# Simple Linear Regression

**Conditions** 

Prof. Maria Tackett



## Click here for PDF of slides



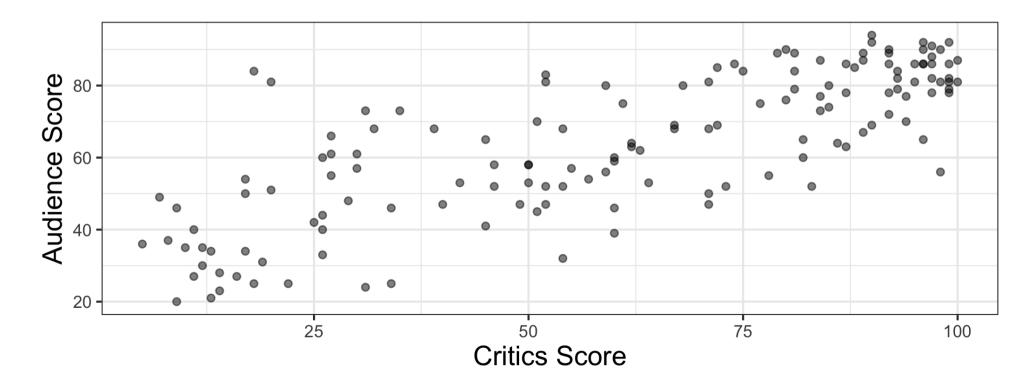
## **Topics**

- List the conditions for simple linear regression
- Use plots of the residuals to check the conditions



## Movie ratings data

The data set contains the "Tomatometer" score (**critics**) and audience score (**audience**) for 146 movies rated on rottentomatoes.com.





### The model

audience = 
$$32.316 + 0.519 \times \text{critics}$$

term	estimate	std.error	statistic	p.value
(Intercept)	32.316	2.343	13.795	0
critics	0.519	0.035	15.028	0



$$Y|X \sim N(\beta_0 + \beta_1 X, \sigma_{\epsilon}^2)$$



#### Model conditions

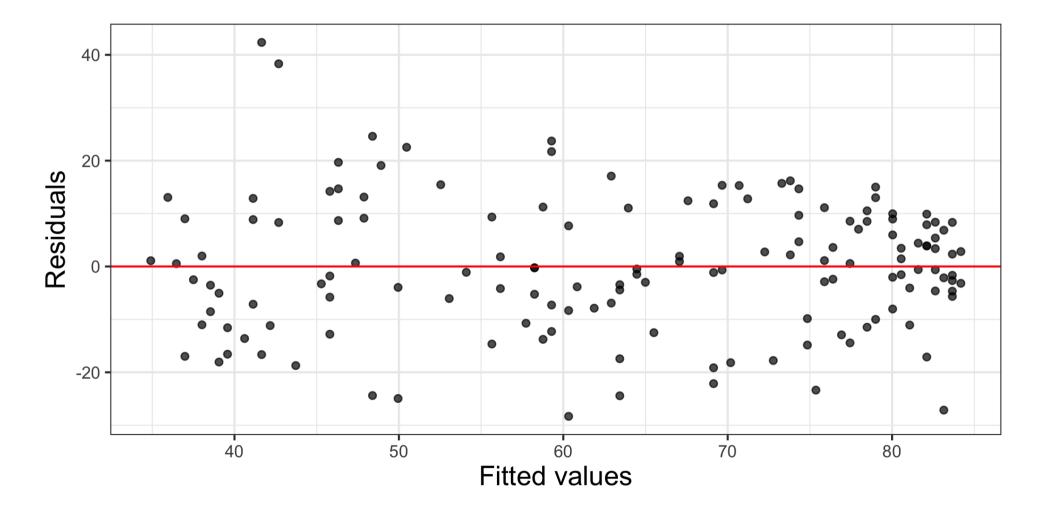
- 1. **Linearity:** There is a linear relationship between the response and predictor variable.
- 2. **Constant Variance:** The variability of the errors is equal for all values of the predictor variable.
- 3. Normality: The errors follow a normal distribution.
- 4. Independence: The errors are independent from each other.



$$residual_i = e_i = y_i - \hat{y}_i$$

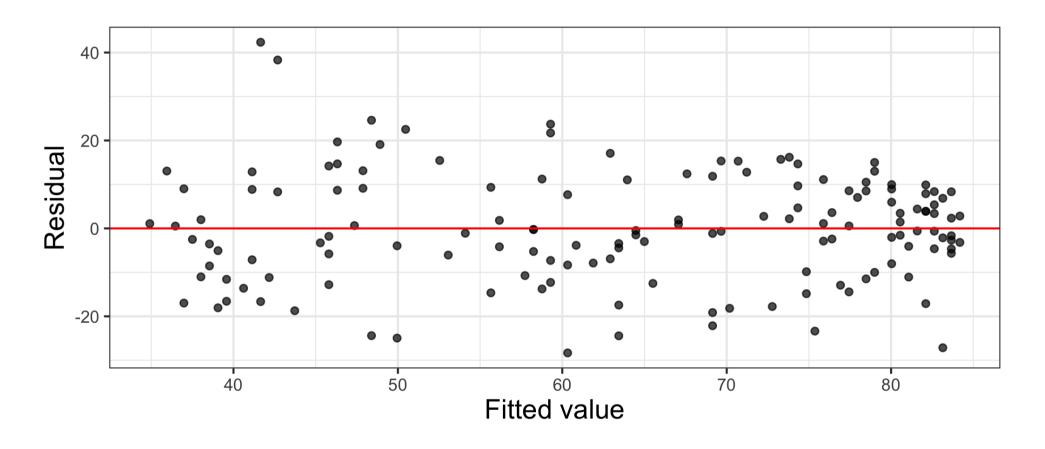


### Residuals vs. fitted values





## **Checking linearity**



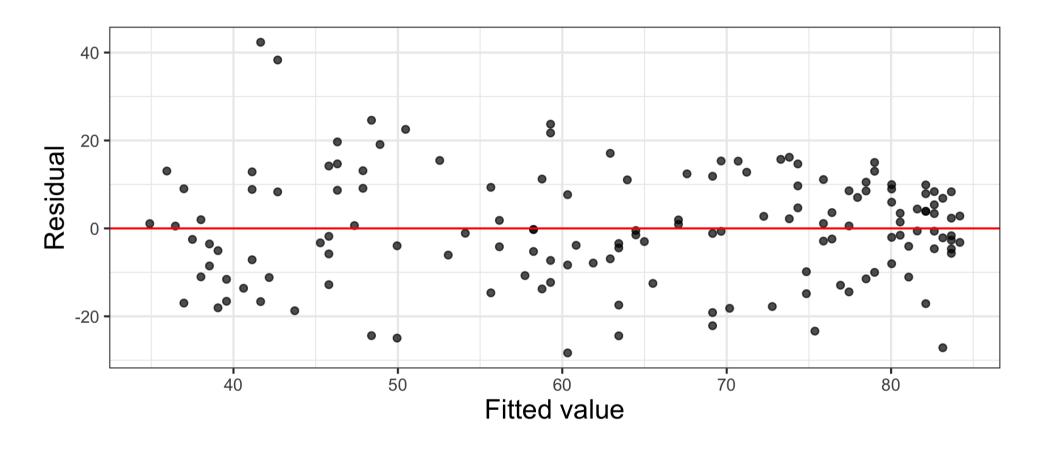


There is no distinguishable pattern or structure. The residuals are randomly scattered.

## X Violation: distinguishable pattern



## Checking constant variance





The vertical spread of the residuals is relatively constant across the plot.

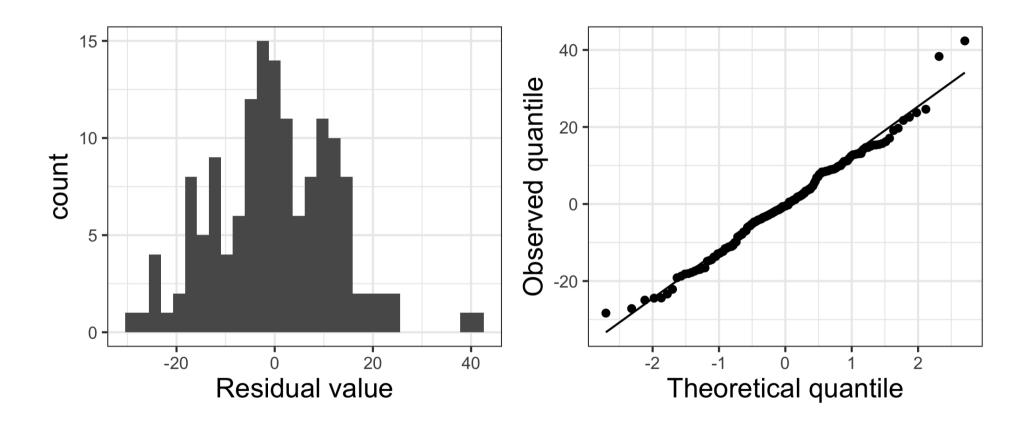
## X Violation: non-constant variance



## Normal quantile plot



## **Checking normality**







## Checking independence

- We can often check the independence assumption based on the context of the data and how the observations were collected.
- If the data were collected in a particular order, examine a scatterplot of the residuals versus order in which the data were collected.
- ☑ Based on available information, the error for one movie does not tell us anything about the error for another movie.



### In practice

As you check the model conditions, ask if any observed deviation from the model conditions are so great that

- a different model should be proposed.
- 2 conclusions drawn from the model should be used with caution.
- ✓ If not, the conditions are sufficiently met and we can proceed with the current model.



### Recap

- Used plots of the residuals to check conditions for simple linear regression:
  - Linearity
  - Constant Variance
  - Normality
  - Independence

