

Perform a Durbin-Watson Test in Excel

Authored by
Mohammed looti

November 5, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Perform a Durbin-Watson Test in Excel*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=10421>

A cornerstone assumption of valid [regression analysis](#) is the statistical independence of error terms, often referred to as [residuals](#). This assumption strictly implies that the error observed at one point in time or sequence should not be correlated with the error observed at any other point. When this condition is violated--a common occurrence in models involving [time series data](#)--our standard error calculations become unreliable, potentially leading to inaccurate statistical inferences about the model's coefficients.

To systematically test for this critical violation, statisticians rely on the [Durbin-Watson test](#). This powerful statistical procedure is explicitly designed to detect the presence of first-order [autocorrelation](#) (or serial correlation) in the residuals generated by ordinary least squares (OLS) regression models. Identifying autocorrelation is essential because its presence compromises the efficiency of the OLS estimators, thereby violating the fundamental Gauss-Markov theorem.

The Durbin-Watson test formally assesses the independence of errors using a pair of competing hypotheses that structure the statistical decision-making process:

H0 (Null Hypothesis): There is **no first-order autocorrelation** among the residuals. The error terms are independent.

HA (Alternative Hypothesis): The residuals exhibit **first-order autocorrelation**, meaning the current error term is dependent on the error term from the previous period.

This comprehensive guide provides a precise, step-by-step methodology for executing and interpreting the Durbin-Watson test statistic using Microsoft **Excel**, a highly accessible tool for preliminary statistical modeling and data analysis.

Step 1: Organizing and Preparing the Data Set in Excel

The initial phase of any robust statistical investigation is the meticulous organization of the underlying data. Before we can generate the necessary components for the Durbin-Watson calculation, we must first structure our data set correctly within the spreadsheet environment. This critical step involves entering the observations for the response variable (Y) and all associated predictor variables (X) into adjacent, clearly labeled columns.

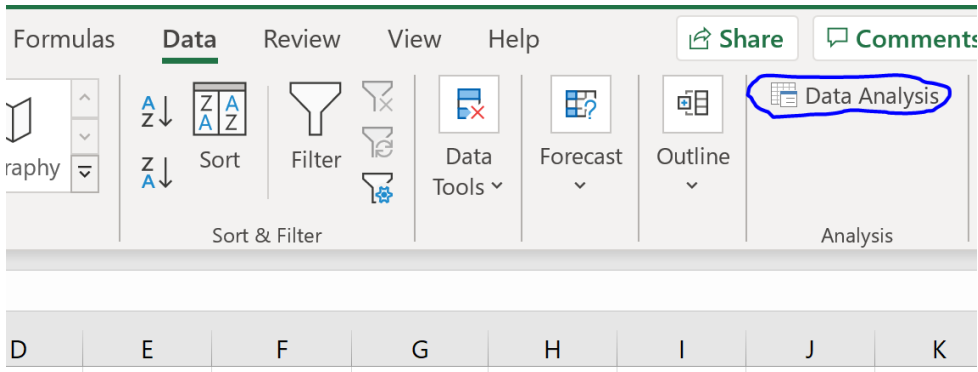
For the purposes of this tutorial, we will utilize a sample data set to construct a [multiple linear regression model](#). It is crucial to label your columns clearly--for instance, designating one column as the response variable 'y' and subsequent columns as the predictor variables, such as 'x1' and 'x2'. Proper data preparation in this manner streamlines the process of defining input ranges when executing the regression function in Excel's analysis tools.

	A	B	C	D	E	F	G
1	x1	x2	y				
2	3	4	45				
3	5	4	47				
4	5	5	49				
5	6	5	54				
6	7	6	50				
7	8	6	47				
8	8	7	50				
9	12	6	59				
10	14	5	64				
11	15	6	60				
12	16	6	58				
13	16	8	59				
14	19	7	52				
15							
16							
17							
18							
19							
20							
21							
22							

