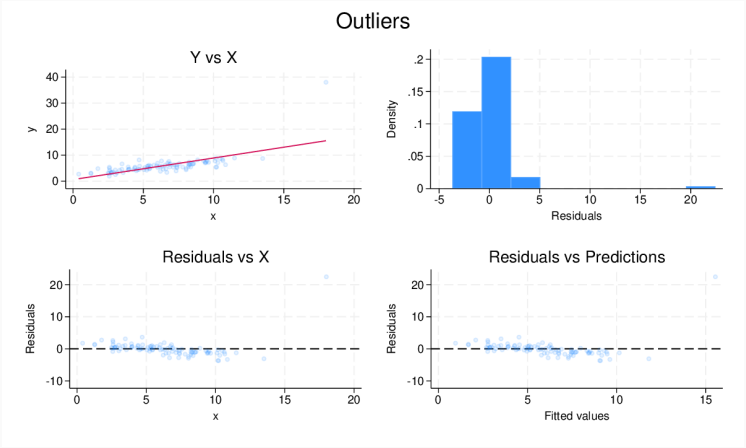
Examining residuals: asymmetric



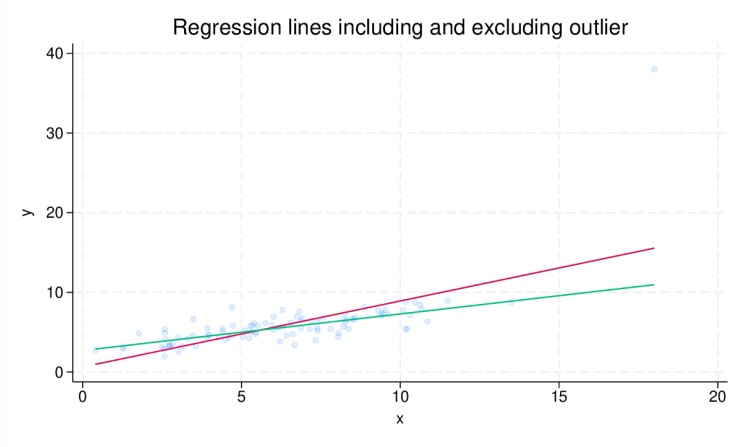
Examining residuals: heteroscedasticity



Examining residuals: Spotting outliers



Examining residuals: Influence of outliers



SO5032 Lecture 6

Influence

Outliers may have undue influence

- `dfbeta`
- Cook's distance

DFBETA

- For each variable in the regression, for each case
- The effect of dropping that case on that variable
- Scaled by the standard error:

$$\frac{b - b^*}{SE}$$

Cook's Distance

- A single number summarising each case's overall influence
- A scaled sum of changes in predicted Y

Outlier interactive app

https://teaching.sociology.ul.ie/apps/influence/