

What are Residuals?



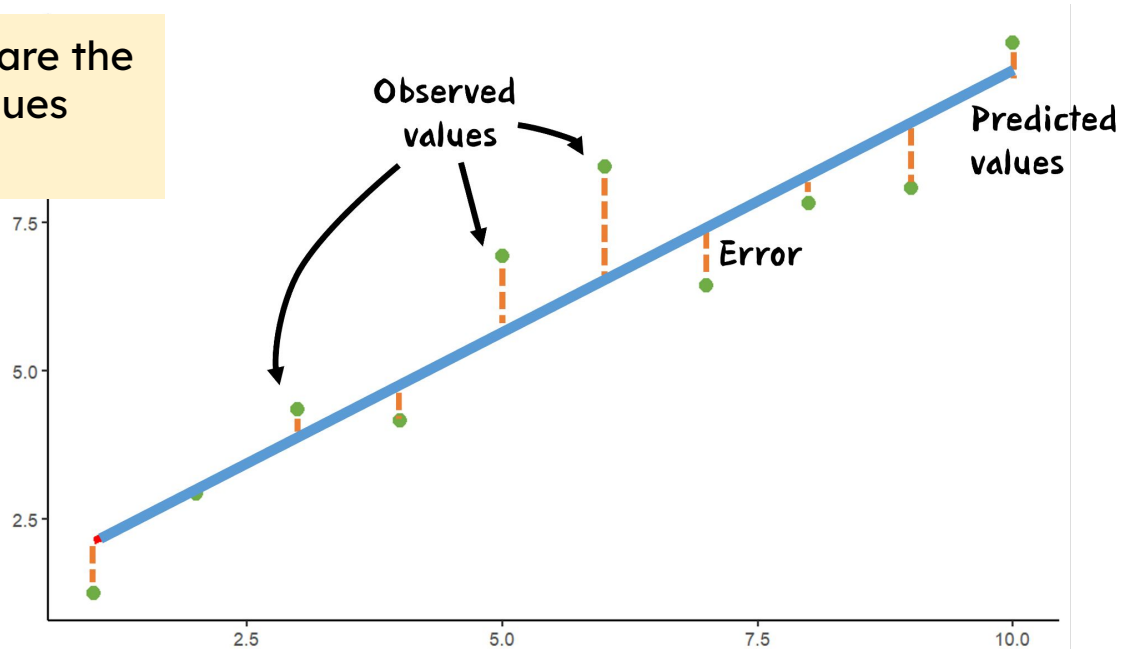
Introduction

In the world of data analysis, we often use models to understand and make predictions. One essential concept is "residuals."

These slides look at what residuals are and why they matter.

Introduction

Definition Residuals (or “errors”) are the differences between observed values and values predicted by a model.



Importance of residuals

- Residuals help us evaluate the accuracy of our predictions.
- They reveal if our model is missing important information
 - a. In agroecology, accurate predictions are vital for sustainable farming practices.
 - b. e.g., if our outcome were plot growth, are we missing key information that affects how well crops grow?



Example with Residuals

Let's say we have the following data:

x	y
-0.3340	-0.05646
1.7824	7.24161
0.6978	2.38251
-0.5383	-0.28679
-0.7619	-0.44329

Then we fit a linear model, which gives the following regression model:

$$y = 1.257 + 3.017 * x$$

Example with Residuals

Let's say we have the following data:

x	y
-0.3340	-0.05646
1.7824	7.24161
0.6978	2.38251
-0.5383	-0.28679
-0.7619	-0.44329

Then we fit a linear model, which gives the following regression model:

$$\hat{y} = 1.257 + 3.017x$$

If we put this to our data we can find the fitted values: that is, the values when the model is fitted to the observed data.

Example with Residuals

Then we fit a linear model, which gives the following regression model:

$$y = 1.257 + 3.017 * x$$

x	y	fitted_values
-0.3340	-0.05646	0.2493
1.7824	7.24161	6.6344
0.6978	2.38251	3.3623
-0.5383	-0.28679	-0.3669
-0.7619	-0.44329	-1.0415

E.g., for row 1:

$$1.257 + (3.017 * -0.3340) = 0.2493$$

Example with Residuals

We can next find the difference between our observed y values (y) and our fitted values (fitted_values). This gives the residuals.

$$y = 1.257 + 3.017 * x$$

Example with Residuals

We can next find the difference between our observed y values (y) and our fitted values (fitted_values). This gives the residuals.

$$y = 1.257 + 3.017 * x$$

x	y	fitted_values	residuals
-0.3340	-0.05646	0.2491	-0.3057
1.7824	7.24161	6.6351	0.6072
0.6978	2.38251	3.3625	-0.9797
-0.5383	-0.28679	-0.3672	0.0801
-0.7619	-0.44329	-1.0419	0.5982

E.g., for row 1:

$$-0.05646 - 0.2493 = -0.3057$$

Interpretation of Residuals

- Positive residuals mean our model underpredicted the outcome
- Negative residuals mean our model overpredicted the outcome.

