# Simple Linear Regression

## Model conditions


Dr. Maria Tackett

# Topics

# Topics

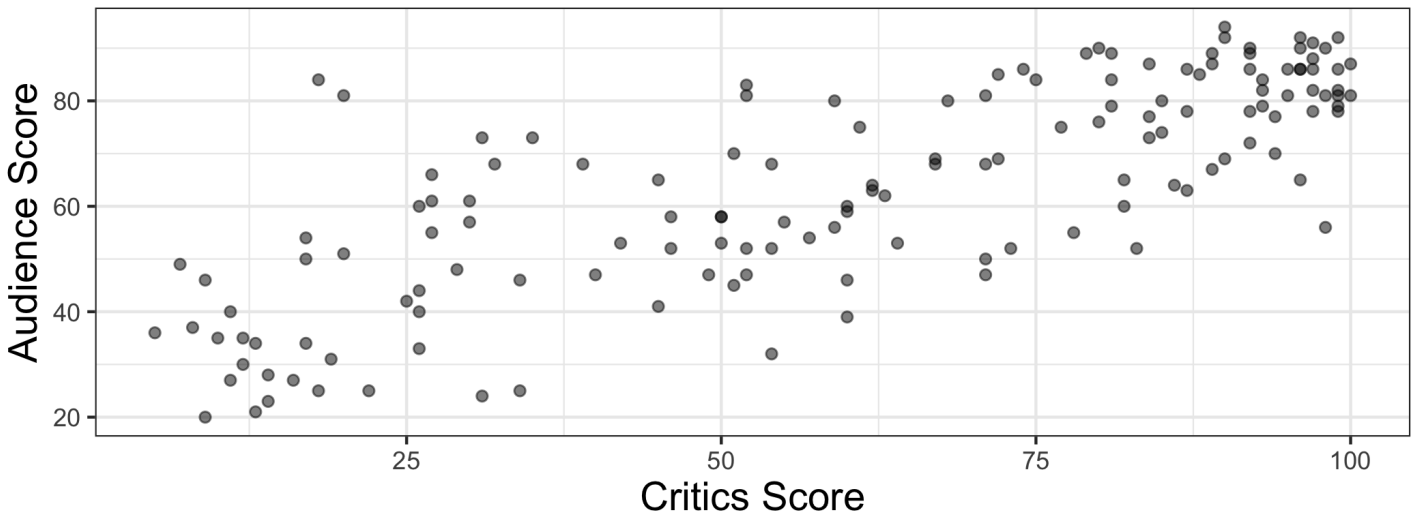
- List the conditions for simple linear regression

# Topics

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

# Movie ratings data

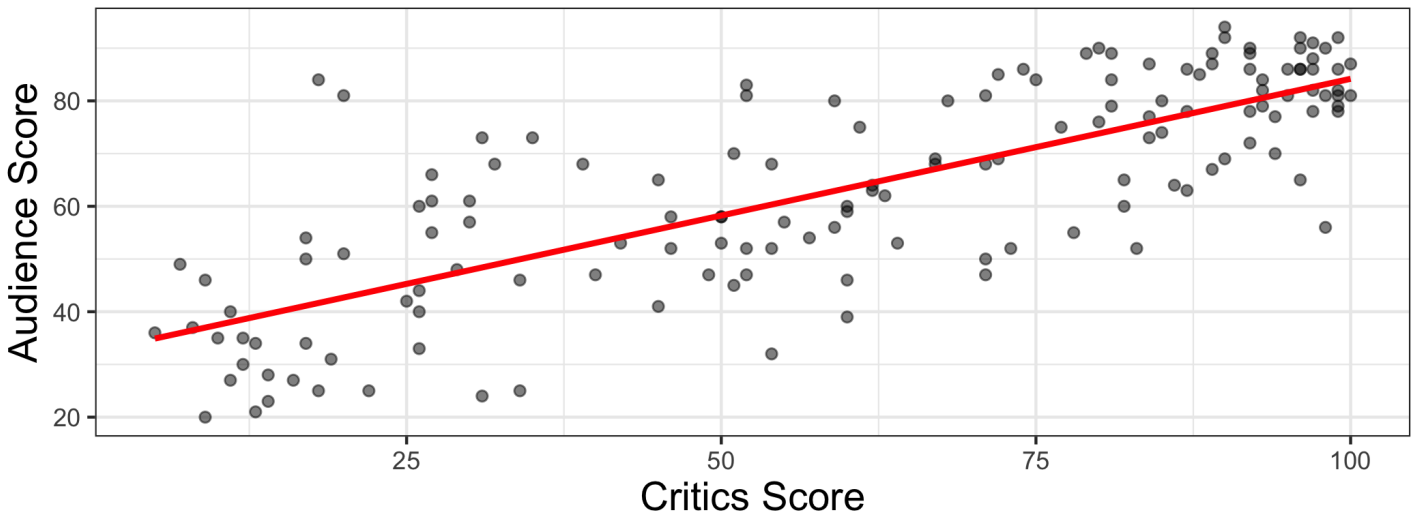
The data set contains the "Tomatometer" score (**`critics`**) and audience score (**`audience`**) for 146 movies rated on rottentomatoes.com.



# The model

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

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



# Model conditions

# Model conditions

1. **Linearity:** There is a linear relationship between the response and predictor variable.

# 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.

# 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.

# Model conditions

1. **Linearity:** There is a linear relationship between the response and predictor variable.

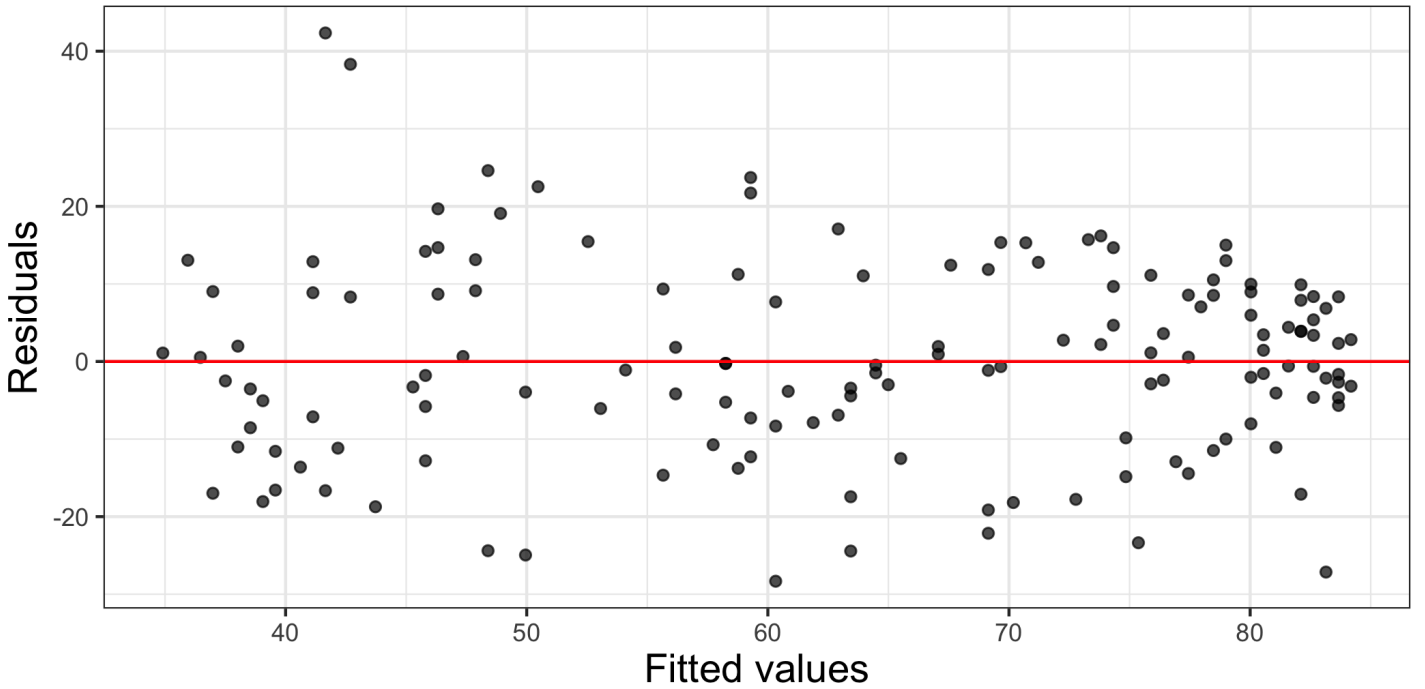
2. **Constant Variance:** The variability of the errors is equal for all values of the predictor variable.

3. **Normality:** The errors follow a normal distribution.

4. **Independence:** The errors are independent from each other.

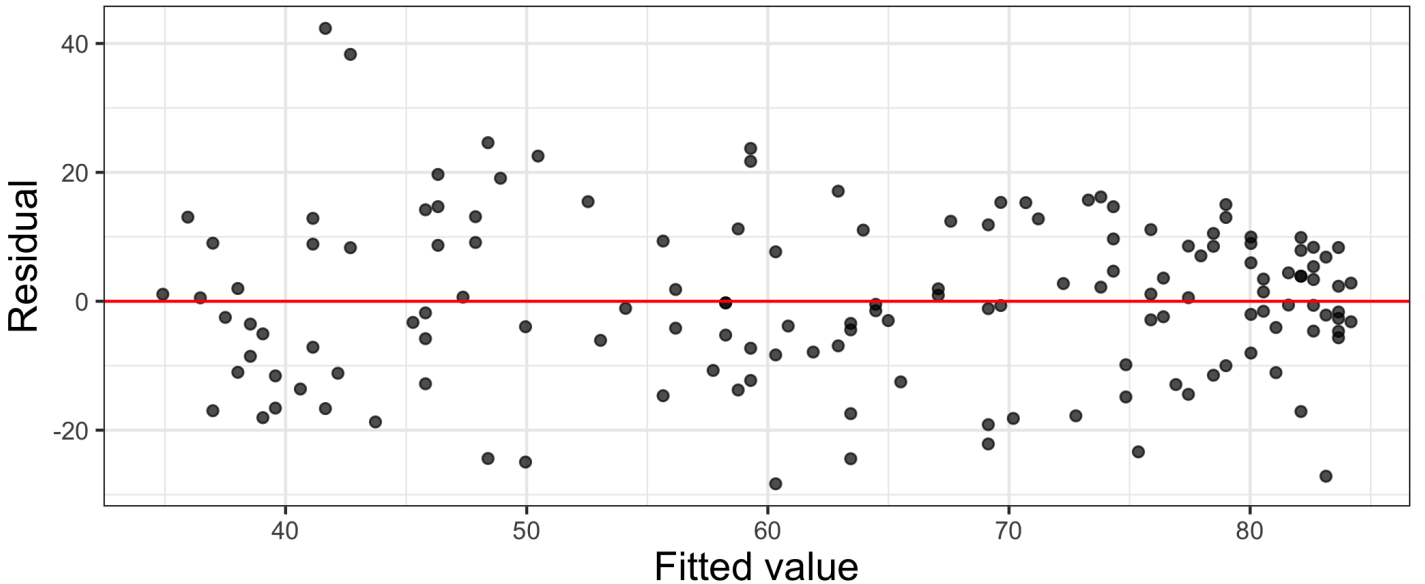
$$\text{residual} = y - \hat{y}$$

# Residuals vs. fitted values

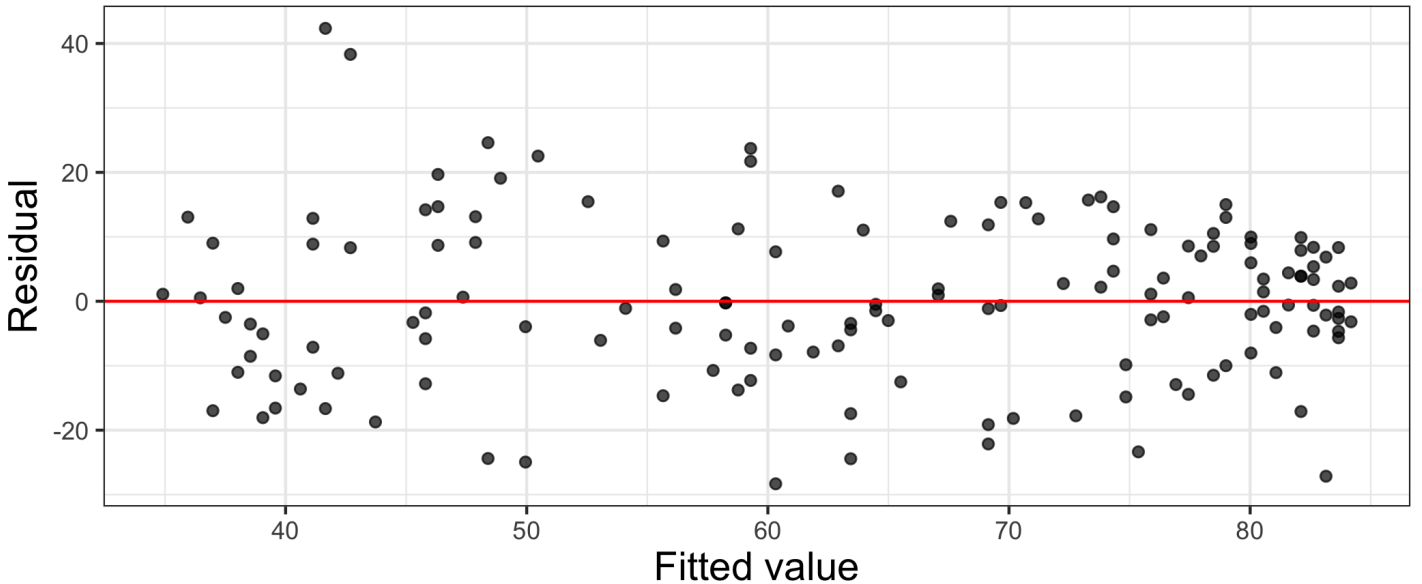


# Checking linearity

# Checking linearity



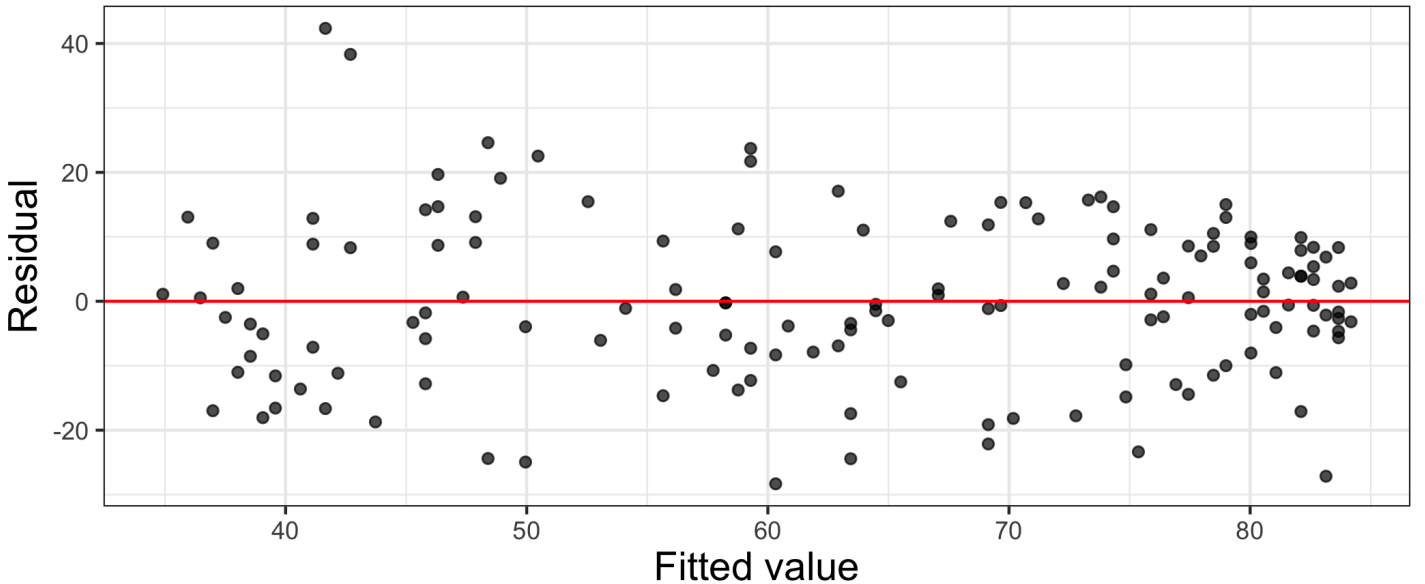
# Checking linearity



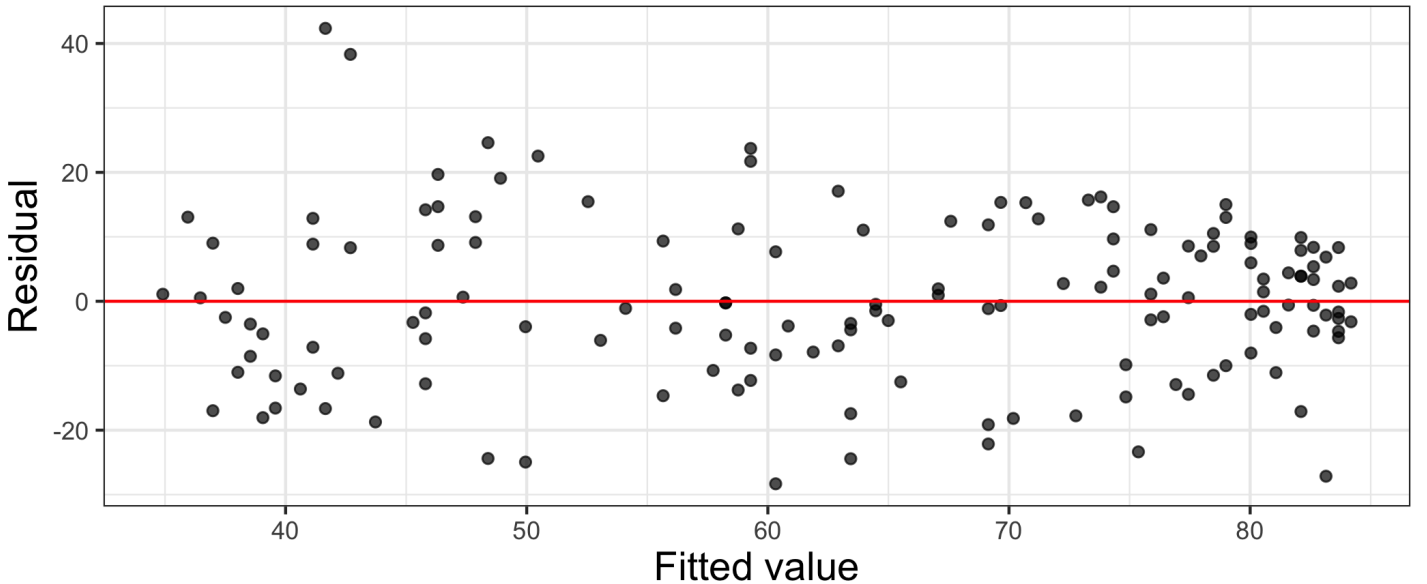

✅ There is no distinguishable pattern or structure. The residuals are randomly scattered.

# Checking constant variance

# Checking constant variance

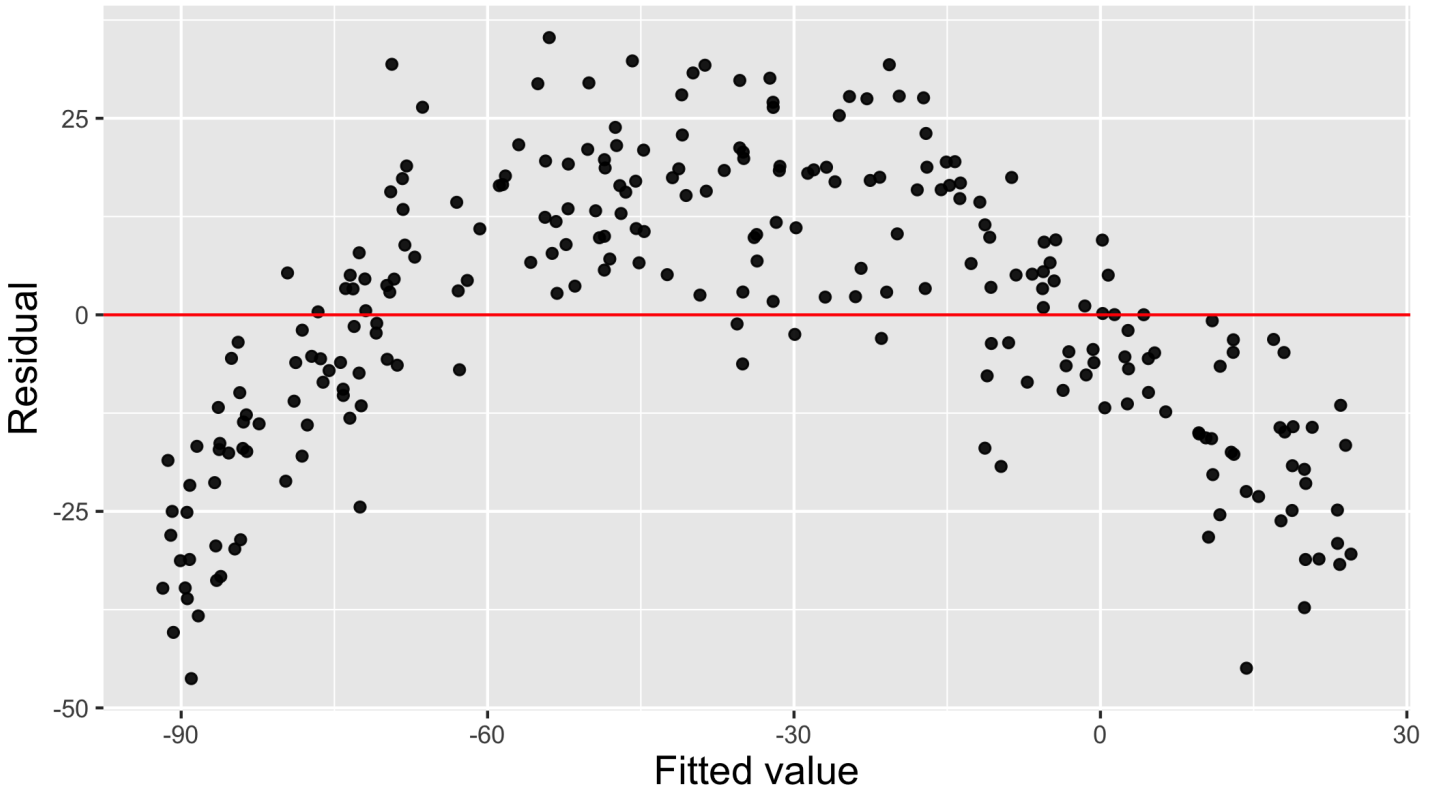


# Checking constant variance

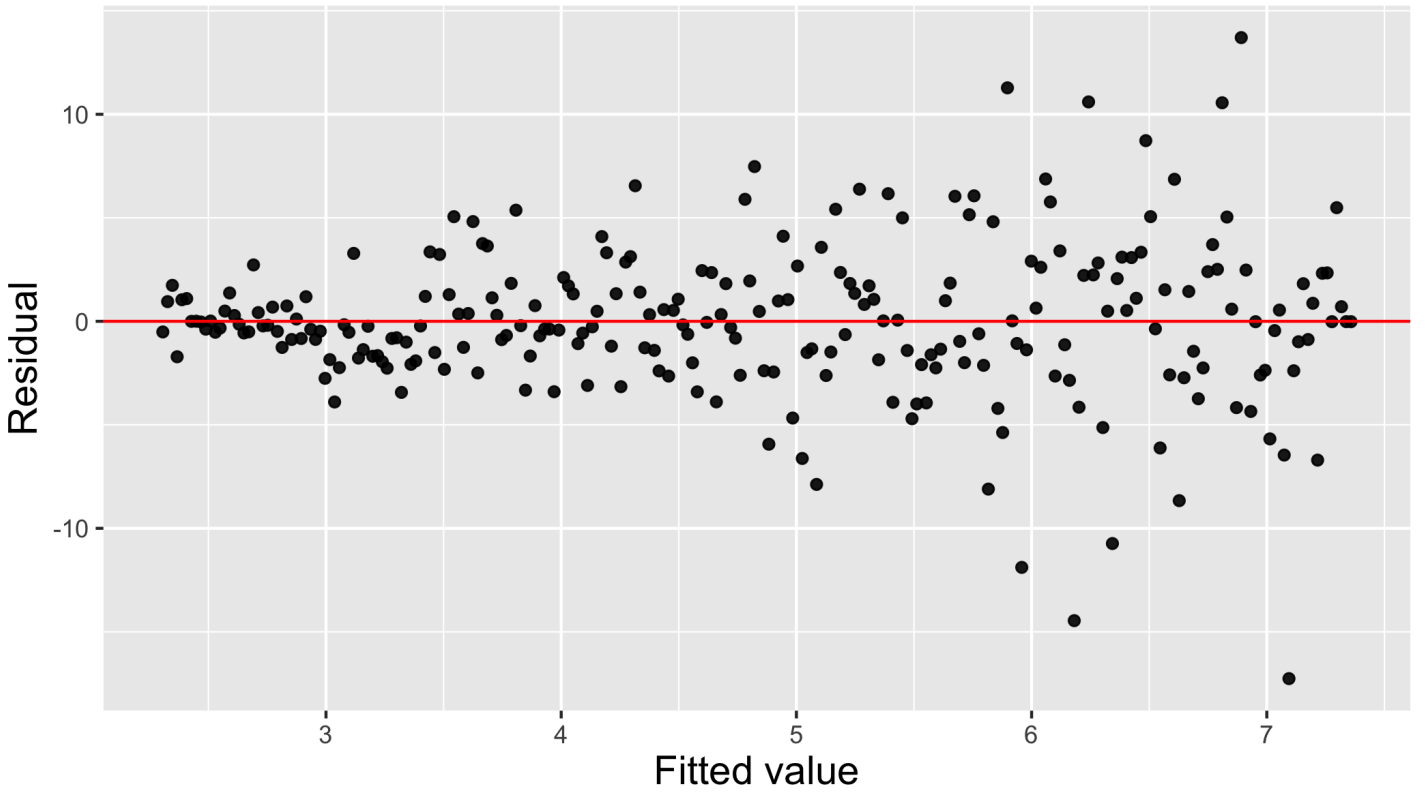



✅ The vertical spread of the residuals is relatively constant across the plot.

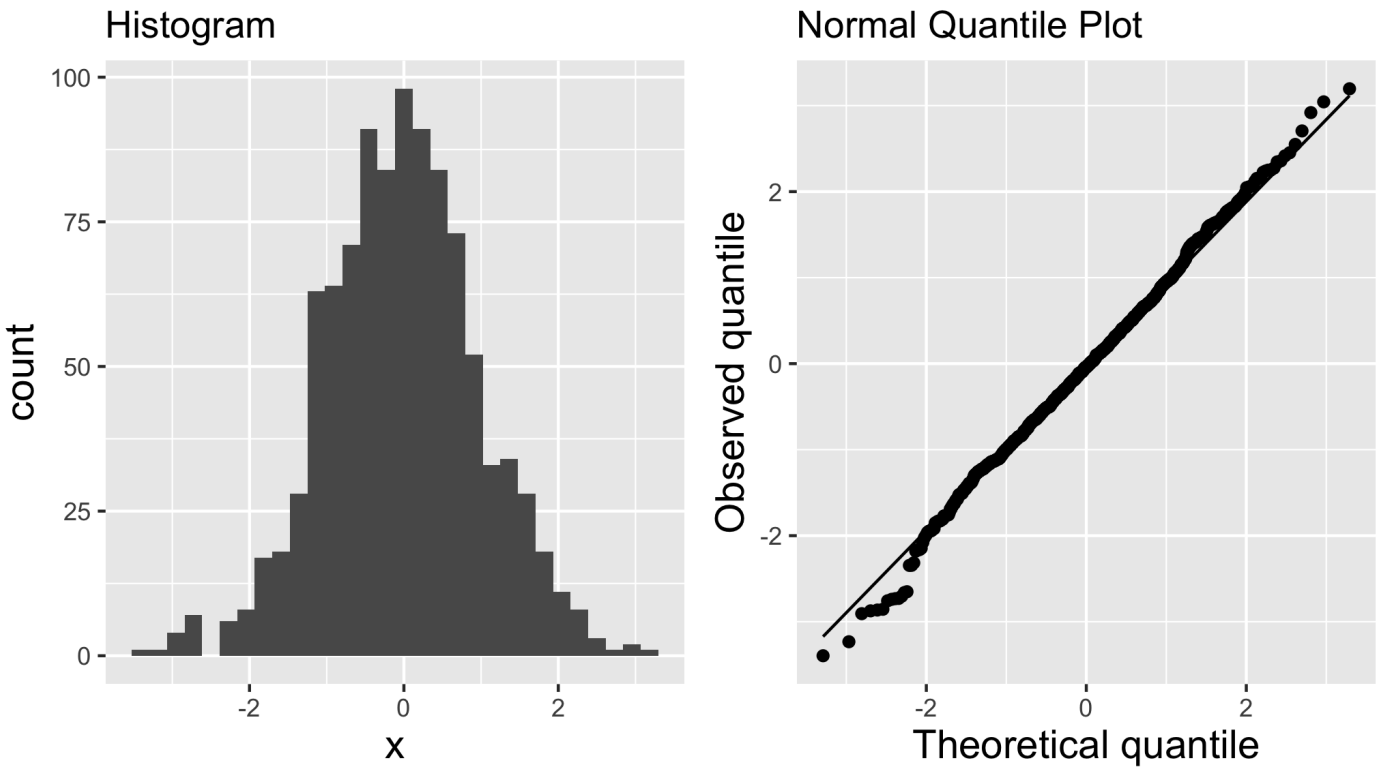
# Violation: distinguishable pattern



# Violation: non-constant variance

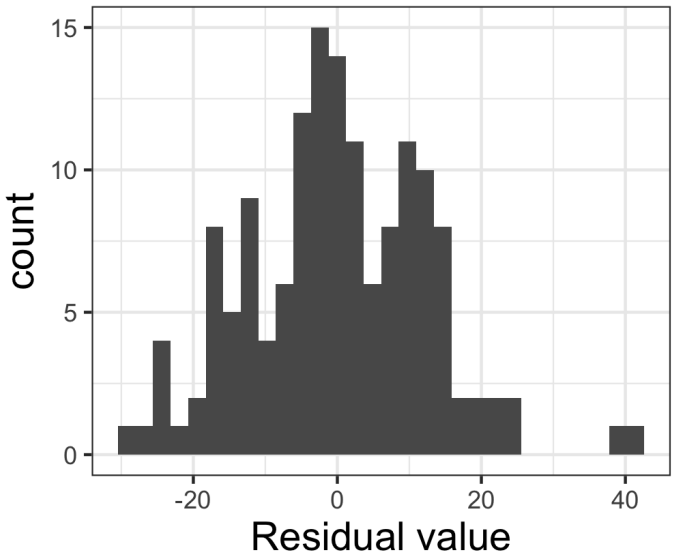


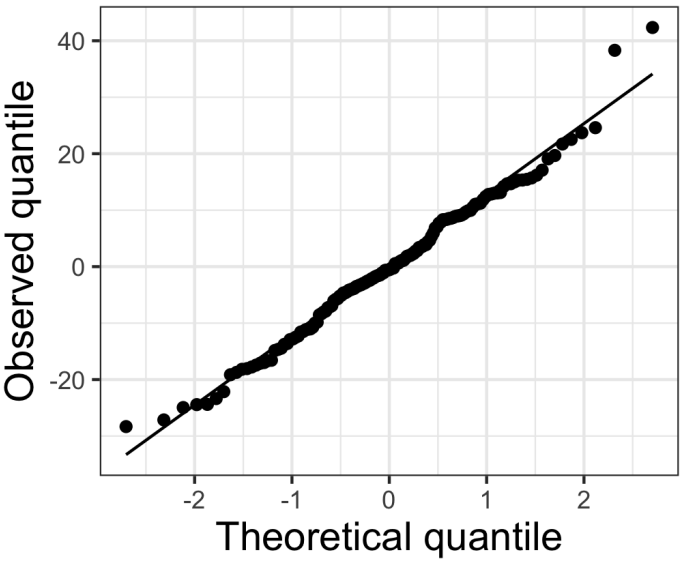
# Normal quantile plot



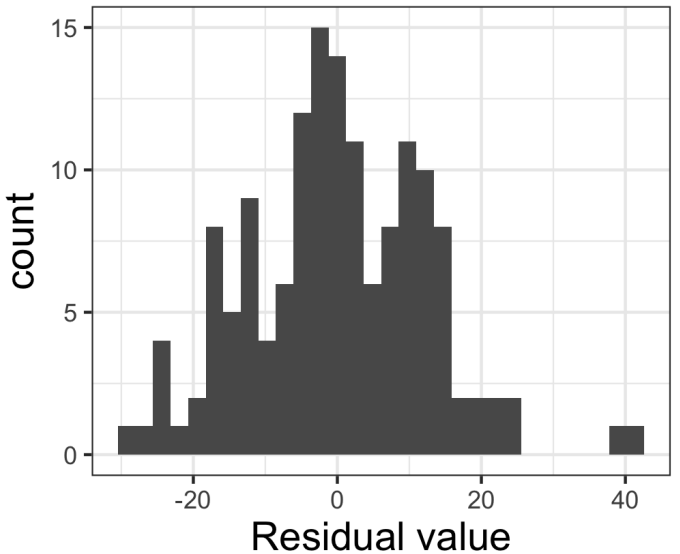
# Checking normality

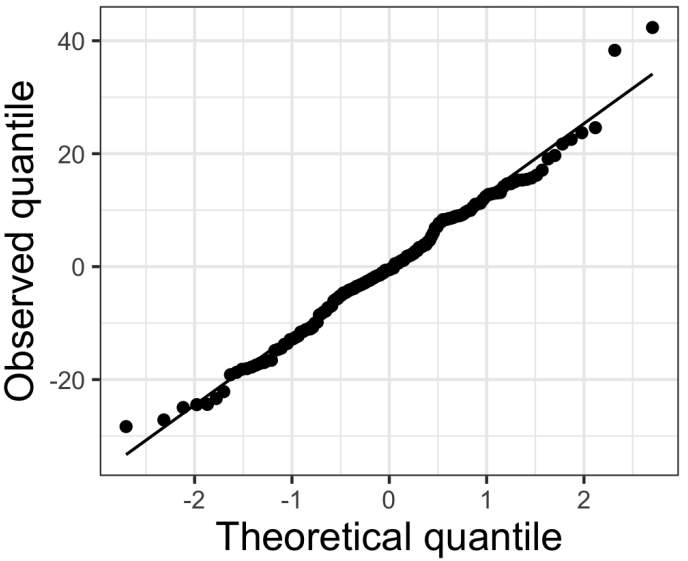
# Checking normality



# Checking normality




✅ Points follow a straight diagonal line on the normal quantile plot.

# Checking independence

- We can often check the independence assumption based on the context of the data and how the observations were collected.

# 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.

# In practice

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

1️⃣ a different model should be proposed.

1️⃣ conclusions drawn from the model should be used with caution.

# In practice

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

1️⃣ a different model should be proposed.

1️⃣ 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

# Recap

- Listed the conditions for simple linear regression:
  - Linearity
  - Constant Variance
  - Normality
  - Independence

# Recap

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