“What is Considered a Good vs. Bad Residual Plot?”

Authored by

stats writer

June 24, 2024

A residual plot is a visual representation of the difference between the observed values and the predicted values in a statistical model. It is used to evaluate the performance of a model and determine whether it is a good fit for the data. A good residual plot should have a random scatter of points around the horizontal axis, indicating that the model has captured the underlying patterns in the data. On the other hand, a bad residual plot will show a clear pattern or trend in the points, suggesting that the model is not able to accurately predict the data. Therefore, a good residual plot is characterized by a lack of pattern or trend, while a bad residual plot shows a clear pattern or trend.

What is Considered a Good vs. Bad Residual Plot?

In regression analysis, a residual plot is a type of plot that displays the fitted values of a regression model on the x-axis and the residuals of the model along the y-axis.

When visually inspecting a residual plot, there are two things we typically look for to determine if the plot is "good" or "bad":

1. Do the residuals exhibit a clear pattern?

In a "good" residual plot, the residuals exhibit no clear pattern.In a "bad" residual plot, the residuals exhibit some type of pattern such as a curve or a wave. This is an indication that the regression model we used is does not provide an appropriate fit to the data.

2. Do the residuals increase or decrease in variance in a systematic way?

In a "good" residual plot, the residuals are randomly scattered about zero with no systematic increase or decrease in variance.In a "bad" residual plot, the variance of the residuals increase or decrease in a systematic way.

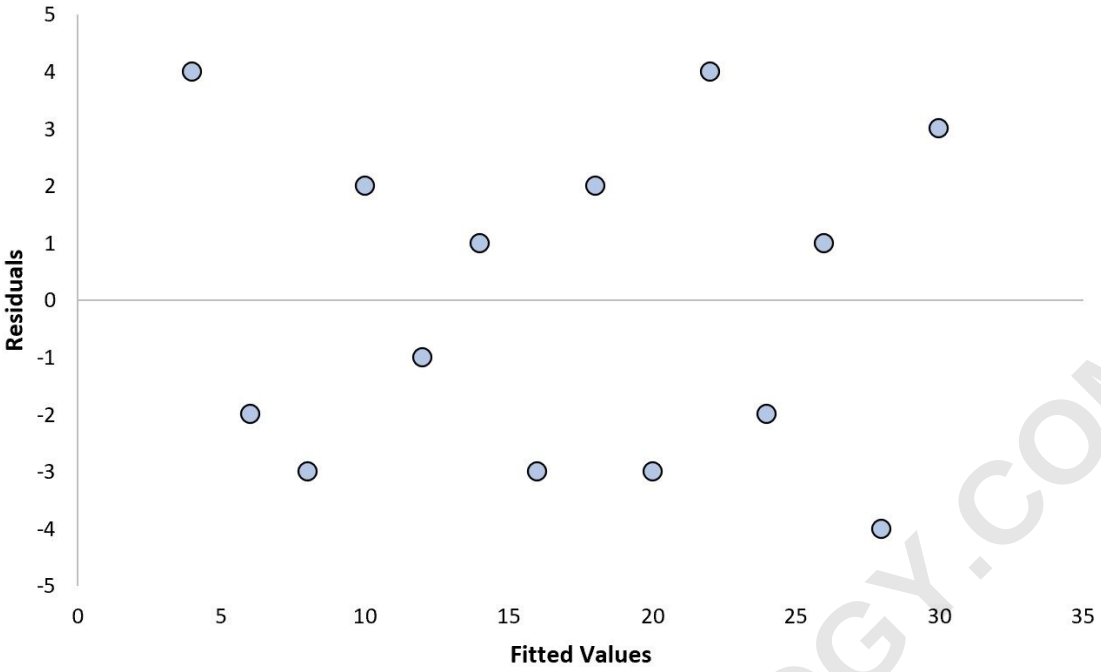
If a residual plot is deemed "good" then it means we can trust the results of the regression model and it's safe to interpret the coefficients in the model.

However, if a residual plot is deemed "bad" then it means the results of the model are untrustworthy and we need to fit a different regression model to the data.

The following examples how to interpret "good" vs. "bad residual plots in practice.

Example 1: A "Good" Residual Plot

Suppose we fit a regression model and end up with the following residual plot:



We can answer the following two questions to determine if this is a "good" residual plot:

1. Do the residuals exhibit a clear pattern?

No. The residuals are randomly scattered about zero with no clear pattern.

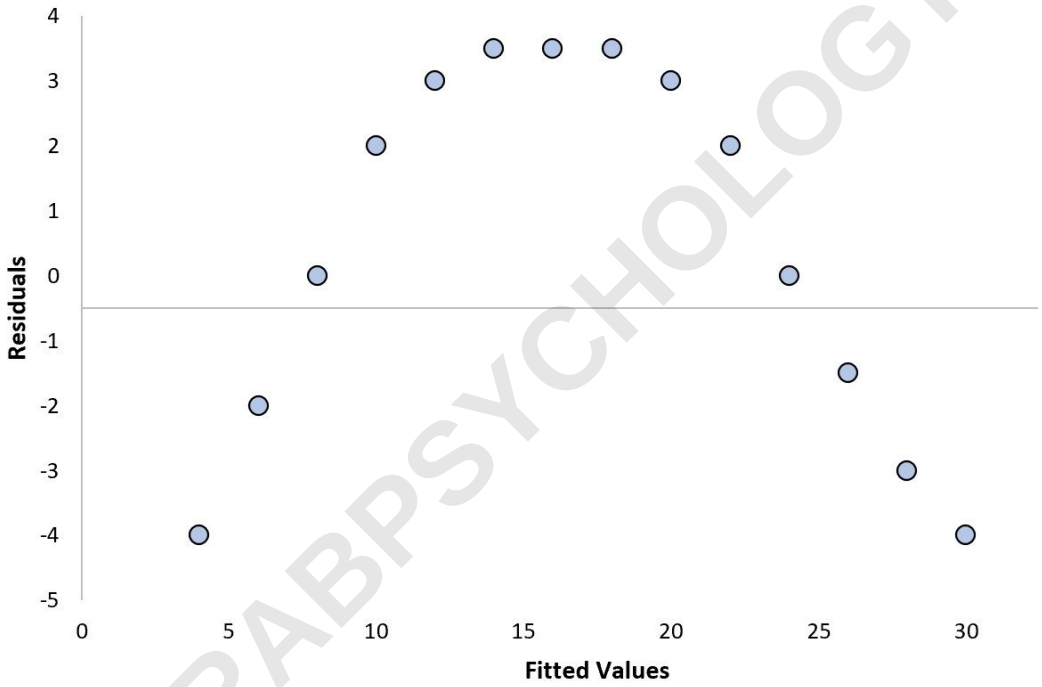
2. Do the residuals increase or decrease in variance in a systematic way?

No. The residuals have fairly constant variance (i.e. the distance between the residuals and the value zero) at each level of the fitted values.

Therefore, we can trust the results of the regression model and it's safe to interpret the coefficients in the model.

Example 2: A "Bad" Residual Plot with a Clear Pattern

Suppose we fit a regression model and end up with the following residual plot:



We can answer the following two questions to determine if this is a "good" residual plot:

1. Do the residuals exhibit a clear pattern?

Yes. The residuals exhibit a curved pattern.

2. Do the residuals increase or decrease in variance in a systematic way?

Yes. The residuals have different levels of variance at different levels of the fitted values.

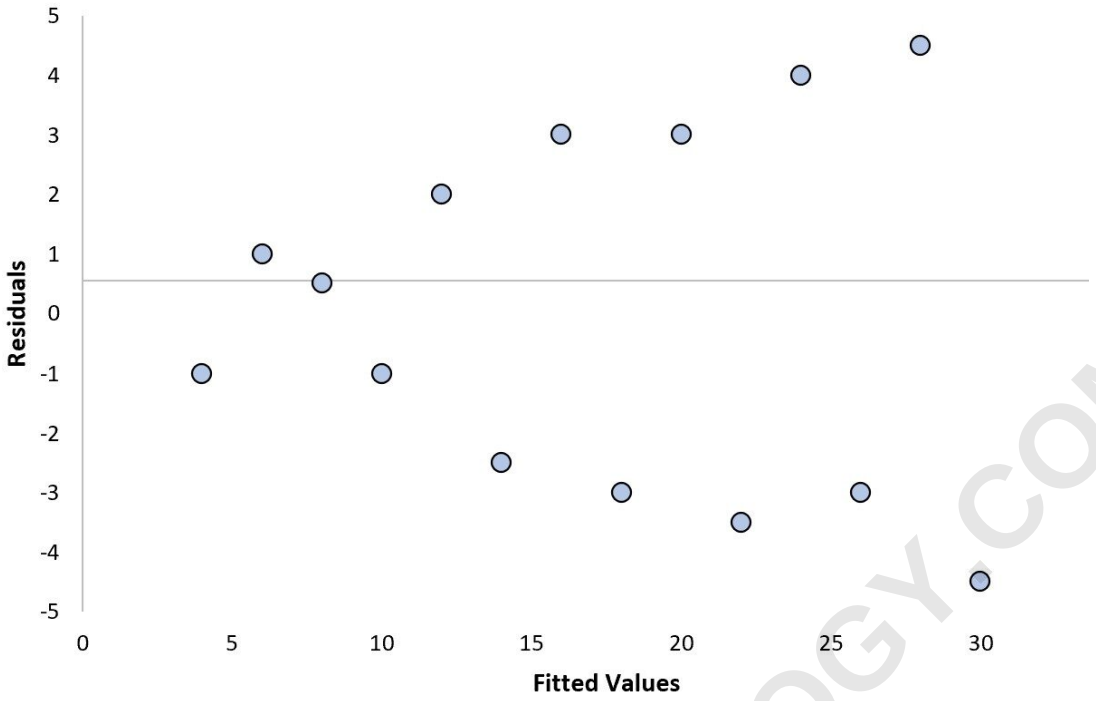
Since we answered "Yes" to at least one of these questions, we would consider this to be a "bad" residual plot.

This means the regression model does not provide a good fit to the data.

In particular, the in the residual plot indicates that a linear regression model does a poor job of fitting the data and that a quadratic regression model would likely do a better job.

Example 3: A "Bad" Residual Plot with Increasing Variance

Suppose we fit a regression model and end up with the following residual plot:



We can answer the following two questions to determine if this is a "good" residual plot:

1. Do the residuals exhibit a clear pattern?

No. There is no clear pattern in the residuals.

2. Do the residuals increase or decrease in variance in a systematic way?

Yes. The variance of the residuals increases as the fitted values increase.

Since we answered "Yes" to at least one of these

questions, we would consider this to be a "bad" residual plot.

In this particular example the residuals suffer from heteroscedasticity, which refers to unequal variance of the residuals at different levels of the fitted values.

This means the results of the regression model may be untrustworthy.

Refer to for various ways that you can address the problem of heteroscedasticity in a regression model.

The following tutorials explain how to create residual plots using different statistical software: