What is the process for interpreting a curved residual plot, and can you provide an example?

Authored by

stats writer

June 26, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the process for interpreting a curved residual plot, and can you provide an example?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=153576

ARABPSYCHOLOGY.COM

The process for interpreting a curved residual plot involves analyzing the pattern and distribution of the residuals, which are the differences between the observed and predicted values in a regression model. A curved residual plot suggests that the relationship between the independent and dependent variables is not linear and that the model may not be the best fit for the data.

To interpret a curved residual plot, one should first check for a clear pattern or trend in the residuals. A curved line or U-shaped pattern indicates that the model is not capturing all the non-linear relationships between the variables. In this case, a different model, such as a polynomial or spline regression, may be more appropriate.

Additionally, the distribution of the residuals should be examined. A curved residual plot with a random distribution of points suggests that the model is adequate. However, if the points are clustered or have a distinct pattern, it indicates that the model is not accurately capturing the variability in the data.

For example, if we are trying to predict housing prices based on square footage, a linear regression model may result in a curved residual plot. This suggests that the relationship between the two variables is not strictly linear and may require a different model, such as a quadratic or cubic regression, to better fit the data. Additionally, if the residual plot shows a U-shaped pattern, it indicates that the model may be overestimating prices for smaller houses and underestimating prices for larger houses. This insight can help improve the accuracy of the model and make more reliable predictions.

ARABPSYCHOLOGY.COM

Interpret a Curved Residual Plot (With Example)

Residual plots are used to assess whether or not the in a regression model are normally distributed and whether or not they exhibit .

Ideally, you would like the points in a residual plot to be randomly scattered around a value of zero with no clear pattern.

If you encounter a residual plot where the points in the

plot have a curved pattern, it likely means that the regression model you have specified for the data is not correct.

In most cases, it means that you attempted to fit a linear regression model to a dataset that instead follows a quadratic trend.

The following example shows how to interpret (and fix) a curved residual plot in practice.

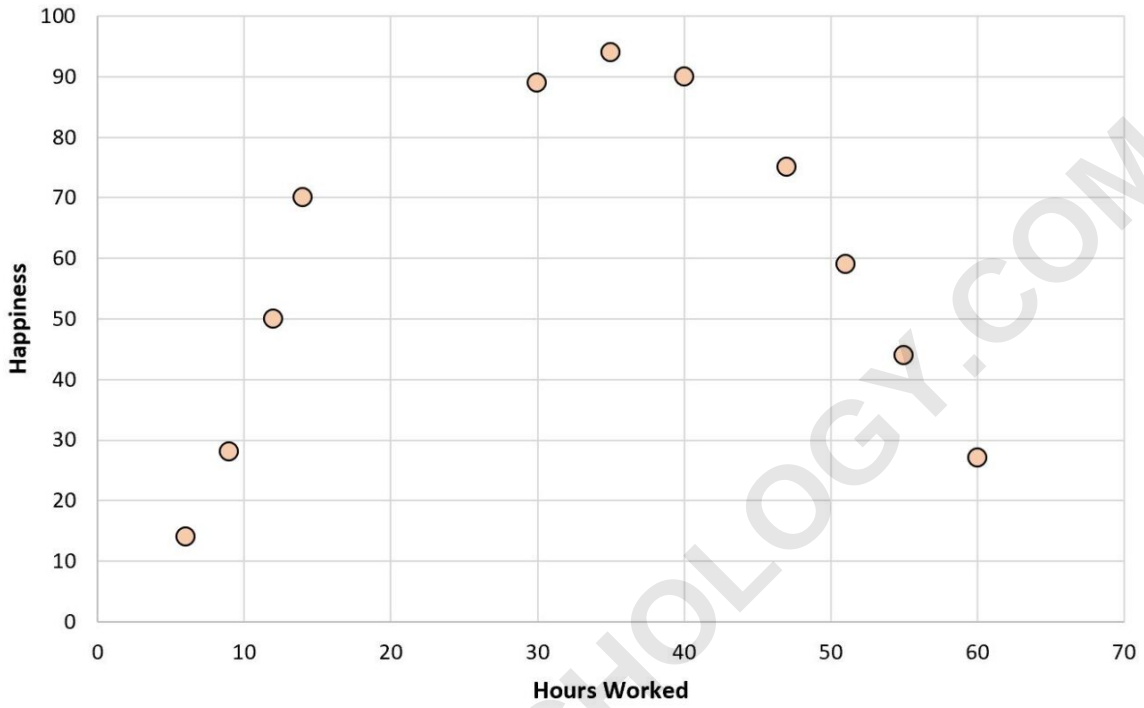
Example: Interpreting a Curved Residual Plot

Suppose we collect the following data on the number of hours worked per week and the reported happiness level (on a scale of 0-100) for 11 different people in some office:

Hours	Happiness
6	14
9	28
12	50
14	70
30	89
35	94
40	90
47	75
51	59
55	44
60	27

ARABPSYCHOLOGY.COM

If we create a simple scatter plot of hours worked vs. happiness level, here's what it would look like:



ARABPSYCHOLOGY.COM

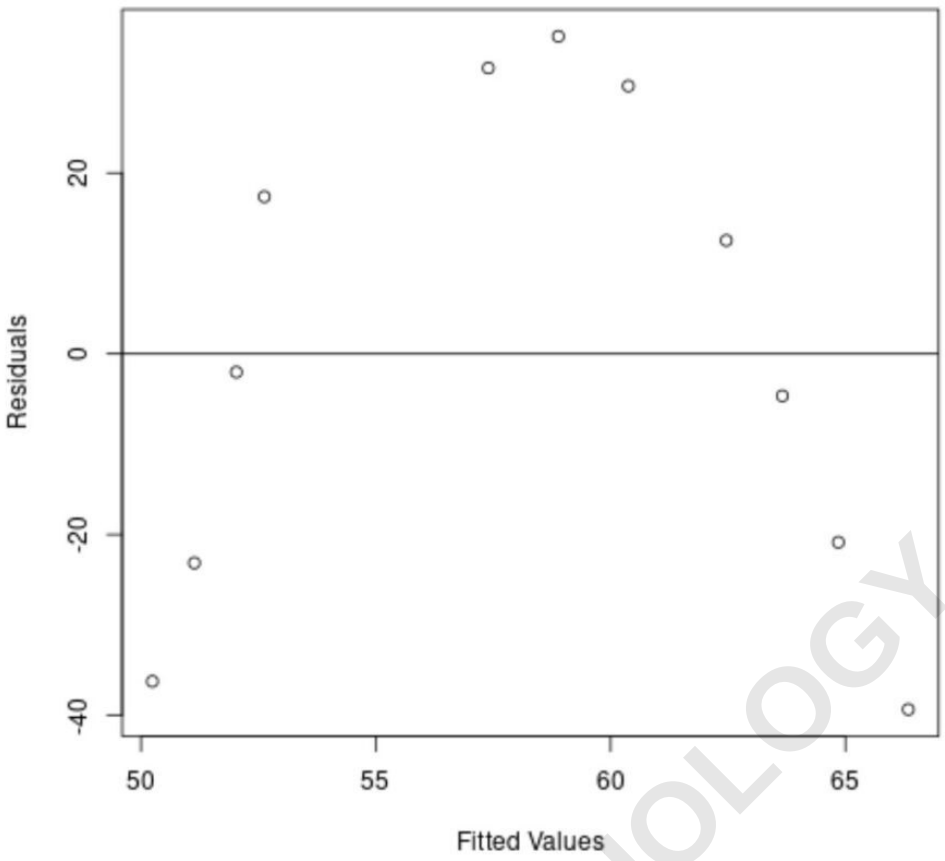
Now suppose we would like to fit a regression model using hours worked to predict happiness level.

The following code shows how fit a simple linear regression model to this dataset and produce a residual plot in R:

```
#create dataframe
df <- data.frame(hours=c(6, 9, 12, 14, 30, 35, 40, 47, 51, 55, 60),
happiness=c(14, 28, 50, 70, 89, 94, 90, 75, 59, 44, 27))
```

```
#fit linear regression model
linear_model <- lm(happiness ~ hours, data=df)

#get list of residuals
res <- resid(linear_model)

#produce residual vs. fitted plot
plot(fitted(linear_model), res, xlab='Fitted Values',
ylab='Residuals')

#add a horizontal line at 0
abline(0,0)
```

ARABPSYCHOLOGY.COM



ARABPSYCHOLOGY.COM

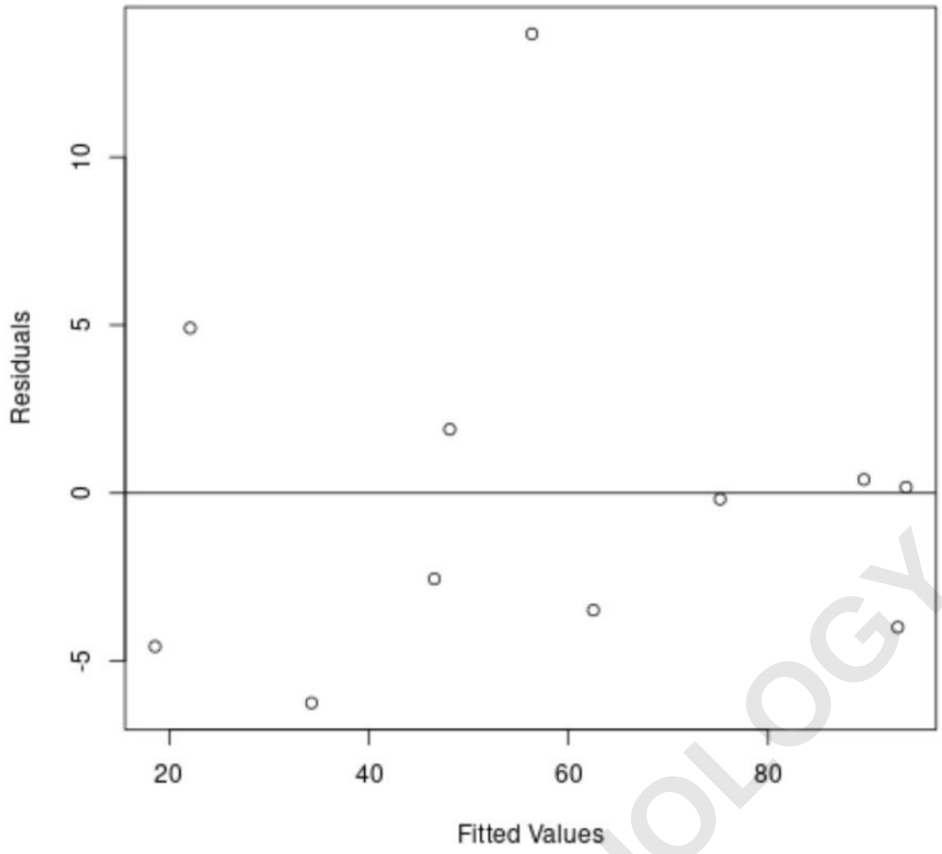
The x-axis displays the fitted values and the y-axis displays the residuals.

From the plot we can see that there is a curved pattern in the residuals, which indicates that a linear regression model does not provide an appropriate fit to this dataset.

The following code shows how fit a quadratic regression model to this dataset and produce a residual plot in R:

```
#create dataframe
df <- data.frame(hours=c(6, 9, 12, 14, 30, 35, 40, 47, 51, 55, 60),
happiness=c(14, 28, 50, 70, 89, 94, 90, 75, 59, 44, 27))
#define quadratic term to use in model
df$hours2 <- df$hours^2

#fit quadratic regression model
quadratic_model <- lm(happiness ~ hours + hours2, data=df)

#get list of residuals
res <- resid(quadratic_model)

#produce residual vs. fitted plot
plot(fitted(quadratic_model), res, xlab='Fitted Values', ylab='Residuals')

#add a horizontal line at 0
abline(0,0)
```

ARABPSYCHOLOGY.COM



ARABPSYCHOLOGY.COM

Once again the x-axis displays the fitted values and the y-axis displays the residuals.

From the plot we can see that the residuals are randomly scattered around zero and there is no clear pattern in the residuals.

This tells us that a quadratic regression model does a much better job of fitting this dataset compared to a linear regression model.

This should make sense considering we saw that the

true relationship between hours worked and happiness level appeared to be quadratic instead of linear.

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

ARABPSYCHOLOGY.COM