

What Are Standardized Residuals?

Authored by
Mohammed loot

November 6, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *What Are Standardized Residuals?*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=11472>

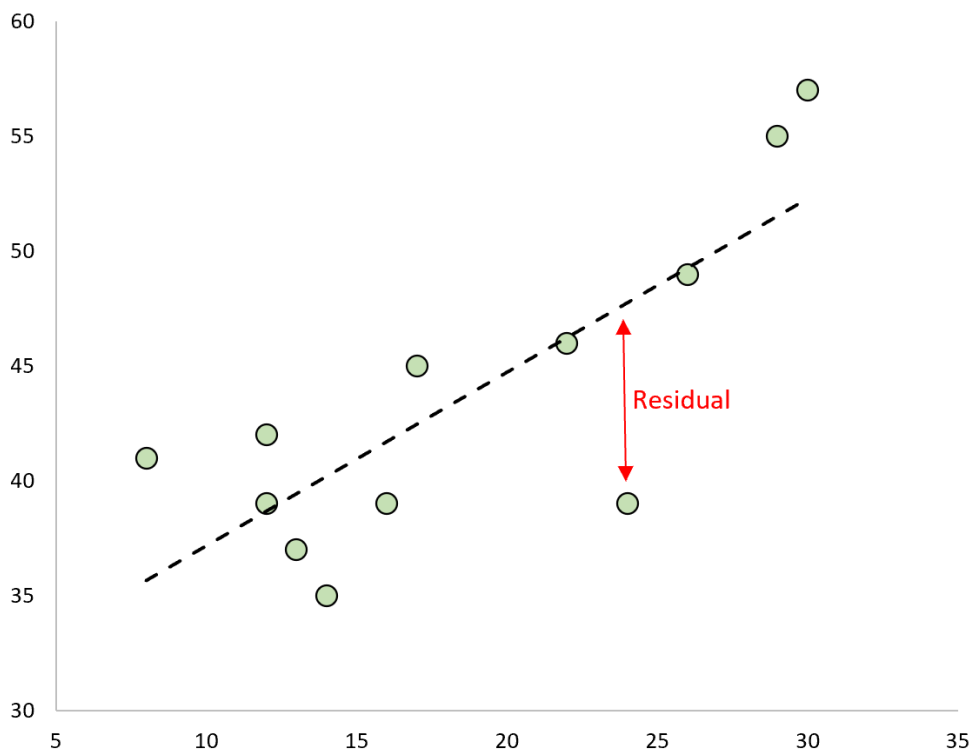
In the field of statistics, particularly within [regression models](#), understanding the discrepancy between actual data points and the model's predictions is crucial. This difference is known as a [residual](#).

A **residual** is fundamentally the vertical distance between an **observed value** and its corresponding **predicted value** generated by the fitted regression line. It quantifies how well the model fits a specific data point.

The calculation is straightforward:

Residual = Observed value - Predicted value

When visualizing a simple linear regression, the residual for each observation is clearly seen as the vertical line segment extending from the data point to the regression line:



What is a Residual in Regression Analysis?

The [residual](#) serves as a core diagnostic tool in statistical modeling. Analyzing these differences helps researchers assess the validity of the model assumptions, such as linearity and homoscedasticity.

If a residual is positive, the model underestimated the actual value; if it is negative, the model

overestimated the actual value. Ideally, residuals should be small and randomly distributed around zero, indicating a good fit.

However, simply looking at raw residuals can be misleading, especially when comparing observations across different scales or models. This is where normalization becomes necessary, leading us to the concept of standardized residuals.

Introducing Standardized Residuals

While raw residuals provide the absolute error, they do not account for the variability inherent in the data or the influence of the observation itself. To address this, we use a normalized form known as the **standardized residual** (sometimes referred to as the "internally studentized residual").

Standardized residuals scale the raw residuals by dividing them by an estimate of their standard deviation. This scaling process makes the errors comparable across all observations, allowing for a more objective identification of influential points or [outliers](#).

These residuals are particularly valuable for diagnostics because they transform the error into a unitless measure, similar to a Z-score, which simplifies the application of standard statistical thresholds.

