

# How to perform a Durbin-Watson Test in Python

Authored by  
**stats writer**

December 24, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to perform a Durbin-Watson Test in Python*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=108704>

The Durbin-Watson Test is a crucial inferential statistic used in econometric and statistical analysis to detect the presence of serial correlation, or autocorrelation, within the residuals of a fitted regression model. Detecting this phenomenon is essential because violated assumptions can lead to inefficient coefficient estimates and misleading hypothesis tests. In the Python data science ecosystem, this test is easily accessible through the statsmodels library. Specifically, the function `statsmodels.stats.stattools.durbin_watson()` allows practitioners to input the sequence of residuals from their model. The function returns a single test statistic, which is bounded between 0 and 4. A value approximately equal to 2 signifies no autocorrelation, while values deviating toward 0 or 4 suggest significant positive and negative autocorrelation, respectively. Understanding this statistic is foundational to validating the robustness of any time-series or regression analysis.

## Understanding the Importance of Residual Analysis

One of the fundamental assumptions underpinning ordinary least squares (OLS) linear regression models is the independence of observation errors. This means that the error term, or residual, for one observation should not be correlated with the error term of any other observation in the sample. If this assumption is violated, particularly in time-series data where errors tend to persist over time, the resulting model parameters may still be unbiased, but the standard errors associated with those estimates will be incorrect. Incorrect standard errors lead directly to flawed confidence intervals and unreliable P-values, potentially causing researchers to make incorrect conclusions about the significance of their predictor variables.

In essence, the independence of residuals implies that any pattern or structure present in the data has been fully captured by the explanatory variables in the model. If the residuals themselves exhibit patterns--such as consecutive positive errors followed by consecutive negative errors--it signals that important information is missing from the model specification or that the modeling technique is inappropriate for the data structure. The presence of autocorrelation means that past errors influence future errors, systematically biasing the estimation of uncertainty. Thus, rigorously testing for independence is a required step in model diagnostics, especially when dealing with data points ordered sequentially in time or space.

## What is the Durbin-Watson Test?

The Durbin-Watson Test is specifically designed to detect first-order serial correlation. First-order correlation means that the residual at time  $t$  is related only to the residual at time  $t-1$ . The test statistic is calculated based on the differences between successive residuals. By comparing the sum of the squared differences of successive residuals to the sum of the squared residuals themselves, the test provides a measure of how clustered or dispersed the errors are relative to each other. When successive residuals are similar (suggesting positive autocorrelation), the

differences are small, leading to a low test statistic. Conversely, when successive residuals alternate signs sharply (suggesting negative autocorrelation), the differences are large, leading to a high test statistic.

The calculation of the Durbin-Watson Test statistic (often denoted as  $d$ ) is approximately equal to 2 multiplied by 1 minus the sample autocorrelation of the residuals ( $r$ ). Specifically, the relationship is  $d \approx 2(1-r)$ . Since the correlation coefficient  $r$  ranges from -1 (perfect negative correlation) to +1 (perfect positive correlation), the test statistic  $d$  is necessarily constrained to the range of 0 to 4. This simple relationship provides an intuitive understanding of the statistic's limits and interpretation, linking it directly back to the underlying correlation structure of the model errors.

## Core Assumptions and Hypotheses

The application of the Durbin-Watson Test is fundamentally aimed at determining if the assumption of independent residuals is met for a given regression model. If this independence holds true, the model is considered more reliable for inference. The test employs a formal framework of hypothesis testing to reach a conclusion regarding the presence of serial correlation:

The test uses the following hypotheses:

**H0 (Null Hypothesis):** There is no serial correlation among the residuals. This is the desired outcome, confirming the independence assumption of the OLS model.

**HA (Alternative Hypothesis):** The residuals are autocorrelated (either positively or negatively). If the data supports the alternative hypothesis, remedial action must be taken to correct the model specification or estimation technique.

## Interpreting the Durbin-Watson Statistic

The value of the Durbin-Watson statistic provides a direct indication of the nature and strength of the serial correlation observed in the residuals. Since the statistic is bound between 0 and 4, its proximity to these bounds determines the diagnostic outcome. Interpreting this value is critical for determining the next steps in the modeling process. The primary interpretations are as follows:

A test statistic of **2** indicates no serial correlation. This is the ideal outcome, suggesting that  $r$  (the sample autocorrelation) is approximately zero, and the errors are independent.

The closer the test statistic is to **0**, the more evidence there is of positive serial correlation. Positive autocorrelation typically occurs when errors persist in the same direction--a positive error tends to be followed by another positive error. This often signals that the model is under-specified or is missing important trend variables.

The closer the test statistic is to **4**, the more evidence there is of negative serial correlation. Negative autocorrelation is characterized by errors that alternate signs rapidly (positive followed by

negative, and so on). While less common than positive correlation, it often indicates potential issues like overdifferencing or improper smoothing techniques applied to the data.

While statistical tables are traditionally used to determine critical bounds ( $d_L$  and  $d_U$ ) for rigorous hypothesis testing, a general rule of thumb is often applied in preliminary analysis. Test statistic values falling within the range of **1.5 and 2.5** are generally considered acceptable, suggesting that autocorrelation is not a severe problem. However, values significantly outside this range warrant further investigation and likely indicate model misspecification or a fundamental violation of the OLS assumptions, requiring adjustment to the regression model structure.

## Step-by-Step Example: Performing the Durbin-Watson Test in Python

To demonstrate the practical application of the Durbin-Watson test, we will use Python and the powerful statsmodels library. We first need a dataset and a fitted OLS model from which we can extract the residuals. Consider a hypothetical dataset describing the attributes of 10 basketball players, which, although not purely time-series, serves as an excellent illustration of the necessary calculation steps:

```
import numpy as np
import pandas as pd

# Create the sample dataset representing player statistics
df = pd.DataFrame({'rating': ,
'points': ,
'assists': ,
'rebounds': })

# Display the created dataset structure
df

rating points assists rebounds
0 90 25 5 11
1 85 20 7 8
2 82 14 7 10
3 88 16 8 6
4 94 27 5 6
5 90 20 7 9
6 76 12 6 6
7 75 15 9 10
8 87 14 9 10
9 86 19 5 7
```

Next, we must fit a multiple linear regression model. We will designate *rating* as the response variable and use *points*, *assists*, and *rebounds* as the predictor variables. The OLS function from the `statsmodels.formula.api` module provides a convenient method for model fitting and subsequent residual extraction:

```
from statsmodels.formula.api import ols
```

```
# Fit the multiple linear regression model
```

```
model = ols('rating ~ points + assists + rebounds', data=df).fit()
```

```
# View the model summary (includes initial Durbin-Watson output, though we calculate it separately below)
```

```
print(model.summary())
```

With the model fitted, we can access the residuals through the `model.resid` attribute. The final step involves importing the necessary function from `statsmodels` and passing the residual array to calculate the test statistic. This single function call yields the required diagnostic measure to determine if the OLS assumption of independent errors holds true for our specific model fit:

```
from statsmodels.stats.stattools import durbin_watson
```

```
# Perform Durbin-Watson test on the model residuals
```

```
durbin_watson(model.resid)
```

```
2.392
```

## Analyzing the Results and Model Validation

The execution of the Durbin-Watson function returns a statistic of **2.392**. We must now interpret this value in the context of our model validity. As established earlier, the ideal value for a model with no serial correlation is 2. Values marginally above 2 suggest a slight negative serial correlation, while values marginally below 2 suggest a slight positive serial correlation.

Given the standard empirical guideline, which suggests that values between 1.5 and 2.5 are generally acceptable, the calculated test statistic of 2.392 falls comfortably within this normal range. Consequently, based on the Durbin-Watson Test, we fail to reject the Null hypothesis ( $H_0$ ): there is no statistically significant first-order autocorrelation present in the residuals of this regression model. This conclusion lends support to the reliability of the standard error estimates derived from the OLS fit, suggesting the model is robust regarding this specific assumption violation.

It is important to remember that while the Durbin-Watson statistic is a powerful tool, it strictly tests for first-order serial correlation. If higher-order autocorrelation (e.g., correlation between residuals at  $t$  and  $t-2$ ) is suspected, other tests like the Breusch-Godfrey test might be more appropriate. However, for most standard linear regression models, the Durbin-Watson test provides a necessary and usually sufficient initial diagnostic check.

## Strategies for Addressing Autocorrelation

If the Durbin-Watson test statistic had fallen outside the acceptable range (e.g., below 1.0 or above 3.0), indicating a significant presence of serial correlation, model modification would be necessary to ensure valid inference. Handling autocorrelation requires understanding whether the correlation stems from model misspecification or if it is an inherent property of the time-series process being modeled.

**Addressing Positive Serial Correlation:** For significant positive serial correlation (test statistic close to 0), the most common approach is to reconsider the model specification. This often involves introducing lags of the dependent variable and/or the independent variables into the model. Including these lagged terms (e.g.,  $Y_{t-1}$  or  $X_{t-1}$ ) helps absorb the temporal dependence that was previously left in the residual term. Alternatively, time-series specific models, such as ARIMA or ARMA models, might be necessary, moving beyond the scope of simple OLS.

**Addressing Negative Serial Correlation:** If the test statistic is close to 4, indicating negative serial correlation, the primary concern is often related to overdifferencing. Differencing is a common technique used to achieve stationarity in time series data, but differencing a variable that is already stationary, or differencing too many times, can introduce negative correlation into the error structure. Analysts should check the order of differencing applied to all variables and potentially reduce it to see if the negative correlation dissipates.

**Utilizing Robust Estimation Methods:** When autocorrelation is present, but model specification cannot be easily changed, analysts often turn to robust standard error estimators. Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors, such as those proposed by Newey and West, adjust the standard error calculations to account for the correlated error structure. While this does not correct the potential inefficiency of the OLS coefficients, it ensures that hypothesis tests and confidence intervals are accurate despite the non-independent residuals. Libraries like [statsmodels](#) fully support HAC estimators.

**Handling Seasonal Correlation:** In datasets with inherent seasonality (e.g., quarterly economic data), autocorrelation might occur at fixed intervals (e.g.,  $e_t$  is correlated with  $e_{t-4}$ ). A targeted approach involves including seasonal dummy variables (indicators for specific months or quarters) or using seasonal lags of the dependent or independent variables to capture these periodic effects directly within the regression framework.

The Durbin-Watson Test remains a cornerstone diagnostic tool in regression analysis, particularly when working with sequential data. Mastery of its implementation in Python, alongside a deep understanding of its interpretation and subsequent corrective strategies, is essential for any professional conducting rigorous statistical modeling.

ARABPSYCHOLOGY.COM