

How to Calculate Residuals in Regression Analysis: A Step-by-Step Guide

Authored by
stats writer

March 5, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Calculate Residuals in Regression Analysis: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=134023>

Residuals in regression analysis refer to the difference between the actual data points and the predicted values from a regression model. To calculate residuals, the predicted values are subtracted from the actual values for each data point. The resulting values represent the error or deviation from the model's prediction. These residuals are then used to assess the accuracy and validity of the regression model, as well as to identify any patterns or trends in the data. This process is essential in evaluating the effectiveness of a regression analysis and making any necessary adjustments to improve the model's performance.

Calculate Residuals in Regression Analysis

is a statistical method you can use to understand the relationship between two variables, x and y .

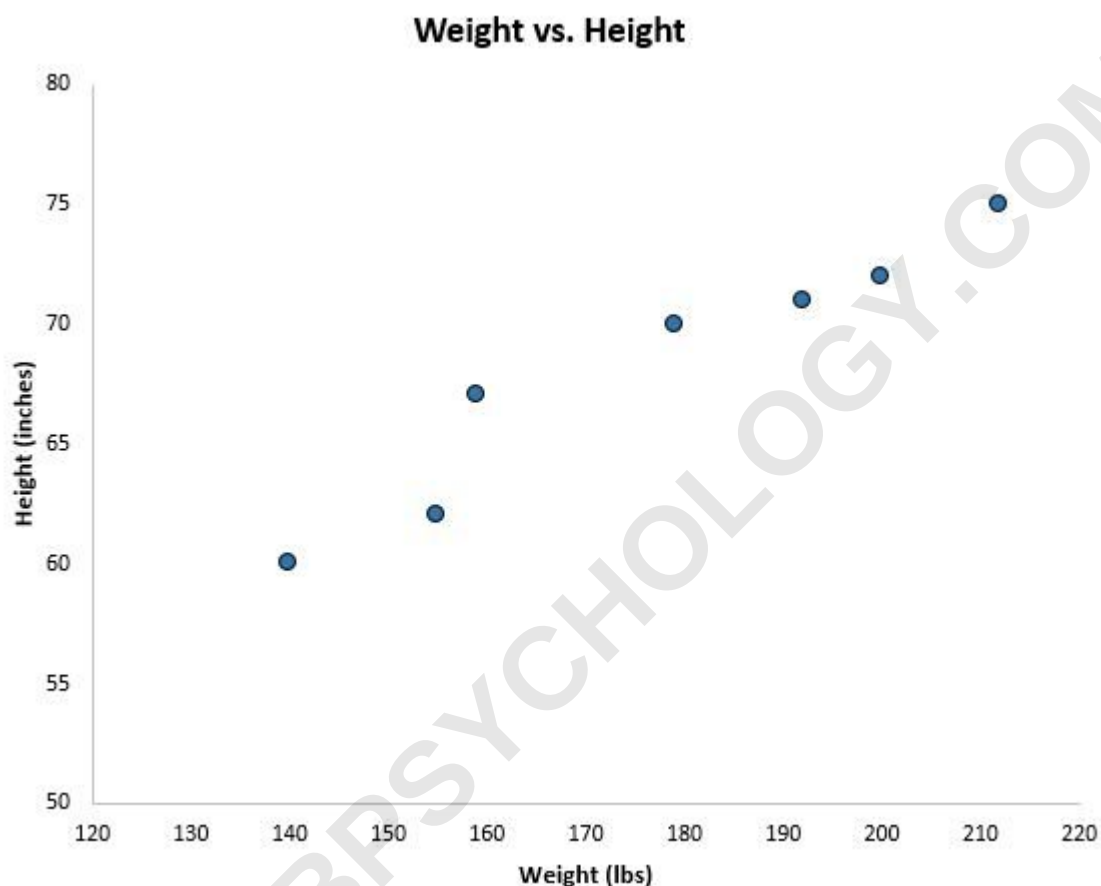
One variable, x , is known as the predictor variable. The other variable, y , is known as the .

For example, suppose we have the following dataset with the weight and height of seven individuals:

Weight (lbs)	Height (inches)
140	60
155	62
159	67
179	70
192	71
200	72
212	75

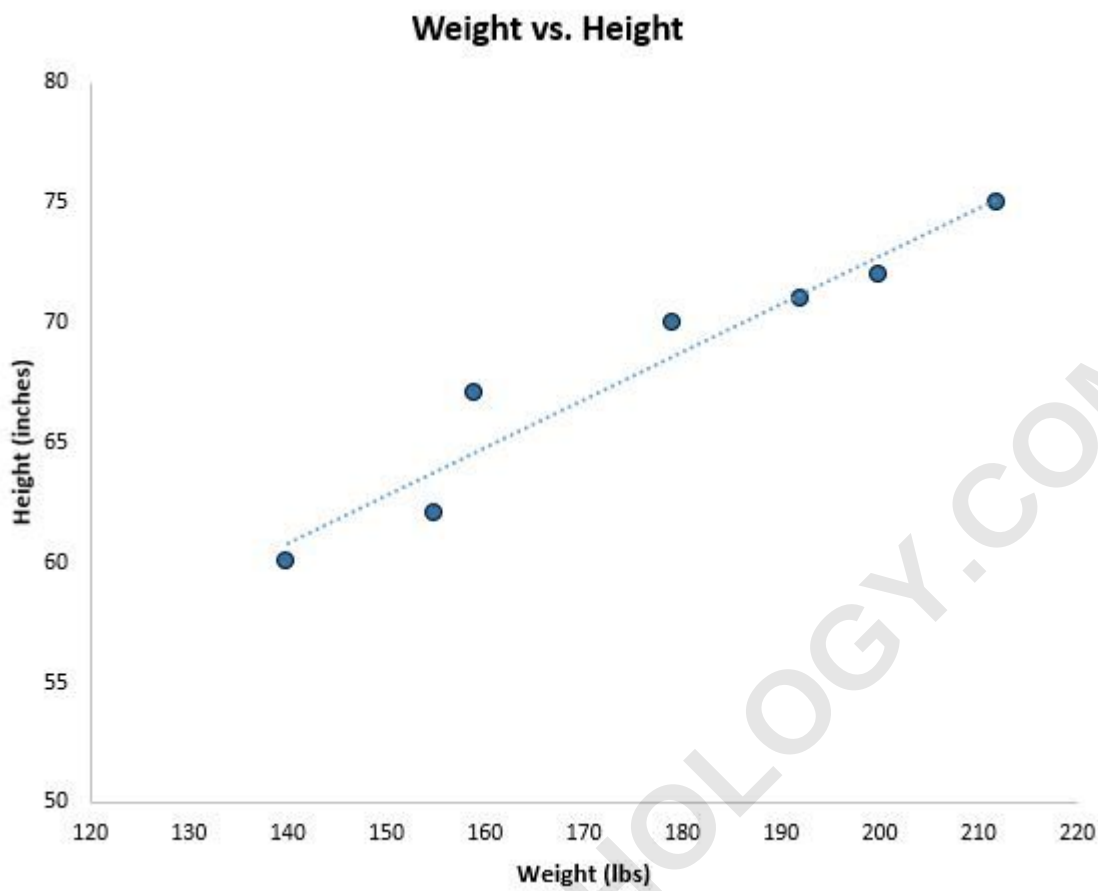
Let *weight* be the predictor variable and let *height* be the response variable.

If we graph these two variables using a , with weight on the x-axis and height on the y-axis, here's what it would look like:



From the scatterplot we can clearly see that as weight increases, height tends to increase as well, but to actually *quantify* this relationship between weight and height, we need to use linear regression.

Using linear regression, we can find the line that best "fits" our data:



The formula for this line of best fit is written as:

$$\hat{y} = b_0 + b_1x$$

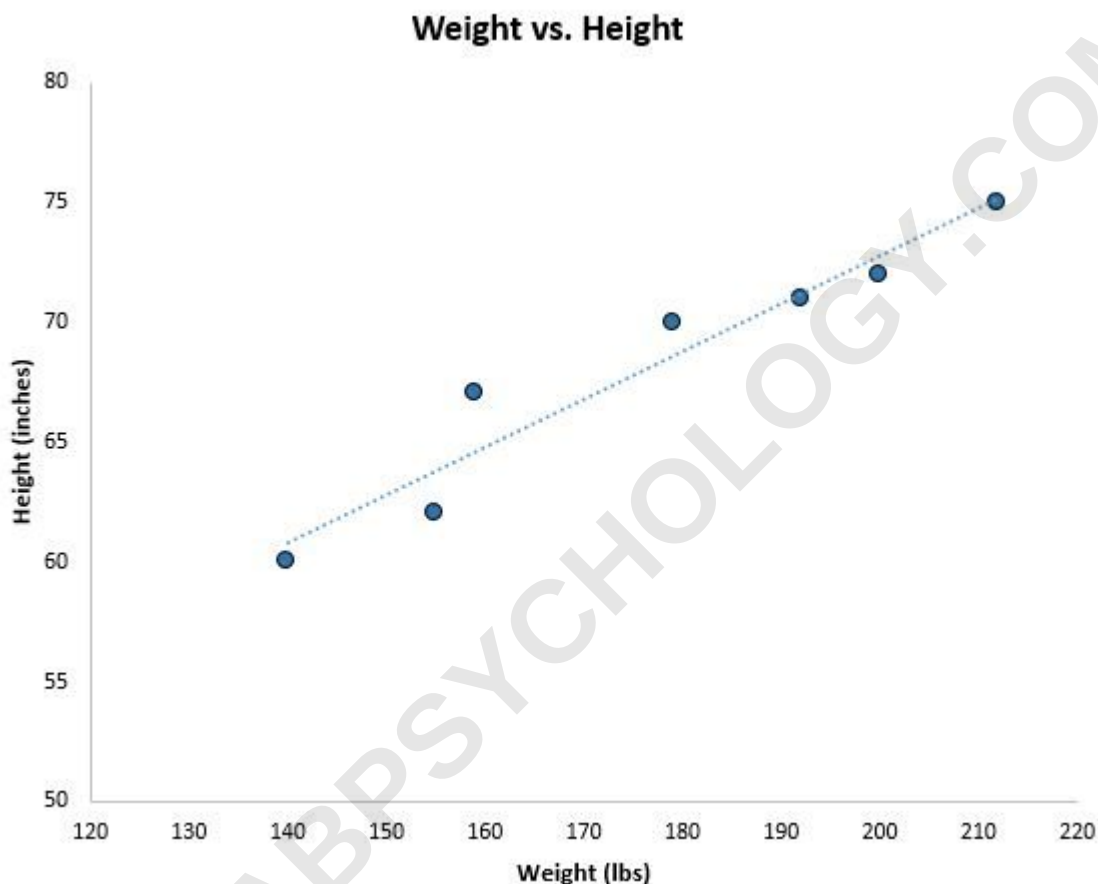
where $\hat{y}$ is the predicted value of the response variable, b_0 is the y-intercept, b_1 is the regression coefficient, and x is the value of the predictor variable.

In this example, the line of best fit is:

$$\text{height} = 32.783 + 0.2001 * (\text{weight})$$

How to Calculate Residuals

Notice that the data points in our scatterplot don't always fall exactly on the line of best fit:



This difference between the data point and the line is called the residual. For each data point, we can calculate that point's residual by taking the difference between its actual value and the predicted value from the line of best fit.

Example 1: Calculating a Residual

For example, recall the weight and height of the seven individuals in our dataset:

Weight (lbs)	Height (inches)
140	60
155	62
159	67
179	70
192	71
200	72
212	75

The first individual has a weight of 140 lbs. and a height of 60 inches.

To find out the predicted height for this individual, we can plug their weight into the line of best fit equation:

$$\text{height} = 32.783 + 0.2001 * (\text{weight})$$

Thus, the predicted height of this individual is:

$$\text{height} = 32.783 + 0.2001 * (140)$$

$$\text{height} = 60.797 \text{ inches}$$

Thus, the residual for this data point is $60 - 60.797 =$

-0.797.

Example 2: Calculating a Residual

We can use the exact same process we used above to calculate the residual for each data point. For example, let's calculate the residual for the second individual in our dataset:

Weight (lbs)	Height (inches)
140	60
155	62
159	67
179	70
192	71
200	72
212	75

The second individual has a weight of 155 lbs. and a height of 62 inches.

To find out the predicted height for this individual, we can plug their weight into the line of best fit equation:

$$\text{height} = 32.783 + 0.2001 * (\text{weight})$$

Thus, the predicted height of this individual is:

$$\text{height} = 32.783 + 0.2001 \times (155)$$

$$\text{height} = 63.7985 \text{ inches}$$

Thus, the residual for this data point is $62 - 63.7985 = -1.7985$.

Calculating All Residuals

Using the same method as the previous two examples, we can calculate the residuals for every data point:

Weight (lbs)	Height (inches)	Predicted Height	Residual
140	60	60.797	-0.797
155	62	63.7985	-1.7985
159	67	64.5989	2.4011
179	70	68.6009	1.3991
192	71	71.2022	-0.2022
200	72	72.803	-0.803
212	75	75.2042	-0.2042

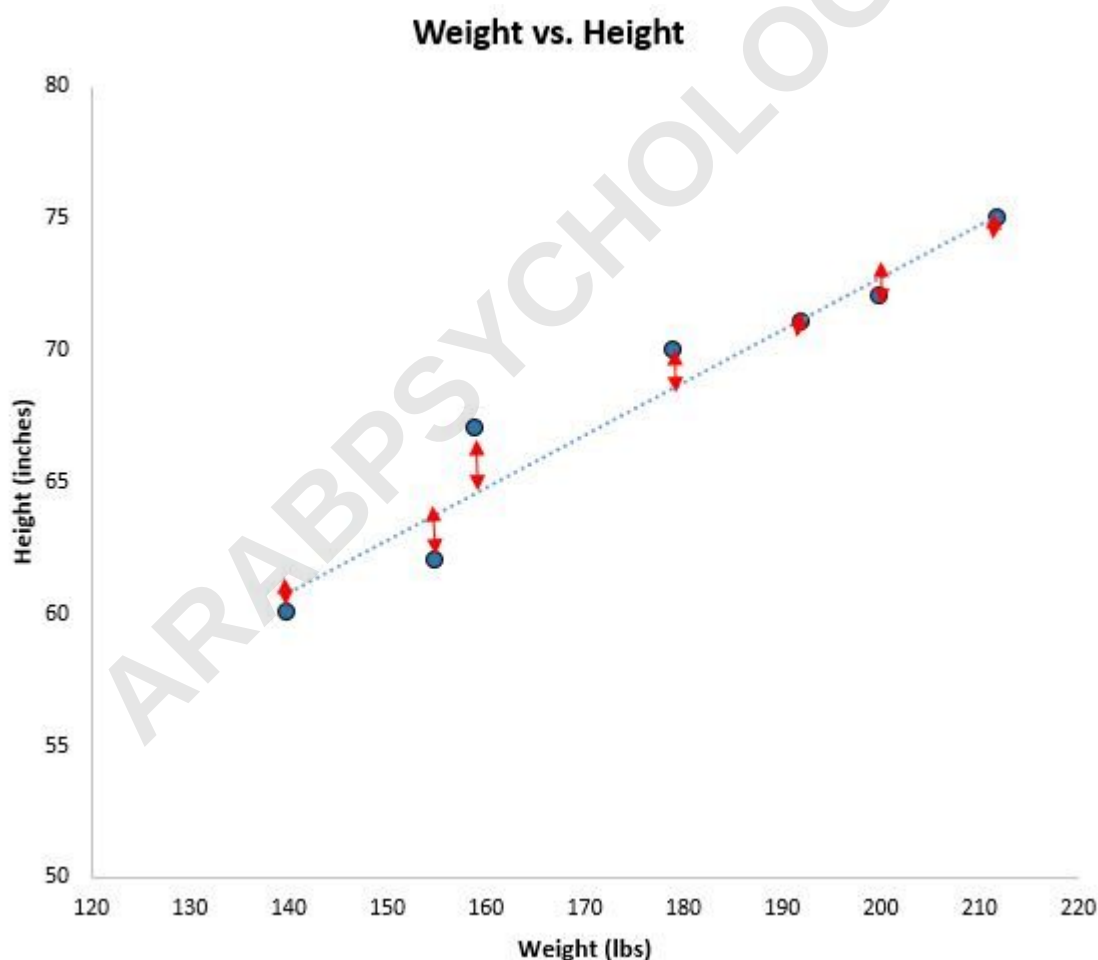
Notice that some of the residuals are positive and some are negative. If we add up all of the residuals, they will add up to zero.

This is because linear regression finds the line that minimizes the total squared residuals, which is why the line perfectly goes through the data, with some of the

data points lying above the line and some lying below the line.

Visualizing Residuals

Recall that a residual is simply the distance between the actual data value and the value predicted by the regression line of best fit. Here's what those distances look like visually on a scatterplot:



Notice that some of the residuals are larger than others.

Also, some of the residuals are positive and some are negative as we mentioned earlier.

Creating a Residual Plot

The whole point of calculating residuals is to see how well the regression line fits the data.

Larger residuals indicate that the regression line is a poor fit for the data, i.e. the actual data points do not fall close to the regression line.

Smaller residuals indicate that the regression line fits the data better, i.e. the actual data points fall close to the regression line.

One useful type of plot to visualize all of the residuals at once is a residual plot. A residual plot is a type of plot that displays the predicted values against the residual values for a regression model.

This type of plot is often used to assess whether or not a linear regression model is appropriate for a given dataset and to check for of residuals.

Check out to find out how to create a residual plot for a

simple linear regression model in Excel.

ARABPSYCHOLOGY.COM