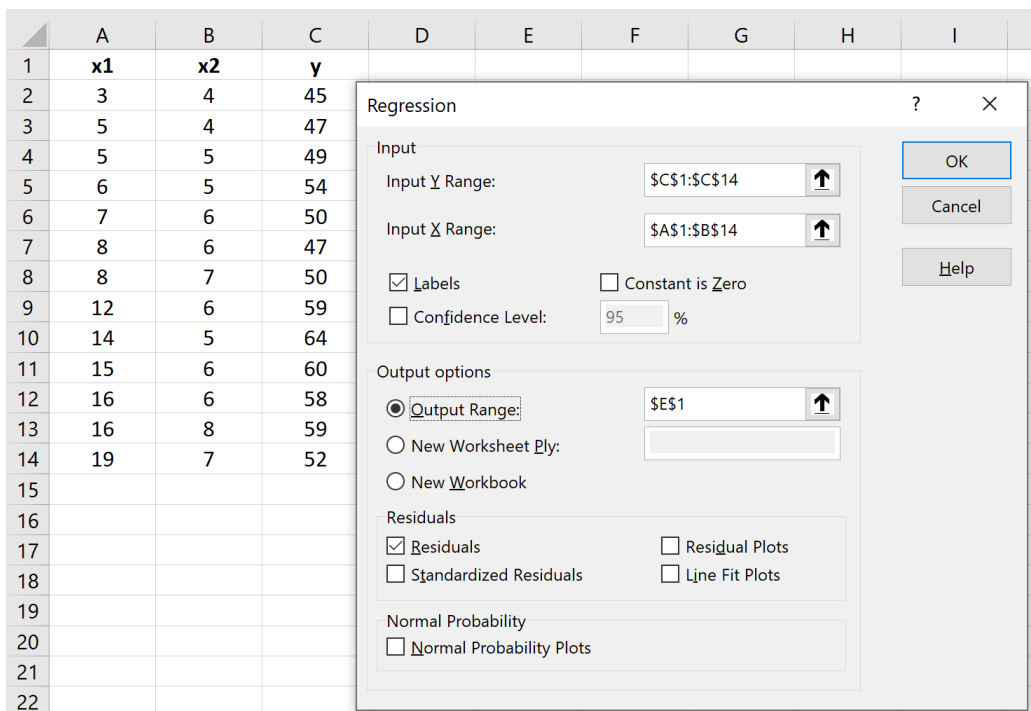
Step 2: Generating Residuals via the Regression Tool

The Durbin-Watson statistic is fundamentally built upon the sequence of error terms derived from the fitted regression model. Therefore, the next critical action is to fit the OLS regression model itself to the prepared data set. In our illustrative example, the variable 'y' serves as the **response variable**, modeled using 'x1' and 'x2' as the two primary **predictor variables**.

To access the regression capabilities in Excel, navigate to the **Data** tab in the main ribbon. Within the Analyze group, usually located on the far right, click on **Data Analysis**. If this option is not immediately visible, you must first [load the Analysis ToolPak add-in](#), as this extension provides the necessary statistical computation capabilities required for running regressions.



In the Data Analysis dialog box that appears, select the **Regression** option and click **OK**. In the ensuing setup box, accurately define the Input Y Range (the dependent variable) and the Input X Range (the independent variables). Crucially, to ensure the necessary components are calculated, navigate to the Output Options section and check the box labeled **Residuals**. This action tells Excel to output the list of numerical residuals required for the subsequent calculation step.



After clicking **OK**, Excel will generate a new sheet containing the comprehensive regression output, which includes the summary statistics, coefficient estimates, and, most importantly for our purposes, a dedicated table listing the calculated residuals for every observation point in the data set.

D	E	F	G	H	I	J	K	L	M
	SUMMARY OUTPUT								
	<i>Regression Statistics</i>								
	Multiple R	0.750703							
	R Square	0.563556							
	Adjusted R	0.476267							
	Standard E	4.367605							
	Observatio	13							
	ANOVA								
		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>ignificance F</i>			
	Regression	2	246.3172	123.1586	6.456215	0.015836			
	Residual	10	190.7597	19.07597					
	Total	12	437.0769						
		<i>Coefficients</i>	<i>andard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>ower 95.0%</i>	<i>pper 95.0%</i>
	Intercept	50.30152	6.748134	7.454138	2.17E-05	35.26574	65.3373	35.26574	65.3373
	x1	1.032967	0.323748	3.190647	0.009644	0.311611	1.754323	0.311611	1.754323
	x2	-1.31117	1.45364	-0.90199	0.388272	-4.55008	1.927748	-4.55008	1.927748
	RESIDUAL OUTPUT								
	<i>Observation</i>	<i>Predicted y</i>	<i>Residuals</i>						
	1	48.15576	-3.15576						
	2	50.2217	-3.2217						
	3	48.91053	0.089467						
	4	49.9435	4.0565						
	5	49.6653	0.334698						
	6	50.69827	-3.69827						
	7	49.3871	0.612897						
	8	54.83014	4.169863						
	9	58.20724	5.792764						
	10	57.92904	2.070962						
	11	58.962	-0.962						
	12	56.33967	2.660326						
	13	60.74974	-8.74974						

Step 3: Calculating the Durbin-Watson (d) Test Statistic

The Durbin-Watson statistic, denoted by the symbol d , provides a precise measure of the relationship between successive error terms. The formal mathematical definition of this statistic is given by the ratio of the sum of the squared differences between consecutive residuals to the total sum of the squared residuals:

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$$

The variables used in this calculation are defined as follows:

T: Represents the **Total number of observations** utilized in the fitted regression model.

et: Denotes the **tth residual** corresponding to the current observation in the time series or sequence.

et-1: Represents the residual from the immediately preceding observation (t-1), highlighting the sequential dependence.

To compute d efficiently in Excel, we must utilize the generated residual column. We calculate the squared difference between each residual and its preceding residual, sum these differences (the numerator), and then divide this total by the sum of all squared residuals (the denominator). The required spreadsheet setup, including these intermediate calculations, is illustrated below:

RESIDUAL OUTPUT		
Observation	Predicted y	Residuals
1	48.15576	-3.15576
2	50.2217	-3.2217
3	48.91053	0.089467
4	49.9435	4.0565
5	49.6653	0.334698
6	50.69827	-3.69827
7	49.3871	0.612897
8	54.83014	4.169863
9	58.20724	5.792764
10	57.92904	2.070962
11	58.962	-0.962
12	56.33967	2.660326
13	60.74974	-8.74974

d	1.347536	=SUMXMY2(G27:G38, G26:G37) / SUMSQ(G26:G38)
-----	----------	---

Following the application of the formula to the summed columns, the Durbin-Watson test statistic for this specific data set is determined to be **1.3475**. This calculated value is now ready for comparison against the critical bounds to draw a statistical conclusion.

Step 4: Interpreting the Durbin-Watson Critical Values

Interpreting the Durbin-Watson statistic differs from standard hypothesis tests because it requires consulting specialized [statistical tables](#) rather than relying on a simple p-value. The Durbin-Watson table provides two critical thresholds: a lower critical value (d_L) and an upper critical value (d_U). These values define three distinct zones--rejection, non-rejection, and inconclusive--and are determined by the significance level (α), the total number of observations (n), and the number of independent variables (k).

For our continuing analysis, we assume standard parameters: a significance level of $\alpha = 0.05$, a total of $n = 13$ observations, and $k = 2$ independent variables (x_1 and x_2). Consulting the appropriate Durbin-Watson table yields the following critical bounds for this specific model configuration:

Lower critical value (dL): **0.86**

Upper critical value (dU): **1.56**

The statistical decision rule is based on where the calculated test statistic (d) falls relative to these critical bounds:

If $d < dL$, we reject H_0 and conclude there is strong evidence of **positive autocorrelation**.

If $d > dU$, we fail to reject H_0 , suggesting **no significant autocorrelation** is present.

If d falls between dL and dU , the test is **inconclusive**, meaning we cannot definitively state whether autocorrelation exists.

Since our calculated test statistic of **1.3475** falls directly within the inconclusive range ($0.86 < 1.3475 < 1.56$), we conclude that we **do not have sufficient statistical evidence** to reject the null hypothesis of the Durbin-Watson test at the 5% significance level. We cannot conclusively determine if autocorrelation is present in this data set.

Step 5: Addressing Autocorrelation When Detected

If the Durbin-Watson test yields a result that leads to the rejection of the null hypothesis--thereby confirming the presence of significant [autocorrelation](#) in the [residuals](#)--it indicates a severe violation of OLS assumptions. This violation results in inefficient coefficient estimates and biased standard errors, making hypothesis testing unreliable. Corrective actions must be taken, and the choice of action depends heavily on the specific nature of the serial correlation observed:

For Positive Serial Correlation: This is the most frequently encountered issue. Effective solutions include modifying the model structure by incorporating **lagged values** of the dependent variable (Y) and/or independent variables (X) as new predictors. Alternatively, specialized estimation techniques, such as the Prais-Winsten or Cochrane-Orcutt procedures, can be used to transform the data, though these usually require dedicated statistical software packages.

For Negative Serial Correlation: Negative autocorrelation is generally less common. When observed, it may sometimes be an unintended consequence of **over-differencing** the time series data. If the series has been subjected to differencing more times than required to achieve stationarity, spurious negative correlation can be introduced. Analysts should carefully review their data transformation steps.

For Seasonal Correlation: In data exhibiting clear seasonal patterns (e.g., quarterly or monthly financial data), the correlation might occur at fixed intervals rather than sequentially. In these

cases, adding specific **seasonal dummy variables** to the regression model is advisable to capture and control for cyclical effects that are not explained by the primary explanatory variables.

Additional Resources for Advanced Econometrics

While Excel serves as an excellent platform for calculating the Durbin-Watson statistic, more complex applications in [econometrics](#) often necessitate dedicated software. Such tools are required for implementing advanced correction methods like generalized least squares (GLS) or computing robust standard errors, such as those derived from the Newey-West procedure. For researchers and students interested in the theoretical foundations and advanced applications of testing and correcting serial correlation, consulting specialized textbooks on time series analysis and econometrics is highly recommended to ensure the robustness of model inferences.

Mastering the methodology for detecting and addressing autocorrelation is an indispensable skill for validating and ensuring the efficiency of any OLS regression model applied to sequential data.