The Standardized Residual Formula Explained

The calculation of a [standardized residual](#), denoted as r_i , requires two additional components beyond the raw residual itself: the model's overall error variance and the observation's statistical influence.

The formula is defined mathematically as:

$$r_i = e_i / s(e_i) = e_i / \text{RSE} \sqrt{1-h_{ii}}$$

Where the components represent the following crucial statistical measures:

e_i : The i th raw [residual](#), calculated as the observed Y minus the predicted Y .

RSE: The [Residual Standard Error](#) of the model. This is an estimate of the standard deviation of the error term (ϵ).

h_{ii} : The [leverage](#) of the i th observation. Leverage measures how far an observation's predictor values are from the mean of the predictor values, indicating its potential influence on the regression line slope.

By incorporating the [RSE](#) and the [leverage](#), the denominator acts as a corrected standard deviation for the specific residual, ensuring the standardization is accurate based on the model's

overall fit and the data point's position.

Identifying Outliers using Standardized Residuals

One of the primary uses of standardized residuals is the robust identification of potential [outliers](#). Since these residuals are standardized, we can apply general statistical rules derived from the normal distribution to determine unusual data points.

A common rule of thumb in statistical practice dictates that any standardized residual with an absolute value greater than 3 is considered a potential outlier. Given that standardized residuals approximate a standard normal distribution (Z-scores), an absolute value exceeding 3 suggests an observation that is highly unlikely to have occurred if the model were correctly specified and the errors were normally distributed.

It is important to note that identifying an outlier does not automatically mean removal. These observations require further investigation to determine if they stem from genuine variation, a data entry error, or a specific anomaly in the process being modeled. Researchers in certain fields may use a slightly stricter threshold, sometimes considering an absolute value greater than 2 to signal observations worthy of closer scrutiny.

Practical Example: Calculating Standardized Residuals

To illustrate the process, let us consider a sample dataset comprising 12 observations, featuring one predictor variable (X) and one response variable (Y). We aim to fit a linear [regression model](#) to this data.

The initial dataset is presented below:

X	Y
8	41
12	42
12	39
13	37
14	35
16	39
17	45
22	46
24	39
26	49
29	55
30	57

Using statistical software (like R, Excel, Python, Stata, etc.), we first fit the optimal linear regression line to this data. For this example, assume the calculated line of best fit is:

$$y? = 29.63 + 0.7553 * X$$

This equation forms the basis for predicting Y values and calculating the initial raw residuals for every data point.

Step-by-Step Calculation Walkthrough

We begin by calculating the predicted value ($y?$) and the raw residual (e_i) for each observation. For instance, considering the first observation where $X = 8$ and $Y = 41$:

$$\text{Predicted } Y (y?) = 29.63 + 0.7553 * (8) = \mathbf{35.67}$$

$$\text{Raw Residual } (e_i) = \text{Observed } Y - \text{Predicted } Y = 41 - 35.67 = \mathbf{5.33}$$

Repeating this procedure for all 12 observations yields the complete list of raw residuals:

X	Y	Predicted Value	Residual
8	41	35.67	5.33
12	42	38.69	3.31
12	39	38.69	0.31
13	37	39.45	-2.45
14	35	40.20	-5.20
16	39	41.71	-2.71
17	45	42.47	2.53
22	46	46.25	-0.25
24	39	47.76	-8.76
26	49	49.27	-0.27
29	55	51.53	3.47
30	57	52.29	4.71

Next, we require the [Residual Standard Error](#) (RSE) and the [leverage](#) (h_{ii}) for each point. For this specific model, the RSE is determined to be **4.44**. While the calculation of the leverage statistic is typically automated by statistical packages, its values are necessary for the final standardization step:

X	Y	Predicted Value	Residual	h_{ii}
8	41	35.67	5.33	0.27
12	42	38.69	3.31	0.16
12	39	38.69	0.31	0.16
13	37	39.45	-2.45	0.14
14	35	40.20	-5.20	0.12
16	39	41.71	-2.71	0.09
17	45	42.47	2.53	0.09
22	46	46.25	-0.25	0.10
24	39	47.76	-8.76	0.13
26	49	49.27	-0.27	0.18
29	55	51.53	3.47	0.27
30	57	52.29	4.71	0.30

Finally, we apply the standardized residual formula, utilizing the raw residual (e_i), RSE, and leverage (h_{ii}). For the first observation, where $e_i = 5.33$ and $h_{ii} = 0.27$:

$$r_i = 5.33 / 4.44\sqrt{1-0.27} = 1.404$$

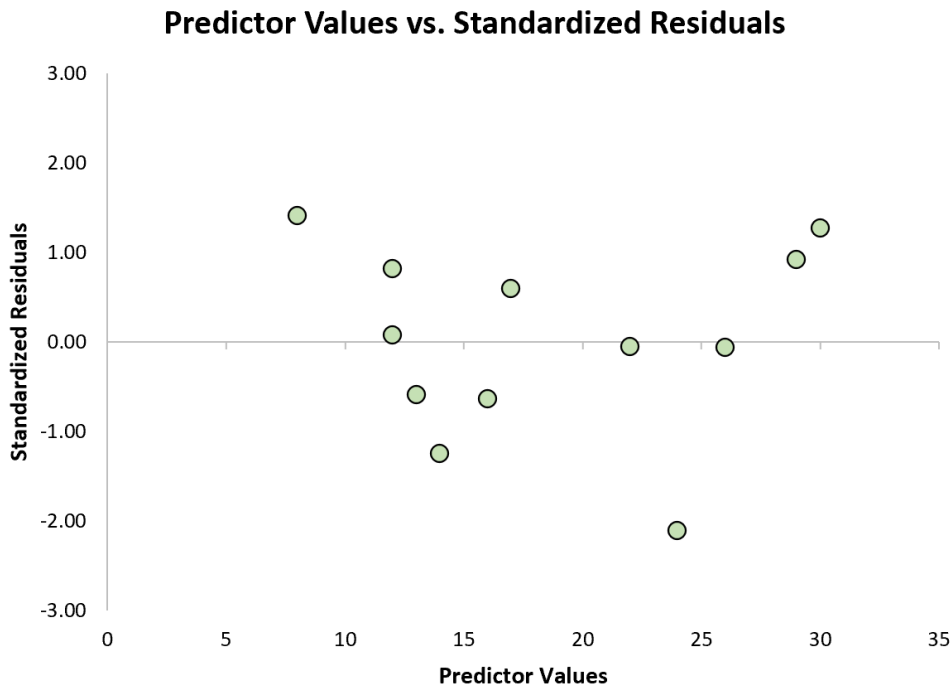
After calculating r_i for every point, the complete list of standardized residuals is compiled:

X	Y	Predicted Value	Residual	h_{ii}	Standardized Residual
8	41	35.67	5.33	0.27	1.40
12	42	38.69	3.31	0.16	0.81
12	39	38.69	0.31	0.16	0.08
13	37	39.45	-2.45	0.14	-0.59
14	35	40.20	-5.20	0.12	-1.25
16	39	41.71	-2.71	0.09	-0.64
17	45	42.47	2.53	0.09	0.60
22	46	46.25	-0.25	0.10	-0.06
24	39	47.76	-8.76	0.13	-2.11
26	49	49.27	-0.27	0.18	-0.07
29	55	51.53	3.47	0.27	0.91
30	57	52.29	4.71	0.30	1.27

Visualizing Results and Interpretation

The most effective way to interpret standardized residuals is through a diagnostic plot, typically plotting the predictor values against the calculated standardized residuals. This visualization allows us to check for patterns, non-linearity, and, most importantly, identify extreme values that cross the predetermined outlier thresholds.

Here is the scatterplot generated from our example data:



By observing the plot, we can confirm that none of the standardized residuals exceed the common absolute threshold of 3. Therefore, based on this metric, none of the observations in this dataset are classified as significant [outliers](#).

The choice between an absolute threshold of 2 or 3 often depends on the field of study and the sensitivity required for the analysis. Regardless of the chosen threshold, standardized residuals remain an indispensable tool for model diagnostics, ensuring the integrity and reliability of the regression analysis.

Additional Resources

The following tutorials provide additional information about standardized residuals and related regression diagnostics: