

# Create a Residual Plot in Google Sheets

Authored by  
**Mohammed looti**

October 31, 2025

## RECOMMENDED CITATION

Mohammed looti (2025). *Create a Residual Plot in Google Sheets*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=6677>

In the realm of statistical analysis, the [residual plot](#) is an indispensable diagnostic tool. It offers a critical visual representation comparing the [predicted values](#) generated by a [regression model](#) against their corresponding error terms, known as [residual values](#). This graphical display is fundamental for thoroughly evaluating the statistical validity and overall appropriateness of the chosen model, especially when dealing with linear relationships.

The primary function of a residual plot is to confirm whether a [linear regression model](#) is the most suitable fit for a given dataset. Crucially, it helps analysts identify potential systematic errors or violations of the core [assumptions of linear regression](#). Two key assumptions examined are linearity and the requirement of [constant variance](#) (homoscedasticity). Deviations from these foundational assumptions, such as a fanning pattern in the plot, can severely compromise the reliability and inferential power of your statistical conclusions.

This comprehensive, step-by-step tutorial will guide you through the precise methodology for constructing a residual plot for a [simple linear regression](#) model directly within [Google Sheets](#). We will detail every necessary calculation, from initial data entry and determining the regression equation to the final visualization and interpretation of the diagnostic plot, enabling you to effectively assess your model quality.

## Step 1: Enter Your Dataset and Define Variables

To commence the analysis, open a new or access an existing spreadsheet document in [Google Sheets](#). The initial task involves accurately inputting the sample dataset. These values represent the paired observations of your independent variable (the [predictor variable](#),  $X$ ) and the dependent variable (the observed outcome,  $Y$ ), which form the basis for the subsequent [regression analysis](#).

It is essential to ensure meticulous accuracy during data entry, as any errors here will propagate through the entire analysis. Carefully input the following sample data into your spreadsheet, typically starting in cells A1 and B1 for the column headers:

	A	B	C	D
1	<b>x</b>	<b>y</b>		
2	8	41		
3	12	42		
4	12	39		
5	13	37		
6	14	35		
7	16	39		
8	17	45		
9	22	46		
10	24	39		
11	26	49		
12	29	55		
13	30	57		
14				
15				
16				
17				
18				

For the purposes of this illustration, Column A will strictly contain the values for our independent variable (X), and Column B will hold the corresponding values for the dependent variable (Y). This setup is critical for applying the correct formulas in the next steps.

## Step 2: Calculate the Simple Linear Regression Coefficients

Once the data is correctly structured, the next crucial step is to determine the specific equation for our [simple linear regression model](#). This model takes the form  $Y = b_0 + b_1X$ , where  $b_0$  is the intercept and  $b_1$  is the slope. These coefficients allow us to quantify the linear relationship and subsequently predict the dependent variable (Y) based on the input of the independent variable (X). [Google Sheets](#) offers powerful, built-in statistical functions that streamline the calculation of these coefficients.

We will leverage the dedicated [SLOPE](#) and [INTERCEPT](#) functions for high-precision calculation. Designate specific cells for these outputs (e.g., B15 for the slope and B16 for the intercept). In the slope cell, input the formula `=SLOPE(B2:B13, A2:A13)`, ensuring that the dependent variable range (Y) is listed first, followed by the independent variable range (X). Similarly, for the intercept cell, input `=INTERCEPT(B2:B13, A2:A13)`. Always double-check that your specified data ranges accurately capture all your X and Y observations.

	A	B	C	D
1	<b>x</b>	<b>y</b>		
2	8	41		
3	12	42		
4	12	39		
5	13	37		
6	14	35		
7	16	39		
8	17	45		
9	22	46		
10	24	39		
11	26	49		
12	29	55		
13	30	57		
14				
15	<b>Slope</b>	0.7552878554	=SLOPE(B2:B13, A2:A13)	
16	<b>Intercept</b>	29.63090069	=INTERCEPT(B2:B13, A2:A13)	
17				
18				
19				

Upon executing these functions, the specific parameter estimates for our dataset are generated. For this example, the calculations yield the following values:

Slope ( $b_1$ ) = 0.755

Intercept ( $b_0$ ) = 29.631

Consequently, the finalized equation for our [simple linear regression equation](#) is:  $\hat{Y} = 29.631 + 0.755X$ . This calculated equation is now ready to be used to generate the necessary [predicted values](#) for every observation in our sample dataset.

### Step 3: Generate Predicted Values for Each Observation

With the [regression equation](#) successfully established, the immediate next task is to calculate the [predicted values](#) (often denoted as  $\hat{Y}$ ) for every data point. These predicted values represent the model's estimate of the dependent variable based on the fitted linear relationship, essentially telling us where each point should theoretically lie on the regression line.

To perform this calculation efficiently across the dataset in [Google Sheets](#), create a new column, perhaps labeled "Predicted Y" (Column C). Navigate to cell **C2** and input a formula that integrates the calculated intercept and slope with the corresponding X-value from cell **A2**. A critical step here

is employing absolute references (using the dollar sign, e.g.,  $\$B\$16$ ) for the slope and intercept cells. This ensures that when the formula is copied down the column, the references to the coefficients remain fixed, while the reference to the independent variable (A2) changes appropriately.

**$=\$B\$16+\$B\$15*A2$**

Once the formula is correctly entered in **C2**, use the fill handle feature (the small box in the lower-right corner of the cell) to drag the formula down through the rest of column **C**. This action automatically computes the predicted Y-value for every corresponding X-value in your dataset, populating the new column with the model's estimations.

	A	B	C	D
1	<b>x</b>	<b>y</b>	<b>Predicted Value</b>	
2	8	41	35.7	
3	12	42	38.7	
4	12	39	38.7	
5	13	37	39.4	
6	14	35	40.2	
7	16	39	41.7	
8	17	45	42.5	
9	22	46	46.2	
10	24	39	47.8	
11	26	49	49.3	
12	29	55	51.5	
13	30	57	52.3	
14				
15	<b>Slope</b>	0.7552878554		
16	<b>Intercept</b>	29.63090069		
17				
18				
19				

#### Step 4: Compute the Residual Values (Errors)

The statistical foundation of the [residual plot](#) rests entirely on the calculation of the residuals themselves. A [residual](#) is formally defined as the vertical distance between the actual [observed value](#) (Y) and the [predicted value](#) (Y-hat) generated by the [simple linear regression](#) model. In essence, the residual quantifies the error of prediction for that specific data point, indicating how

much the model missed the actual observation.

Understanding these residuals is paramount because they directly measure the quality of the model's fit at an individual observation level. A residual close to zero signifies an excellent fit, whereas a large positive or negative residual suggests a significant deviation or poor fit. The required calculation is straightforward:

### Residual = Observed Value – Predicted Value

To compute the residual for the first observation, enter the following formula into cell **D2** (assuming Column D is your new "Residuals" column):

**=B2-C2**

Just as in the previous step, apply the fill handle feature to drag this formula down to the end of your dataset, populating column **D** with the computed [residual values](#) for every observation. This final column contains all the necessary data points needed to construct the diagnostic plot.

D2    fx    =B2-C2					
	A	B	C	D	E
1	<b>x</b>	<b>y</b>	<b>Predicted Value</b>	<b>Residual</b>	
2	8	41	35.7	5.3	
3	12	42	38.7	3.3	
4	12	39	38.7	0.3	
5	13	37	39.4	-2.4	
6	14	35	40.2	-5.2	
7	16	39	41.7	-2.7	
8	17	45	42.5	2.5	
9	22	46	46.2	-0.2	
10	24	39	47.8	-8.8	
11	26	49	49.3	-0.3	
12	29	55	51.5	3.5	
13	30	57	52.3	4.7	
14					
15	<b>Slope</b>	0.7552878554			
16	<b>Intercept</b>	29.63090069			
17					
18					
19					
20					

## Step 5: Visualize the Residual Plot in Google Sheets

With the calculation phase complete, the next logical step is to visualize the relationship between the independent variable and the calculated error terms. The construction of the [residual plot](#) is the key step for visually diagnosing the effectiveness and validity of the [regression model](#).

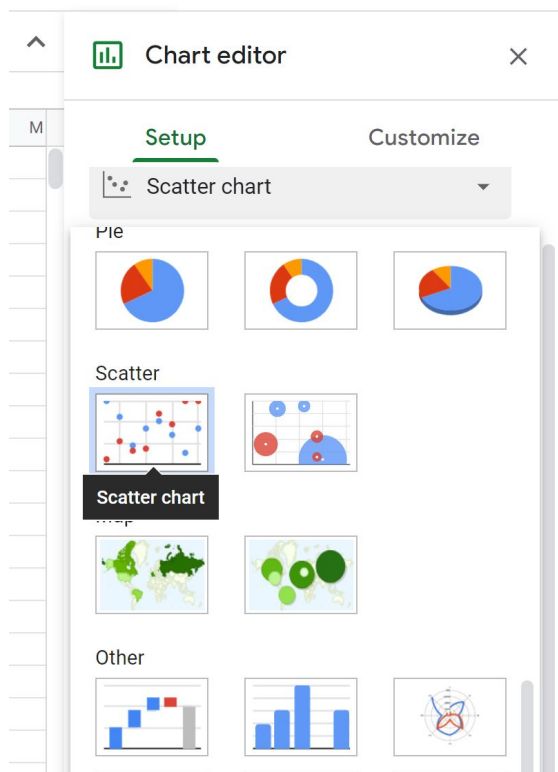
Begin by selecting the data ranges required for the plot. First, highlight the range containing your independent variable (X-values), which is **A2:A13** in this example. Next, hold down the "Ctrl" key (or "Command" on Mac) and simultaneously select the range containing the calculated residual values, **D2:D13**. Selecting these two non-contiguous ranges tells [Google Sheets](#) which data series to plot against each other.

