How can I calculate standardized residuals in R?

Authored by

stats writer

April 23, 2024

Standardized residuals are a statistical measure used to assess the difference between observed and expected values in a regression model. In R, these residuals can be calculated using the "residuals" function, which takes into account the mean and standard deviation of the dependent variable. To obtain standardized residuals, these values are divided by the standard deviation of the residuals. This process allows for a standardized and more meaningful interpretation of the residuals, making it easier to identify any outliers or influential data points in the model. The resulting values can also be used to check for model assumptions and assess the overall fit of the regression model.

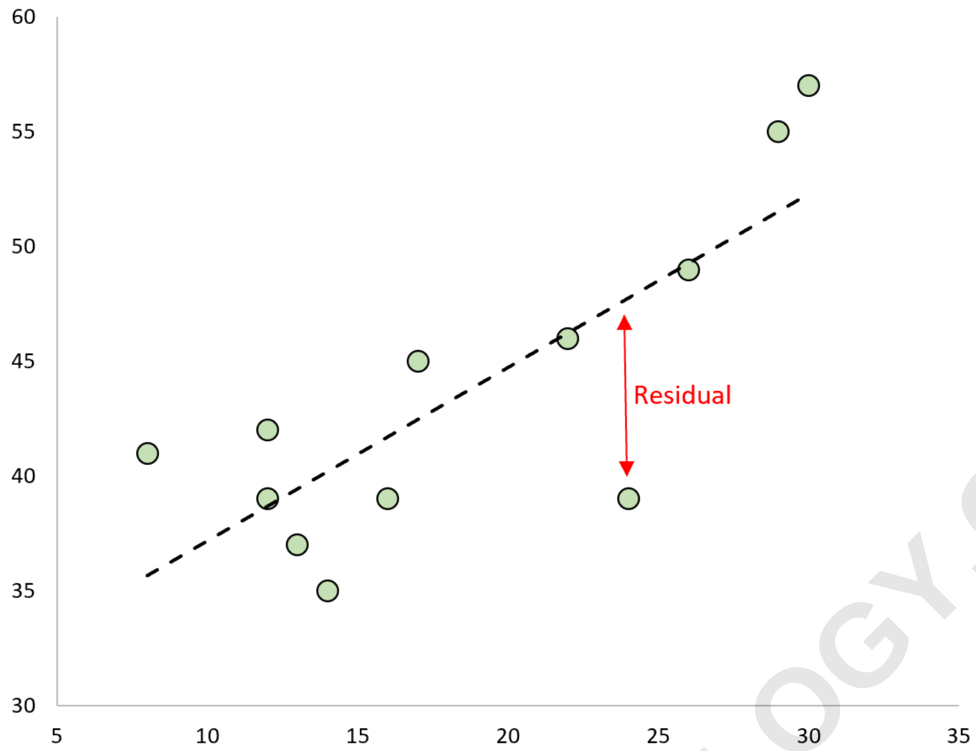
Calculate Standardized Residuals in R

A residual is the difference between an observed value and a predicted value in a regression model.

It is calculated as:

Residual = Observed value - Predicted value

If we plot the observed values and overlay the fitted regression line, the residuals for each would be the vertical distance between the observation and the regression line:



One type of residual we often use to identify outliers in a regression model is known as a standardized residual.

It is calculated as:

ri = ei / s(ei) = ei / RSE√1-hii

where:

ei: The ith residualRSE: The residual standard error of the modelhii: The leverage of the ith observation

In practice, we often consider any standardized residual

with an absolute value greater than 3 to be an outlier.

This tutorial provides a step-by-step example of how to calculate standardized residuals in R.

Step 1: Enter the Data

First, we'll create a small dataset to work with in R:

```
#create data
data <- data.frame(x=c(8, 12, 12, 13, 14, 16, 17, 22, 24, 26, 29, 30),
y=c(41, 42, 39, 37, 35, 39, 45, 46, 39, 49, 55, 57))

#view data
data

x y
1 8 41
2 12 42
3 12 39
4 13 37
5 14 35
6 16 39
7 17 45
8 22 46
```

9 24 39

10 26 49

11 29 55

12 30 57

Step 2: Fit the Regression Model

Next, we'll use the lm() function to fit a :

```
#fit model
model <- lm(y ~ x, data=data)

#view model summary
summary(model)Call:
lm(formula = y ~ x, data = data)

Residuals:
Min 1Q Median 3Q Max
-8.7578 -2.5161 0.0292 3.3457 5.3268

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.6309 3.6189 8.188 9.6e-06 ***
x 0.7553 0.1821 4.148 0.00199 **
---
```

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Residual standard error: 4.442 on 10 degrees of freedom
Multiple R-squared: 0.6324, Adjusted R-squared: 0.5956
F-statistic: 17.2 on 1 and 10 DF, p-value: 0.001988

Step 3: Calculate the Standardized Residuals

Next, we'll use the built-in rstandard() function to calculate the standardized residuals of the model:

#calculate the standardized residuals
standard_res <- rstandard(model)

#view the standardized residuals
standard_res

1 2 3 4 5 6
1.40517322 0.81017562 0.07491009 -0.59323342 -1.24820530 -0.64248883
7 8 9 10 11 12
0.59610905 -0.05876884 -2.11711982 -0.06655600 0.91057211 1.26973888

We can add the standardized residuals back to the

original data frame if we'd like:

#column bind standardized residuals back to original data frame
final_data <- cbind(data, standard_res)

#view data frame
x y standard_res
1 8 41 1.40517322
2 12 42 0.81017562
3 12 39 0.07491009
4 13 37 -0.59323342
5 14 35 -1.24820530
6 16 39 -0.64248883
7 17 45 0.59610905
8 22 46 -0.05876884
9 24 39 -2.11711982
10 26 49 -0.06655600
11 29 55 0.91057211
12 30 57 1.26973888

We can then sort each observation from largest to smallest according to its standardized residual to get an idea of which observations are closest to being outliers:

#sort standardized residuals descending
final_data

x y standard_res
1 8 41 1.40517322
12 30 57 1.26973888
11 29 55 0.91057211
2 12 42 0.81017562
7 17 45 0.59610905
3 12 39 0.07491009
8 22 46 -0.05876884
10 26 49 -0.06655600
4 13 37 -0.59323342
6 16 39 -0.64248883
5 14 35 -1.24820530
9 24 39 -2.11711982

From the results we can see that none of the standardized residuals exceed an absolute value of 3. Thus, none of the observations appear to be outliers.

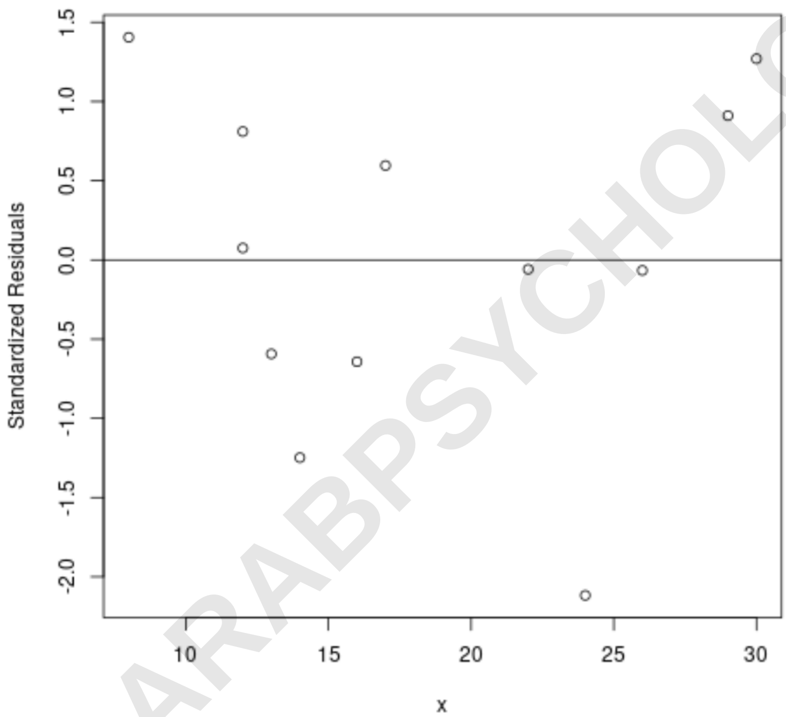
Step 4: Visualize the Standardized Residuals

Lastly, we can create a scatterplot to visualize the values for the predictor variable vs. the standardized

residuals:

```
#plot predictor variable vs. standardized residuals
plot(final_data$x, standard_res, ylab='Standardized Residuals', xlab='x')

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



What Are Standardized Residuals?