

How to Easily Perform a KPSS Unit Root Test in R

Authored by
stats writer

December 1, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Perform a KPSS Unit Root Test in R*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=102950>

Analyzing time series data is a foundational element of econometrics and predictive modeling. A crucial first step in any robust analysis is determining whether the series is stationary. Non-stationary time series can lead to spurious regression results and unreliable forecasts, necessitating formal statistical testing before model development.

Among the various unit root tests available, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test provides a unique perspective by testing the null hypothesis of stationarity, contrasting with more common tests like the Augmented Dickey-Fuller (ADF) test, which assume non-stationarity as the null. This inversion of hypotheses makes the **KPSS test** an essential tool for comprehensive stationarity assessment, especially when dealing with data that exhibits potential trend or level shifts.

To effectively perform the **KPSS test** within the R statistical environment, analysts rely on specialized packages. Specifically, the tseries package provides the necessary functionality through the `kpss.test()` function. This function simplifies the process, taking the time series data as input and returning key output metrics, including the test statistic, the associated p-value, and necessary critical values for manual comparison, allowing for immediate assessment of whether the data is **trend stationary** or requires differencing.

The Core Concept of Stationarity in Time Series

Stationarity is a critical property of a time series where the statistical properties, such as the mean, variance, and autocorrelation, remain constant over time. When a series is stationary, it allows us to reliably model the data using established statistical techniques like ARIMA, as the relationships observed in the past are assumed to hold true in the future. Conversely, non-stationary data requires transformation, typically through differencing, before meaningful modeling can occur.

The **KPSS test** specifically focuses on assessing two common forms of stationarity: level stationarity (where the mean is constant) and **trend stationarity** (where fluctuations occur around a deterministic trend line). Understanding which type of non-stationarity is present dictates the appropriate corrective measures. A series that is trend stationary can often be made stationary by simply detrending, whereas a series with unit roots (difference stationary) requires differencing.

The choice of stationarity test is crucial. While tests like ADF and PP (Phillips-Perron) operate under the assumption that a unit root exists (non-stationarity), the KPSS test reverses this logic. By having stationarity as the default assumption (the null hypothesis), the KPSS test provides robust evidence when non-stationarity is indeed detected, helping researchers avoid mistakenly concluding that a series is stationary when it is not.

Defining the KPSS Test Hypotheses and Interpretation

The unique feature of the KPSS test lies in the structure of its hypotheses. Unlike unit root tests, the KPSS framework assumes the series is stable unless sufficient evidence suggests otherwise. This approach is highly useful when the goal is to confirm stability rather than to search for instability, providing complementary information to other tests.

The formal hypotheses are defined as follows:

H0 (Null Hypothesis): The time series is trend stationary.

HA (Alternative Hypothesis): The time series is *not* trend stationary (it contains a unit root).

The decision to reject or fail to reject the null hypothesis hinges entirely on the relationship between the calculated **test statistic** and the critical values, or more commonly, the resulting p-value. If the calculated test statistic exceeds the critical value at a chosen significance level (or if the p-value falls below that level), the evidence favors the alternative hypothesis, suggesting the series is non-stationary.

In practice, analysts typically select a standard significance level, denoted as alpha (e.g., $\alpha = 0.05$). If the computed **p-value** is less than this predetermined significance level, we confidently **reject the null hypothesis** (H0). Rejecting the null hypothesis means we conclude that the time series is *not* trend stationary and likely contains a unit root or is difference stationary. Conversely, if the p-value is greater than or equal to the significance level, we **fail to reject the null hypothesis**, thereby supporting the conclusion that the time series is **trend stationary**.

Setting Up R for the KPSS Analysis

To execute the **KPSS test**, the first step involves ensuring the necessary packages are installed and loaded into the **R** working environment. The core functionality for this specific test is housed within the `tseries` package, which is a collection of functions dedicated to time series analysis and computational finance. Installation is a one-time process, but the library must be called upon every time a new R session is initiated.

Once the package is available, the primary function used is `kpss.test()`. This versatile function requires at least one argument: the time series data object itself. However, it also allows for optional arguments, notably the `null` parameter, which specifies the type of stationarity being tested. By default, `kpss.test()` often tests for level stationarity (stationarity around a constant mean). For comprehensive analysis, especially in financial and economic datasets, it is often critical to specify `null="Trend"` to test for stationarity around a deterministic trend line, aligning precisely with the definition provided in the hypotheses above.

It is important to remember that R operates on objects, and the time series data must be correctly formatted--ideally as a numeric vector or a dedicated time series object (like `ts`). The flexibility of R allows for easy data simulation and transformation, facilitating the demonstration of the test with both manufactured and real-world datasets, as shown in the subsequent examples.

Example 1: Testing Stationary Data

For our initial demonstration, we will simulate a dataset that inherently possesses stationary characteristics. This is often achieved by generating observations randomly from a distribution, such as a standard normal distribution, where the mean and variance are stable over time. This approach allows us to confirm that the KPSS test correctly identifies stationary data when it is truly present, thereby validating its functionality.

We begin by setting a seed for reproducibility, ensuring that anyone replicating this analysis generates the exact same sequence of pseudo-random numbers. We then use the `rnorm()` function to create 100 observations drawn from a normal distribution. Subsequently, plotting this raw data visually confirms the series fluctuates around a constant mean without any noticeable long-term upward or downward trend, providing preliminary evidence of stationarity.

#make this example reproducible

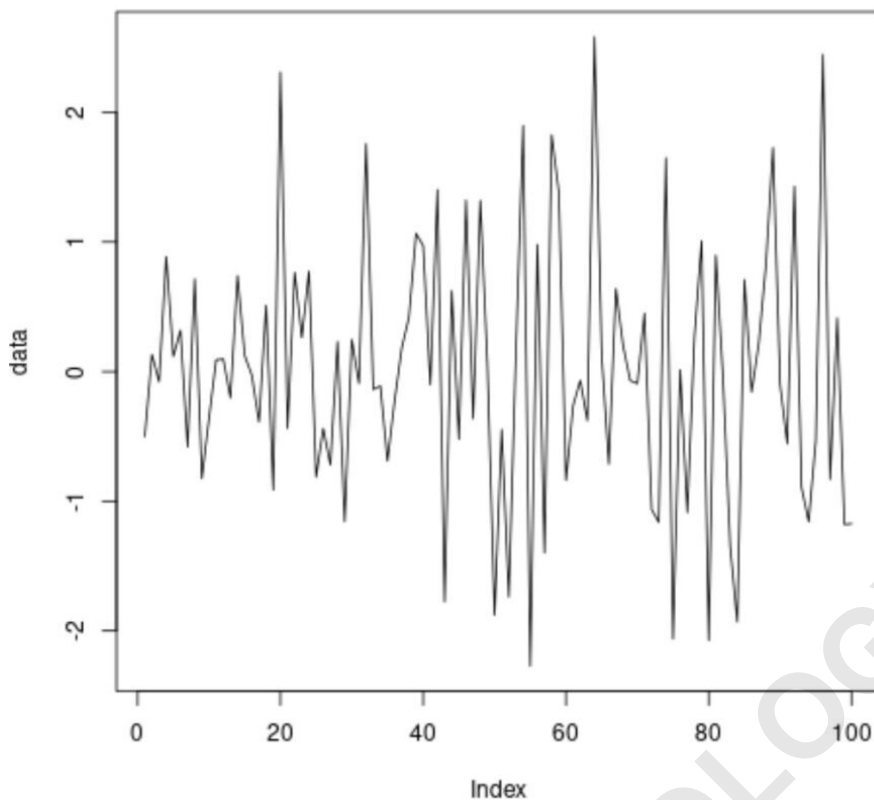
set.seed(100)

#create time series data (randomly generated, expected to be stationary)

data<-rnorm(100)

#plot time series data as line plot

plot(data, type='l')



Using the generated data, we execute the `kpss.test()` function, explicitly instructing R to test the null hypothesis of **Trend Stationarity** using the `null="Trend"` argument. This action triggers the calculation of the KPSS test statistic based on the cumulative sum of residuals from a regression of the series on a constant and a time trend. The output then provides the necessary metrics for statistical inference.

library(tseries)

```
#perform KPSS test  
kpss.test(data, null="Trend")
```

KPSS Test for Trend Stationarity

```
data: data  
KPSS Trend = 0.034563, Truncation lag parameter = 4, p-value = 0.1
```

Warning message:

```
In kpss.test(data, null = "Trend") : p-value greater than printed p-value
```

The results show a calculated KPSS Trend statistic of 0.034563. Crucially, the reported p-value is **0.1**. Considering our significance level (α) is typically set at 0.05, the p-value (0.1) is significantly

greater than 0.05. According to the decision rule, because the p-value is not less than the significance level, we **fail to reject the null hypothesis**. This outcome strongly supports the conclusion that the simulated time series is indeed **trend stationary**, confirming our initial visual assessment of the randomly generated data.

A crucial detail to note in this output is the accompanying warning message: `p-value greater than printed p-value`. The `kpss.test()` function in the `tseries` package has an inherent limitation where the highest p-value it reports is 0.1. In cases where the true p-value is, for instance, 0.5 or 0.9, the output will still display 0.1, alongside this warning. Therefore, seeing a p-value of 0.1 in the context of this warning confirms that the evidence for stationarity is very strong, as the true p-value is even higher than the reported ceiling value.

Example 2: Analyzing Non-Stationary Data

To provide a complete understanding of the KPSS test, it is essential to examine its behavior when applied to a non-stationary time series. Non-stationary data, often exhibiting clear trends or structural breaks, should result in the rejection of the null hypothesis. For this example, we will manually define a small dataset that clearly demonstrates a strong upward trajectory, simulating a basic non-stationary process that likely contains a unit root.

This deliberately constructed series shows a general tendency to increase over time, implying that the mean is not constant--a classic violation of the stationarity assumption. Visual inspection of the resulting line plot immediately suggests that the statistical properties are evolving, making it highly probable that the KPSS test will correctly flag the series as non-stationary.

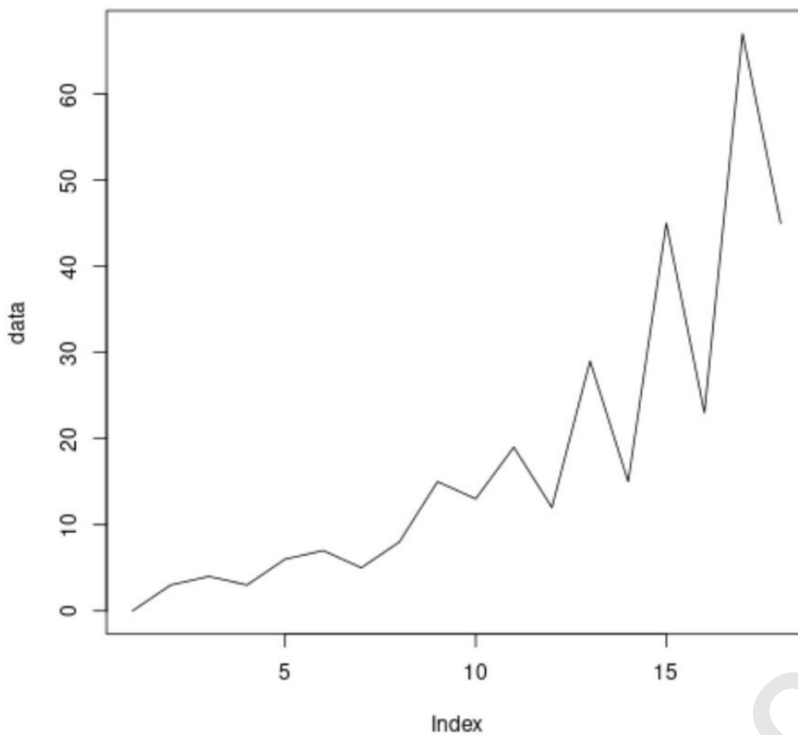
#make this example reproducible

```
#create time series data (manually defined, expected to be non-stationary)
```

```
data <-c(0, 3, 4, 3, 6, 7, 5, 8, 15, 13, 19, 12, 29, 15, 45, 23, 67, 45)
```

```
#plot time series data as line plot
```

```
plot(data, type='l')
```



Once again, we apply the `kpss.test()` function, ensuring that the `null="Trend"` parameter is specified to test for stationarity around a deterministic trend. Given the visible upward movement in the plot, we anticipate a small p-value, indicating strong statistical evidence against the null hypothesis of stationarity.

library(tseries)

```
#perform KPSS test  
kpss.test(data, null="Trend")
```

KPSS Test for Trend Stationarity

```
data: data  
KPSS Trend = 0.149, Truncation lag parameter = 2, p-value = 0.04751
```

The output for this non-stationary series yields a KPSS Trend statistic of 0.149 and a p-value of **0.04751**. When compared against the standard significance level of $\alpha = 0.05$, the calculated p-value (0.04751) is indeed less than 0.05. This result dictates that we **reject the null hypothesis** (H_0). Rejecting the null hypothesis leads us to the formal statistical conclusion that the time series is **not trend stationary** and therefore requires appropriate processing, such as differencing, before advanced time series modeling can be conducted.

Critical Considerations for Interpreting KPSS Results

While the **KPSS test** is a powerful diagnostic tool, its results should always be interpreted within the broader context of time series analysis. One critical consideration is the choice between testing for level stationarity (`null="Level"`, the default) versus testing for **trend stationarity** (`null="Trend"`). Testing for level stationarity means assuming the series fluctuates around a fixed mean, while trend stationarity means it fluctuates around a stable, deterministic trend. Misinterpreting the underlying data structure can lead to incorrect conclusions regarding the need for differencing.

Furthermore, the KPSS test is often susceptible to issues related to determining the appropriate truncation lag parameter, which controls the number of lagged autocovariances included in the calculation of the test statistic. The `tseries` package attempts to select this lag parameter automatically, but poor selection can impact the robustness of the test result, potentially biasing the **p-value**. Analysts should ideally cross-validate the KPSS result with other tests, such as ADF, to ensure a comprehensive understanding of the unit root dynamics.

When both the ADF and KPSS tests are performed, four possible outcomes emerge regarding the series' stationarity status. For example, if ADF suggests non-stationarity and KPSS suggests non-stationarity (both reject their respective nulls), the evidence is strong for non-stationarity. However, if ADF fails to reject non-stationarity and KPSS fails to reject stationarity, the results are ambiguous and suggest closer examination of the data structure, possibly indicating stationarity around a deterministic trend.

Next Steps in Time Series Data Preparation

If the KPSS test leads to the rejection of the null hypothesis, indicating that the series is non-stationary, the next phase of data preparation involves applying transformations to achieve stationarity. The most common and effective technique is **differencing**, where the difference between consecutive observations is calculated. This process, often repeated until the series becomes stationary, eliminates the unit root and stabilizes the mean.

Alternatively, if the series is found to be **trend stationary** (meaning we fail to reject H_0 with `null="Trend"`), analysts may opt for detrending the data by regressing the series against a time variable and using the residuals. These residuals, which represent the stochastic component of the data, should then be stationary and suitable for modeling. The decision between differencing and detrending is fundamental and depends on whether the non-stationarity is driven by a deterministic trend or a stochastic unit root process.

Mastering the application and interpretation of tests like the KPSS in R is indispensable for generating reliable results in forecasting and econometric modeling. The clear distinction made by

the KPSS test regarding the presence or absence of stationarity provides the necessary statistical confirmation before transitioning into advanced modeling techniques such as ARMA, ARIMA, or GARCH models, ensuring the fundamental assumptions of these models are met.

Resources for Advanced Time Series Analysis

The examples provided demonstrate the straightforward execution and critical interpretation of the **KPSS test** using the powerful statistical capabilities of R. By correctly applying the `kpss.test()` function and understanding the implications of the resulting p-value, analysts can confidently assess whether a time series requires further transformation, such as differencing, prior to modeling.

This test is a cornerstone of robust time series methodology, guaranteeing that subsequent modeling efforts are based on statistically sound data. For those looking to deepen their expertise, further tutorials and documentation on the `tseries` package, as well as general resources on unit root testing and ARIMA methodology, are highly recommended to expand proficiency in econometric analysis.

Effective implementation of stationarity checks ensures the reliability and accuracy of predictive models, turning raw data into actionable insights for forecasting and decision-making.