	A	B	C	D
1	<b>x</b>	<b>y</b>	<b>Predicted Value</b>	<b>Residual</b>
2	8	41	35.7	5.3
3	12	42	38.7	3.3
4	12	39	38.7	0.3
5	13	37	39.4	-2.4
6	14	35	40.2	-5.2
7	16	39	41.7	-2.7
8	17	45	42.5	2.5
9	22	46	46.2	-0.2
10	24	39	47.8	-8.8
11	26	49	49.3	-0.3
12	29	55	51.5	3.5
13	30	57	52.3	4.7
14				
15	<b>Slope</b>	0.7552878554		
16	<b>Intercept</b>	29.63090069		
17				
18				
19				
20				

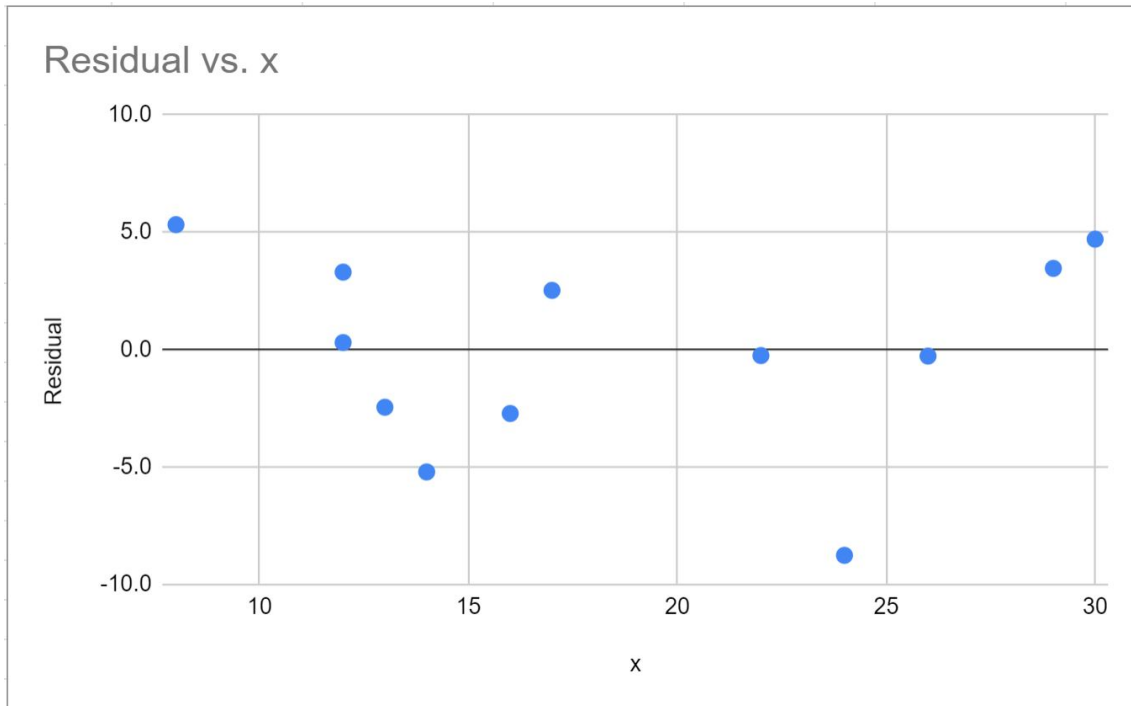
After selecting the data, navigate to the [Insert](#) menu located in the Google Sheets toolbar and select the **Chart** option from the resulting dropdown menu. This action will automatically launch the Chart editor panel on the right side of your screen, where customization occurs.

Within the **Chart editor**, it is imperative to set the Chart type correctly. Ensure that the **Scatter chart** option is selected. A scatter chart is the statistically appropriate visualization tool for

displaying the relationship between two numerical variables, allowing us to clearly see the distribution of residuals relative to the predictor variable.



Once the scatter chart type is confirmed, the residual plot will be generated instantly and displayed directly on your spreadsheet:



## Step 6: Interpretation of the Residual Plot

In the resulting plot, the horizontal x-axis represents the values of your [predictor variable](#) (X), while the vertical y-axis displays the magnitude of the corresponding [residual values](#). The primary analytical goal when examining this plot is to rigorously assess the fulfillment of key [assumptions of linear regression](#), most notably the assumption of [constant variance](#).

The assumption of [constant variance](#) mandates that the variability (spread) of the residuals must remain consistent across the entire range of the predictor variable. When this critical assumption is satisfied, the points in the residual plot should exhibit a patternless, random scattering centered closely around the horizontal line at zero. There should be no discernible structure, such as a curvature (suggesting non-linearity) or a fan shape (suggesting increasing or decreasing error variance).

In the residual plot we generated for our sample data, the plotted points demonstrate an ideal random scattering around the zero reference line. Crucially, there is an absence of any obvious patterns--no funneling, no upward or downward trends, and no clear curve. This strong visual evidence confirms that the assumption of [constant variance](#) is met for this specific dataset and [simple linear regression](#) model, thereby reinforcing the statistical reliability and validity of our overall linear analysis.

## Additional Resources for Google Sheets Statistical Analysis

Mastering both the construction and the accurate interpretation of [residual plots](#) is a fundamental skill for reliable statistical analysis. To further develop your expertise in leveraging [Google Sheets](#) for advanced statistical computations and diagnostics, we recommend exploring these related tutorials: