What is the definition of the constant variance assumption and can you provide an example?

Authored by

stats writer

May 11, 2024

The constant variance assumption is a statistical concept that states that the variance of a variable remains consistent across all values of that variable. This means that the spread of data points around the mean should be similar throughout the dataset. An example of this would be a study measuring the heights of different age groups. The constant variance assumption would suggest that the variability in height within each age group should be similar, rather than significantly increasing or decreasing with age. This assumption is important in various statistical methods, such as regression analysis, as it helps ensure the validity of the results.

The Constant Variance Assumption: Definition & Example

Linear regression is a technique we use to quantify the relationship between one or more predictor variables and a .

One of the key assumptions of linear regression is that the residuals have constant variance at every level of the predictor variable(s).

If this assumption is not met, the residuals are said to suffer from . When this occurs, the estimates for the model coefficients become unreliable.

How to Assess Constant Variance

The most common way to determine if the residuals of a regression model have constant variance is to create a fitted values vs. residuals plot.

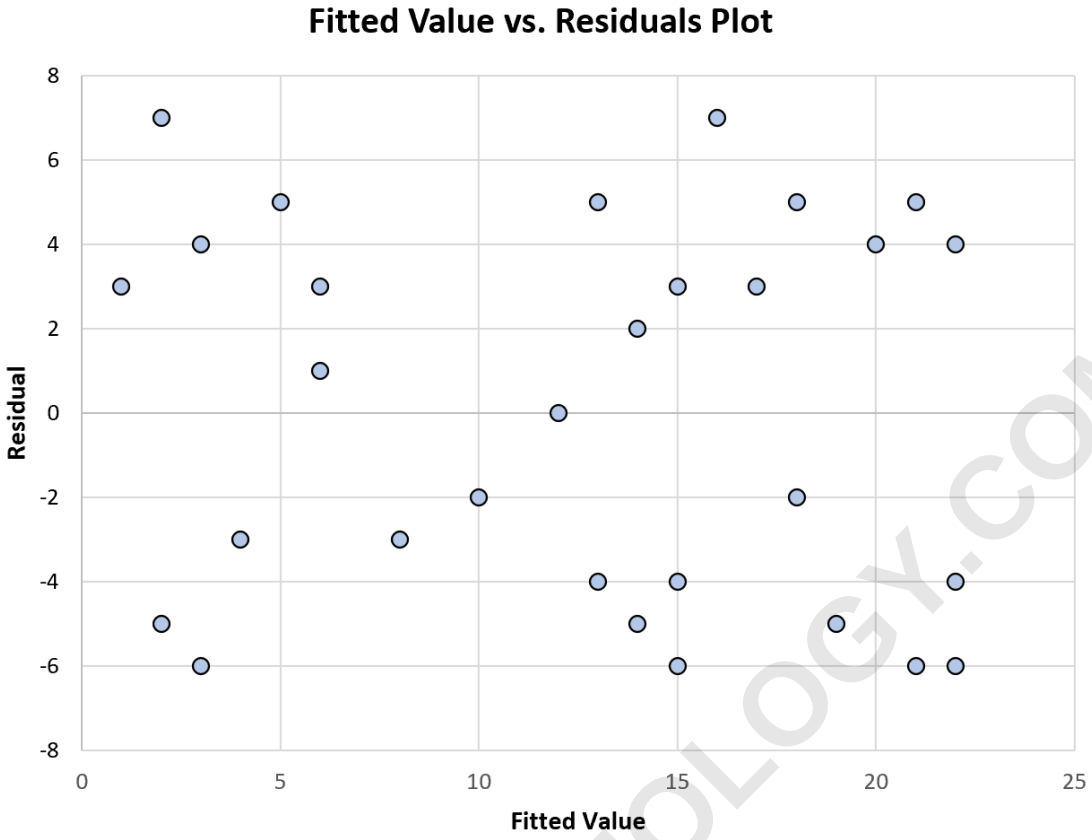
This is a type of plot that displays the fitted values of the regression model along the x-axis and the residuals of those fitted values along the y-axis.

If the spread of the residuals is roughly equal at each level of the fitted values, we say that the constant variance assumption is met.

Otherwise, if the spread of the residuals systematically increases or decreases, this assumption is likely violated.

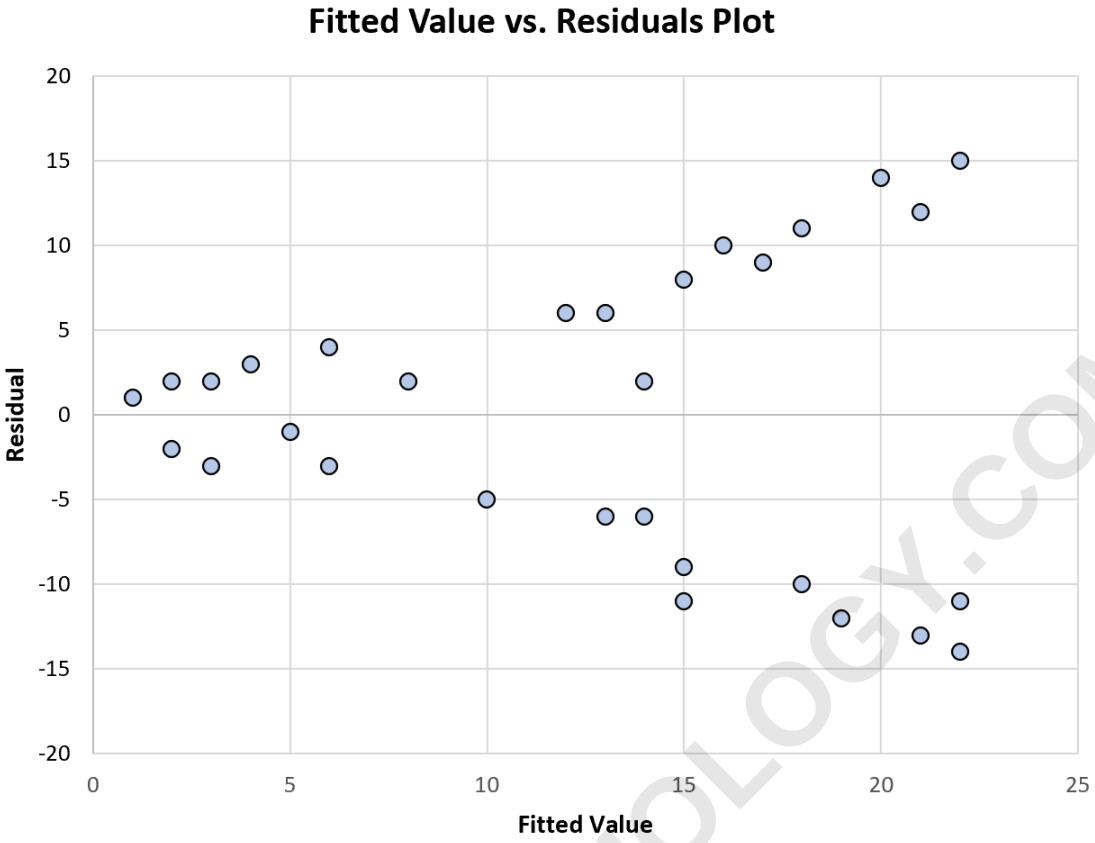
Note: This type of plot can only be created *after* fitting a regression model to the dataset.

The following plot shows an example of a fitted values vs. residual plot that displays constant variance:



Notice how the residuals are scattered randomly about zero in no particular pattern with roughly constant variance at every level of the fitted values.

The following plot shows an example of a fitted values vs. residual plot that displays non-constant variance:



Notice that the spread of the residuals grows larger and larger as the fitted values increase. This is a typical sign of non-constant variance.

This tells us that our regression model suffers from non-constant variance of residuals and thus the estimates for the model coefficients aren't reliable.

How to Fix a Violation of Constant Variance

1. Log Transformation: Transform the response variable from y to log(y)

2. Square Root Transformation: Transform the response variable from y to $\sqrt{y}$

3. Cube Root Transformation: Transform the response variable from y to y1/3

By performing these transformations, the problem of non-constant variance typically goes away.

The following tutorials provide additional information about linear regression and residual analysis: