Simple Linear Regression

Assessing Conditions

Dr. Maria Tackett

Topics

Topics

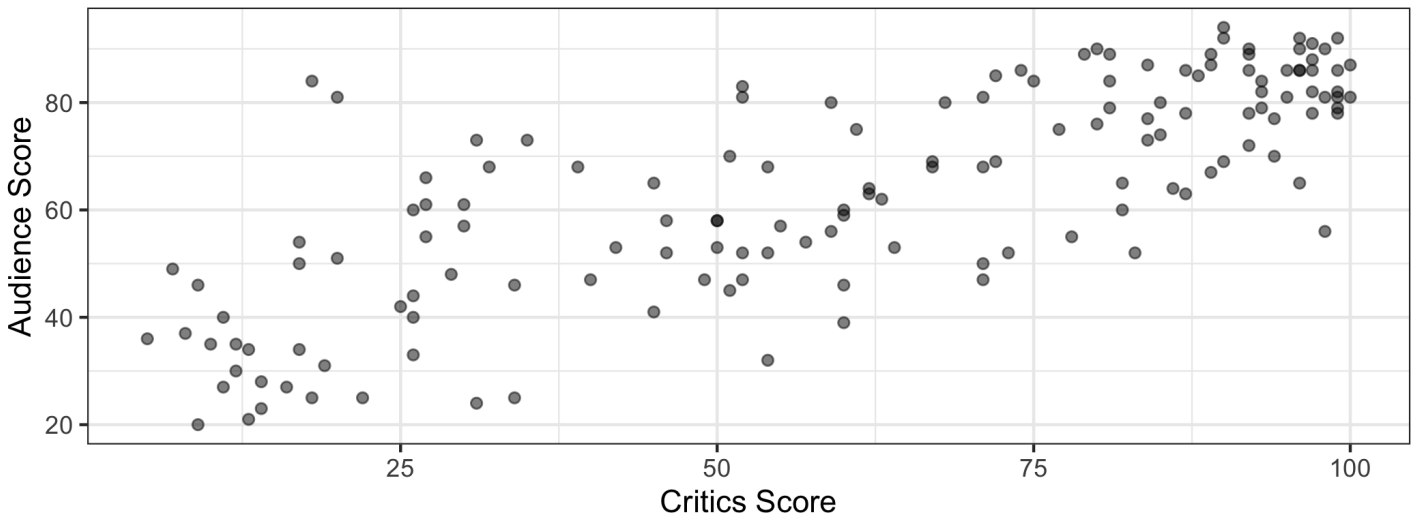
- Identify the conditions for simple linear regression

Topics

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

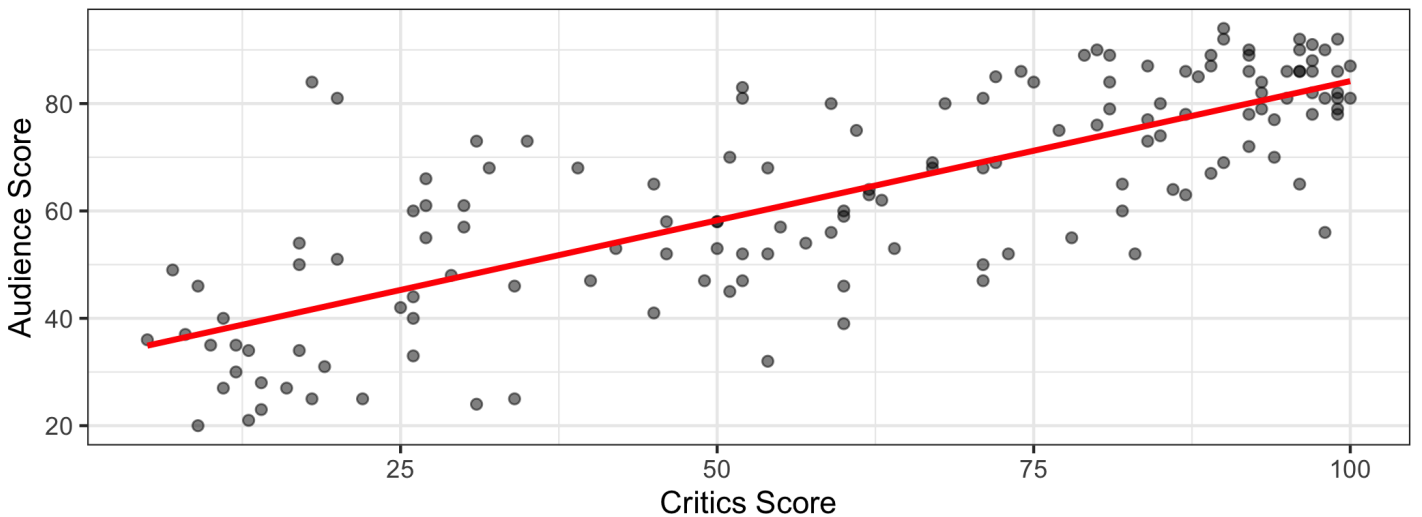
Movie ratings data

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



We fit a line to describe the relationship between the critics score and audience score.

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.

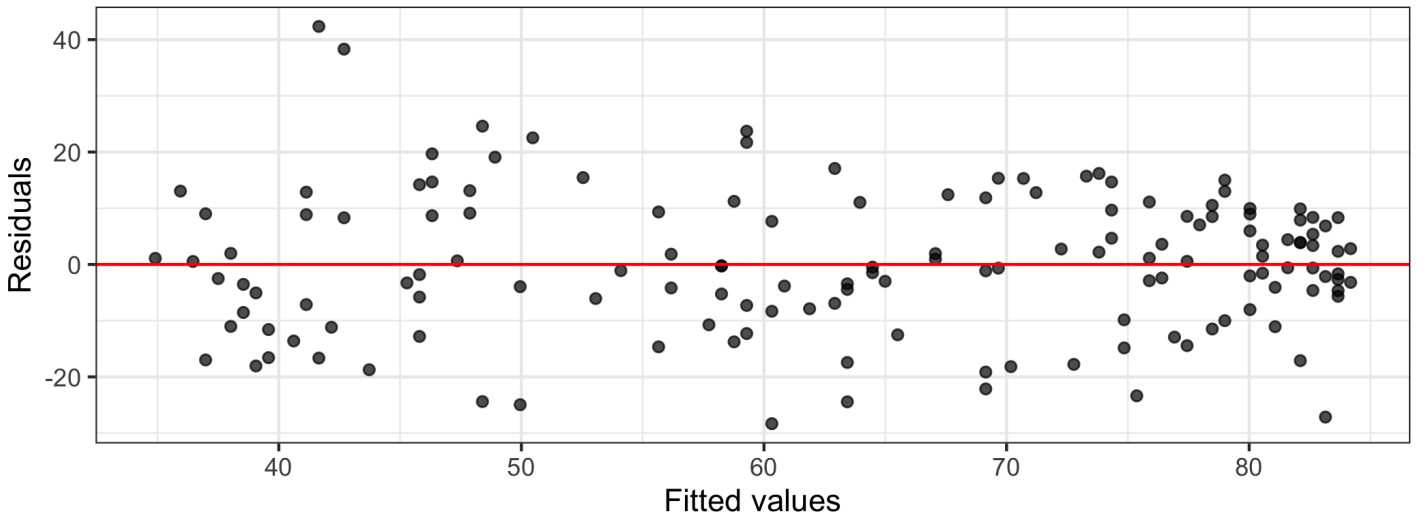
2. **Constant Variance:** The variability of the errors is equal for all values of the predictor variable.

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

4. **Independence:** The errors are independent from one another.

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

Plot of residuals vs. fitted values

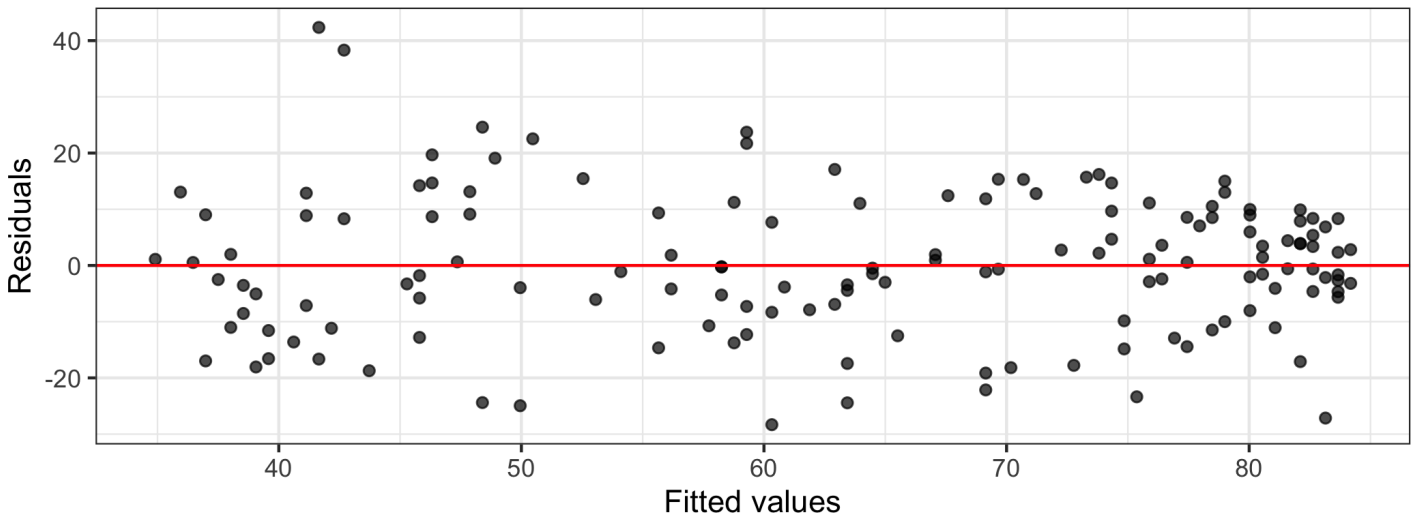


Assessing linearity

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

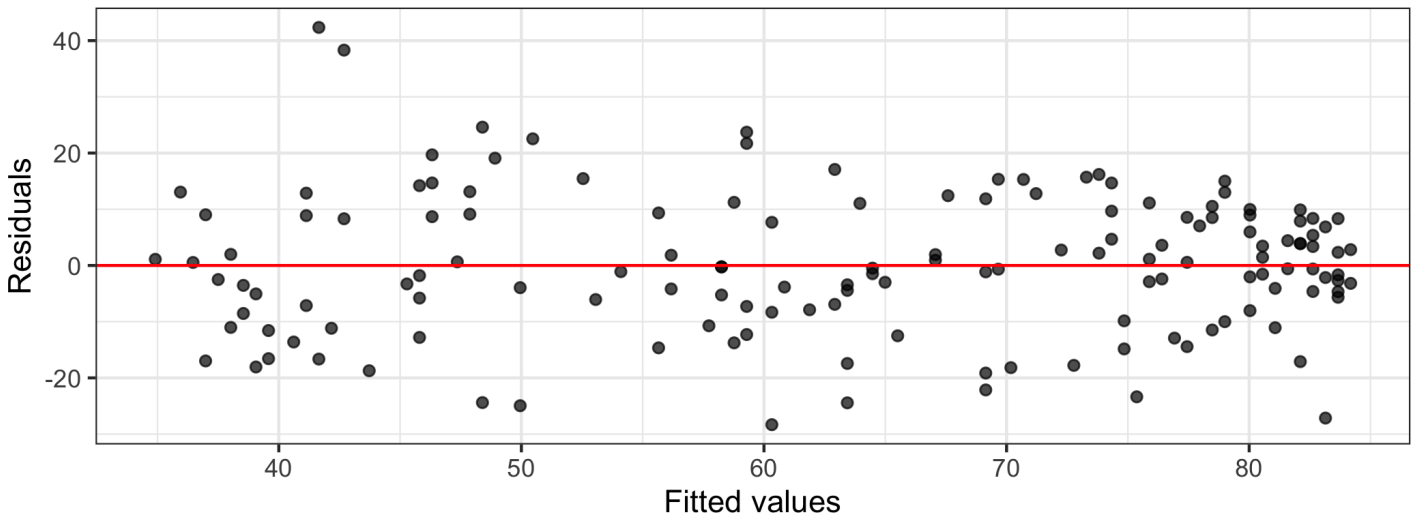
Assessing linearity

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



Assessing linearity

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



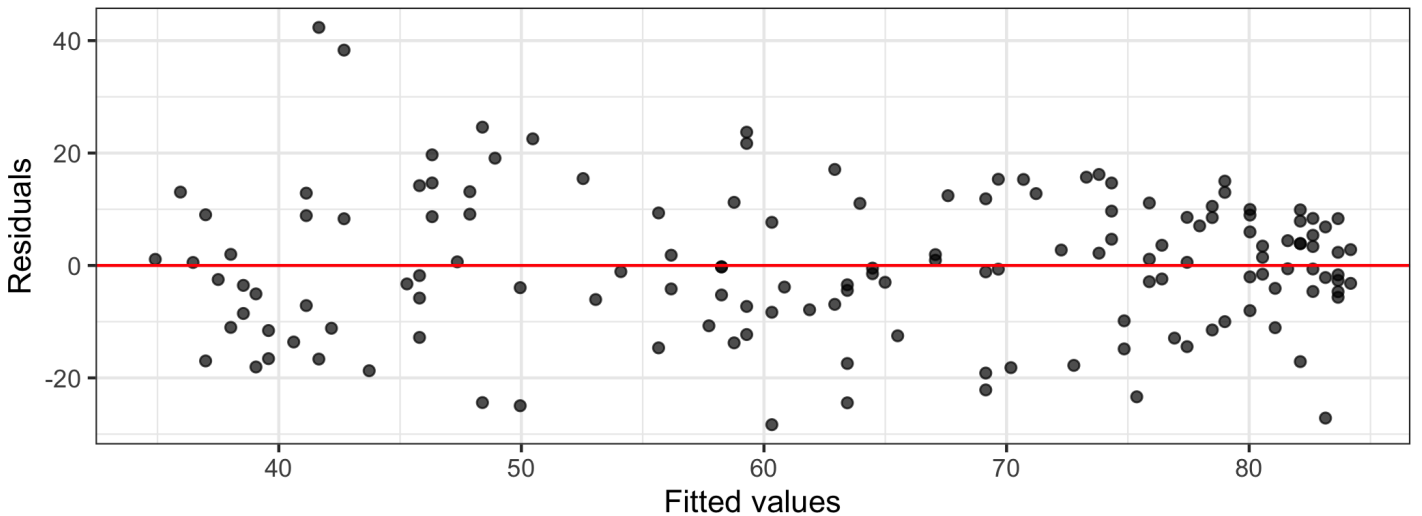
✅ There is no distinguishable pattern or structure. The residuals are randomly scattered.

