

How to Easily Create a Residual Plot in Google Sheets

Authored by
stats writer

November 30, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Create a Residual Plot in Google Sheets*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=102334>

A residual plot is an essential diagnostic tool in statistical analysis, particularly within the context of regression models. It provides a visual representation of the difference between the observed values in a dataset and the values predicted by the model. This difference is formally known as the residual.

The primary purpose of generating a residual plot is to assess the appropriateness of the chosen model--in our case, a simple linear regression. By visualizing how the errors (residuals) are distributed across the range of predictor variables, analysts can check critical assumptions such as linearity, independence, and the homogeneity of variance (homoscedasticity). A properly created residual plot in Google Sheets allows for quick and effective model validation, moving beyond simple metrics like R-squared.

To effectively create this plot, we must systematically execute several steps within the spreadsheet environment. This detailed guide walks through the process, from initial data entry and calculation of the regression line parameters to the final visualization and interpretation of the results. Mastering this technique ensures that your statistical conclusions drawn from the regression analysis are robust and reliable.

Understanding the Role of the Residual Plot

A **residual plot** is a type of scatter plot that displays the fitted values (or the predictor variable values) on the x-axis against the corresponding residual values on the y-axis for a regression model. This configuration is specifically designed to highlight patterns in the error term.

This type of plot is often used to assess whether or not a simple linear regression model is appropriate for a given dataset. If the relationship between variables is truly linear and the model assumptions hold, the residuals should exhibit random scatter around the zero line. Deviations from this pattern--such as a curved relationship or a fan-like spread--signal that the model may be misspecified or that key assumptions have been violated.

The following step-by-step example explains exactly how to calculate the necessary components and create a diagnostic residual plot for a simple linear regression model in Google Sheets.

Step 1: Enter the Data

The foundational step for any regression analysis is organizing your predictor (independent) and response (dependent) variables. For this example, we will work with a small dataset consisting of twelve paired observations. Ensure that the independent variable (X) is placed in column A and the dependent variable (Y) is placed in column B.

First, let's enter the following values for a dataset, labeling the columns clearly for readability and

ease of reference in later steps:

	A	B	C	D
1	x	y		
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				

It is crucial to verify the accuracy of the data entry, as any transcription errors at this stage will cascade through all subsequent calculations, leading to an inaccurate regression line and consequently, a misleading residual plot. We will use these original observation pairs to derive both the predicted values and the error terms (residuals).

Step 2: Calculate the Equation of the Regression Model

To calculate the predicted values for the linear regression, we first need to determine the slope (m) and the y-intercept (b) of the line of best fit. The linear equation is represented as $Y = mx + b$. Fortunately, Google Sheets offers built-in statistical functions that automate these complex calculations.

We will use the **SLOPE** function to find the angular coefficient (m) and the **INTERCEPT** function to find the constant term (b). The syntax for both requires specifying the 'data_y' range (column B) first, followed by the 'data_x' range (column A). Input these formulas into separate cells, perhaps B15 and B16, as shown below:

	A	B	C	D
1	x	y		
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	Slope	0.7552878554	=SLOPE(B2:B13, A2:A13)	
16	Intercept	29.63090069	=INTERCEPT(B2:B13, A2:A13)	
17				
18				
19				

Using these calculated values, where the slope is approximately 0.755 and the intercept is 29.631, we can formalize the simple linear regression equation for this specific dataset:

$$y = 29.631 + 0.755x$$

This equation represents the line that minimizes the sum of squared errors between the observed data points and the line itself. These parameters are fundamental to calculating the predicted values in the next step.

Step 3: Calculate the Predicted Values

Once we have established the regression equation, we can calculate the predicted dependent variable (Y?) for every observation in our dataset. The predicted value is what our model estimates Y to be, given the input X value. This calculation involves substituting each X value (from column A) into the derived equation: $Y = \text{Intercept} + (\text{Slope} * X)$.

To implement this efficiently in Google Sheets, we need to ensure that the references to the Slope (\$B\$15) and Intercept (\$B\$16) remain fixed when copying the formula down the column. This is achieved using absolute references (the dollar sign prefix). We'll type the following formula into cell C2, which represents the predicted Y value for the first observation:

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

We can then copy and paste this formula down to every remaining cell in column **C** (down to C13). Column C now contains the predicted values generated by our linear regression model, which are necessary for calculating the core component of our plot: the residuals.

C2		<i>fx</i>	=B\$16+B\$15*A2	
	A	B	C	D
1	x	y	Predicted Value	
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	Slope	0.7552878554		
16	Intercept	29.63090069		
17				
18				
19				

Step 4: Calculate the Residuals

The residual is the cornerstone of the residual plot. A residual represents the vertical distance between an actual observed data point and the corresponding point on the regression line. It quantifies the error of the model's prediction for that specific observation.

The calculation is straightforward but critical:

Residual = Observed value (Y) - Predicted value (Y?)

To calculate the residual for each observation in our dataset, we will create a new column (D) and type the following simple subtraction formula into cell **D2**, referencing the observed Y value in B2 and the predicted Y value in C2:

=B2-C2

We can then copy and paste this formula down to every remaining cell in column **D** (down to D13). These values--some positive (under-predicted), some negative (over-predicted)--will form the vertical axis of our diagnostic plot.

D2		$\sum x$	=B2-C2		
	A	B	C	D	E
1	x	y	Predicted Value	Residual	
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	Slope	0.7552878554			
16	Intercept	29.63090069			
17					
18					
19					
20					

Step 5: Create the Residual Plot

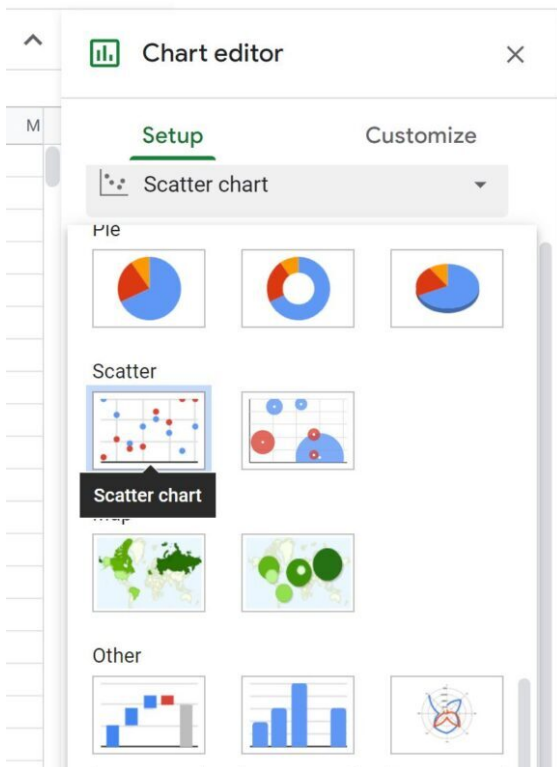
With the residuals calculated, we are ready to construct the plot. The residual plot requires two variables: the independent variable (X) from column A, and the residuals (error term) from column D.

To create the residual plot in Google Sheets, we must select these two non-contiguous columns. Highlight the values in the range **A2:A13** (X values), then hold the Ctrl key (or Command on Mac) and highlight the values in the range **D2:D13** (Residuals). This selection pairs the input variable with its corresponding error:

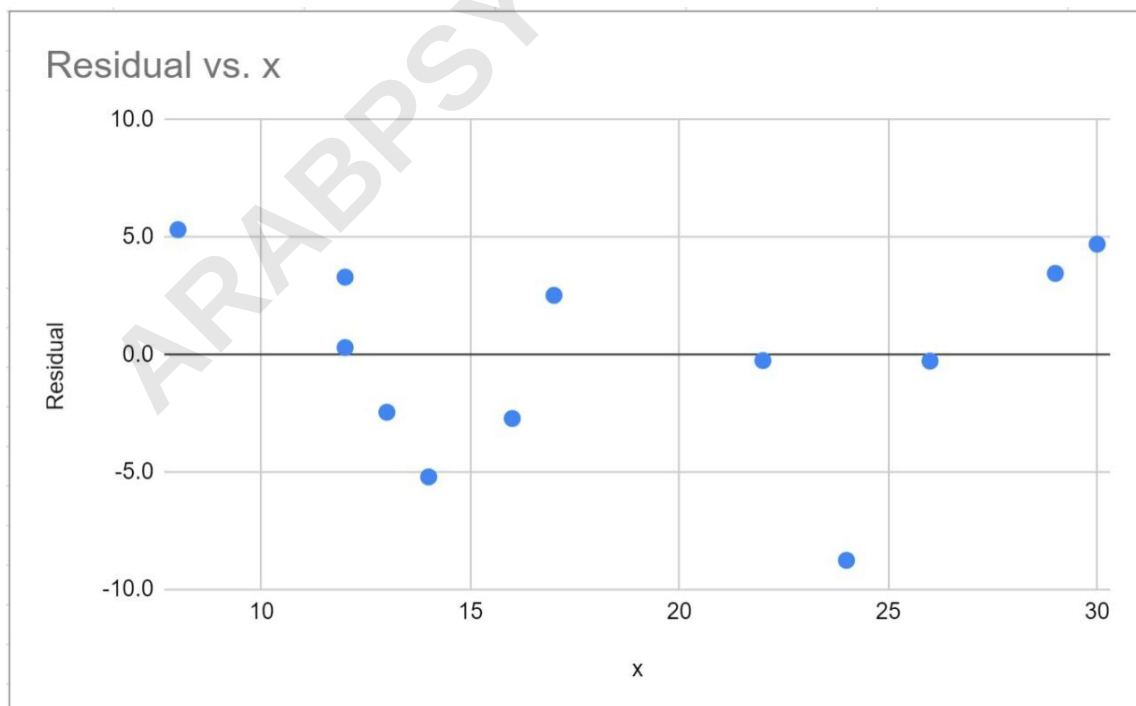
	A	B	C	D
1	x	y	Predicted Value	Residual
2	8	41	35.7	5.3
3	12	42	38.7	3.3
4	12	39	38.7	0.3
5	13	37	39.4	-2.4
6	14	35	40.2	-5.2
7	16	39	41.7	-2.7
8	17	45	42.5	2.5
9	22	46	46.2	-0.2
10	24	39	47.8	-8.8
11	26	49	49.3	-0.3
12	29	55	51.5	3.5
13	30	57	52.3	4.7
14				
15	Slope	0.7552878554		
16	Intercept	29.63090069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

Next, click the **Insert** tab, then click **Chart** in the dropdown menu. Google Sheets often defaults to a line or bar chart, which is incorrect for this purpose.

In the **Chart editor** panel that appears on the right side of the screen, navigate to the Setup tab and choose **Scatter chart** as the Chart type. This will display the individual data points necessary for pattern recognition:



Upon selecting the scatter chart type, the following residual plot will automatically appear, visually mapping the errors of your model:



Interpreting the Results of the Residual Plot

The resulting plot displays the values for the predictor variable (X) on the x-axis and the residuals (Y) on the y-axis. The zero line on the y-axis represents a perfect prediction (where the observed value equals the predicted value). The interpretation of this plot is critical for determining the validity of the simple linear regression model.

One key assumption of linear regression is that the residuals have constant variance at every level of X (known as homoscedasticity). We use a residual plot primarily to determine if this assumption is met. If the residuals are roughly evenly scattered around zero in the plot, exhibiting no discernible pattern or widening/narrowing trend, then the assumption of constant variance is generally considered met.

In our residual plot above, we can observe that the data points appear to be randomly scattered both above and below the zero line, with no obvious curvature (suggesting linearity is met) and without a funnel or cone shape (suggesting constant variance is met). Thus, based on this visual diagnostic, we would conclude that the assumption of constant variance is met for this particular regression model, reinforcing confidence in the model's fitted line.

Summary and Best Practices

Creating a residual plot is an indispensable step beyond simply generating a regression line. While the line itself provides the best fit for the data, the residual plot offers powerful diagnostic insights into the quality and statistical viability of that fit. If a plot reveals a curved pattern, it suggests a non-linear relationship exists that the linear model fails to capture. If the spread of residuals increases or decreases as X increases (heteroscedasticity), then the standard errors and confidence intervals derived from the model will be unreliable.

Always pair the quantitative results of your regression (coefficients, R-squared) with the qualitative diagnostic assessment provided by the residual plot. The visual evidence of random scatter around zero is the strongest indicator that the core assumptions required for reliable inference using ordinary least squares (OLS) regression have been satisfied. By following this meticulous process in Google Sheets, you ensure a higher standard of statistical rigor in your analysis.