Assessing constant variance

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

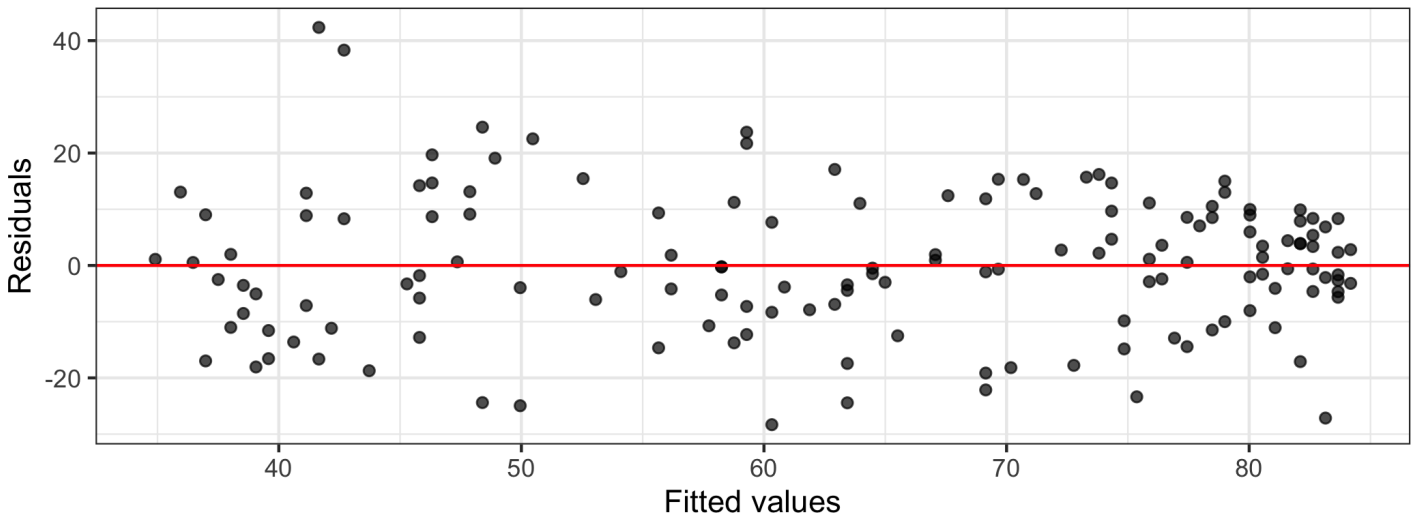
Assessing constant variance

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



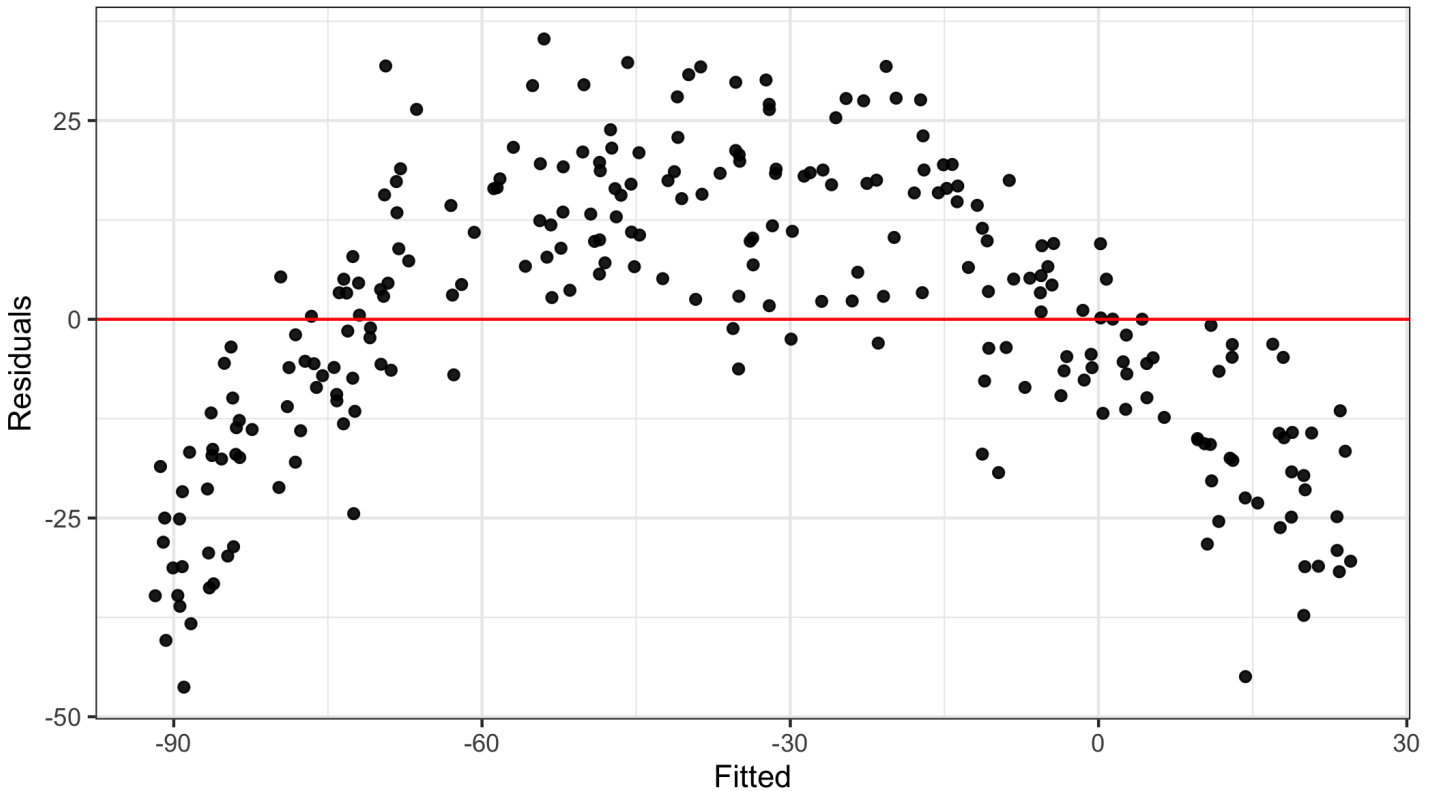
Assessing constant variance

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

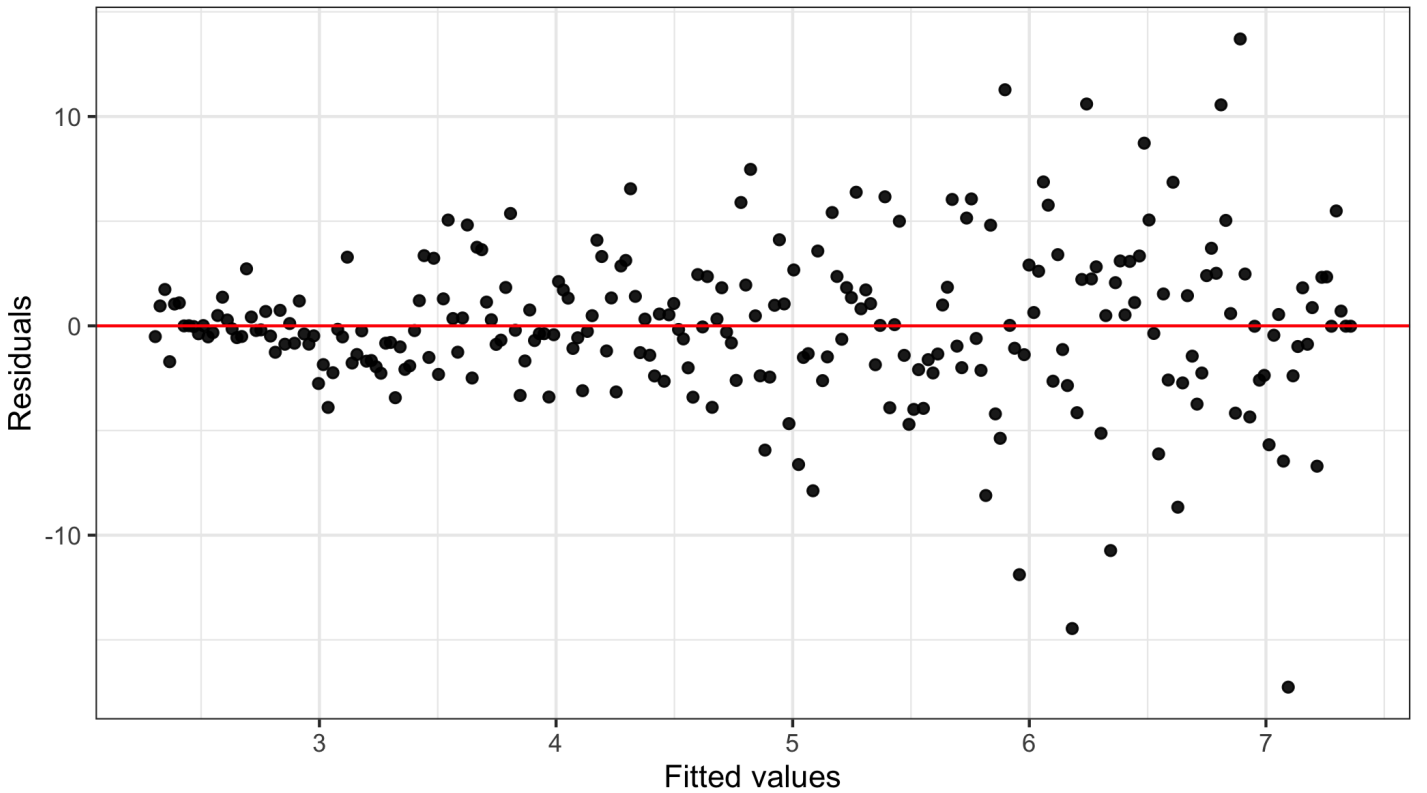


✅ The vertical spread of the residuals is relatively constant.

Violation: distinguishable pattern



Violation: non-constant variance

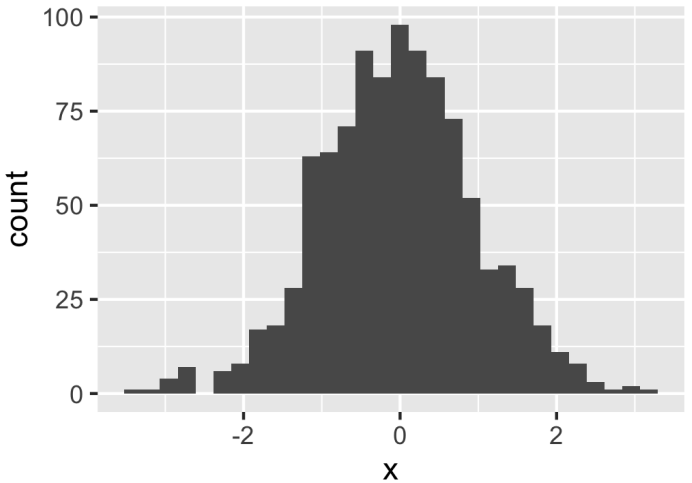


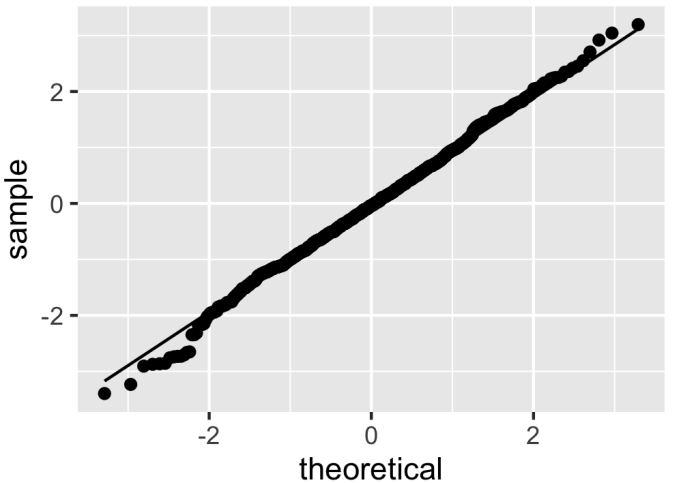
Normal quantile plot

A **normal quantile plot** is a scatterplot of the quantiles of the observed data (x-axis) versus the theoretical quantiles from a sample of the same size that is perfectly normal.

Normal quantile plot

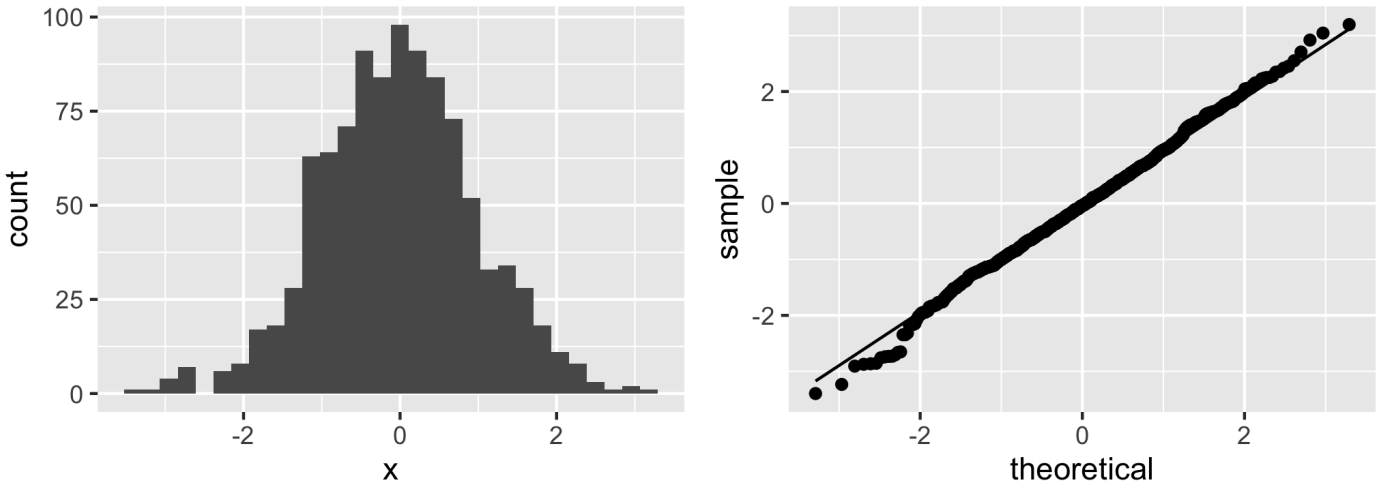
A **normal quantile plot** is a scatterplot of the quantiles of the observed data (x-axis) versus the theoretical quantiles from a sample of the same size that is perfectly normal.





Normal quantile plot

A **normal quantile plot** is a scatterplot of the quantiles of the observed data (x-axis) versus the theoretical quantiles from a sample of the same size that is perfectly normal.



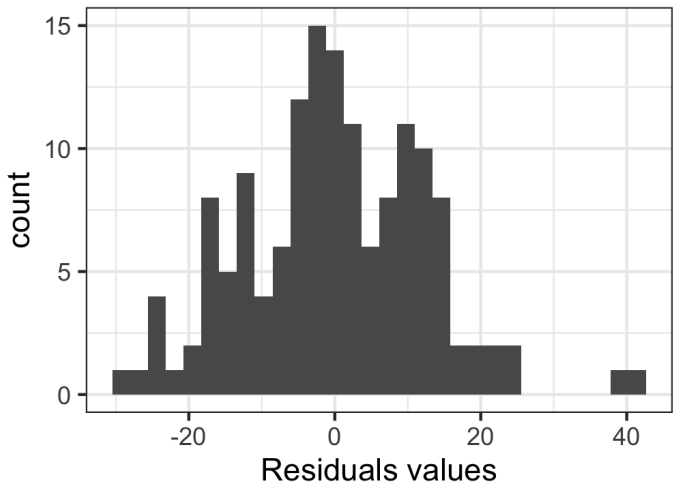
If the distribution of the observed data is approximately normal, the points will follow a straight diagonal line.

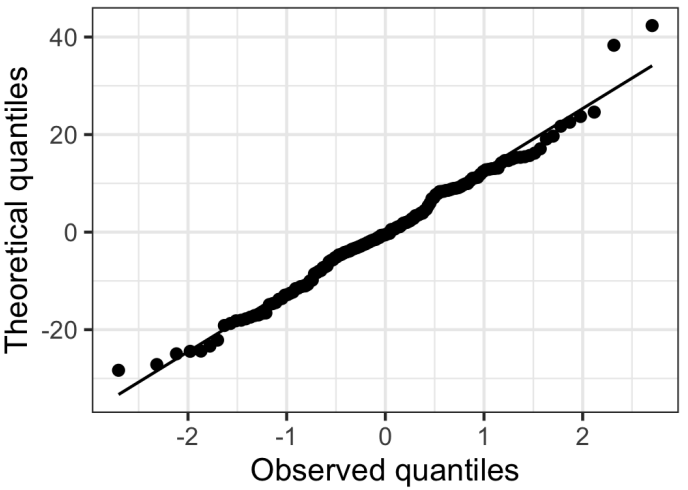
Assessing normality

Normality: The errors follow a normal distribution.

Assessing normality

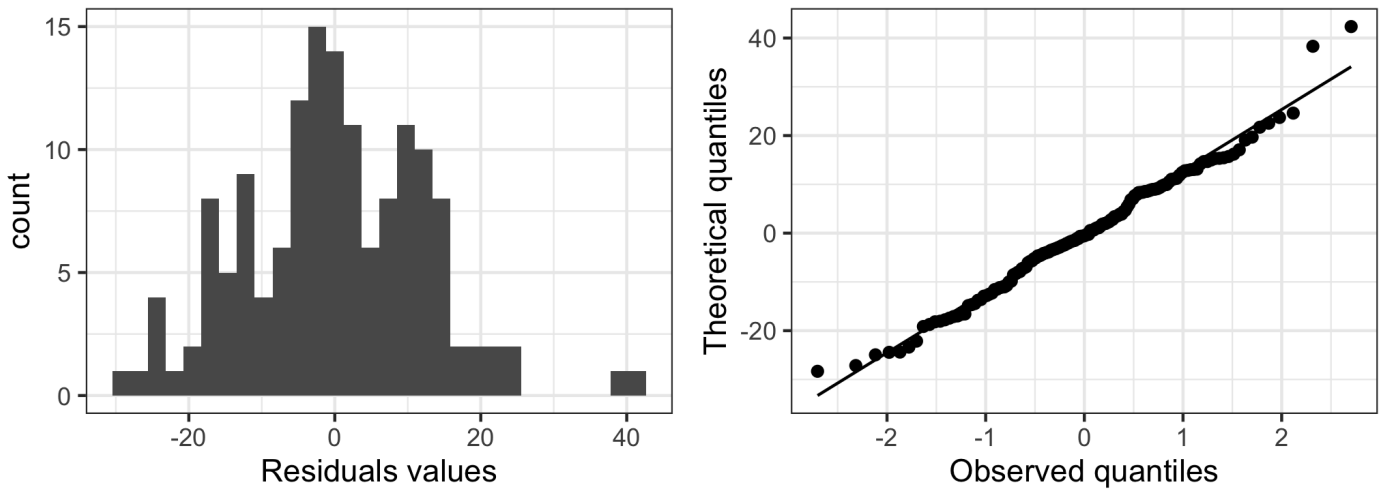
Normality: The errors follow a normal distribution.





Assessing normality

Normality: The errors follow a normal distribution.



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

Assessing independence

Independence: The errors are independent from one another.

Assessing independence

Independence: The errors are independent from one another.

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

Assessing independence

Independence: The errors are independent from one another.

- We can often assess 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 (e.g., over time), you can 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 provide sufficient evidence 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 provide sufficient evidence that

1️⃣ a different model should be proposed.

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

If not, the conditions are satisfied sufficiently enough to proceed.

Recap

Recap

- Identified the conditions for simple linear regression:
  1. Linearity
  2. Constant Variance
  3. Normality
  4. Independence

Recap

- Identified the conditions for simple linear regression:
  1. Linearity
  2. Constant Variance
  3. Normality
  4. Independence
- Used plots of the residuals to assess the conditions.

Simple Linear Regression

Assessing Conditions

Dr. Maria Tackett