

[Go to Courses](#)[ADF Test](#)[P-Value](#)[Python](#)[Significance Tests](#)[Stationarity](#)[Time Series](#)

Augmented Dickey Fuller Test (ADF Test) – Must Read Guide

📅 November 2, 2019 by Selva Prabhakaran



Augmented Dickey Fuller test (ADF Test) is a common statistical test used to test whether a given Time series is stationary or not. It is one of the most commonly used statistical test when it comes to analyzing the stationarity of a series.

1. Introduction

In [ARIMA time series forecasting](#), the first step is to determine the number of differencing required to make the series stationary.

Since testing the stationarity of a time series is a frequently performed activity in autoregressive models, the ADF test along with [KPSS test](#) is something that you need to be fluent in when performing [time series analysis](#).

Another point to remember is **the ADF test is fundamentally a statistical significance test**. That means, **there is a hypothesis testing involved with a null**

[Feedback](#)

and alternate hypothesis and as a result a test statistic is computed and p-values get reported.

It is from the test statistic and the p-value, you can make an inference as to whether a given series is stationary or not.

So, how exactly does the ADF test work? let's see the mathematical intuition behind the test with clear examples.

Let's begin.

2. What is a Unit Root Test?

The ADF test belongs to a category of tests called 'Unit Root Test', which is the proper method for testing the stationarity of a time series.

So what does a 'Unit Root' mean?

Unit root is a characteristic of a time series that makes it non-stationary.

Technically speaking, a unit root is said to exist in a time series of the value of $\alpha = 1$ in the below equation.

$$Y_t = \alpha Y_{t-1} + \beta X_e + \epsilon$$

where, Y_t is the value of the time series at time 't' and X_e is an exogenous variable (a separate explanatory variable, which is also a time series).

What does this mean to us?

The presence of a unit root means the time series is non-stationary. Besides, the number of unit roots contained in the series corresponds to the number of differencing operations required to make the series stationary.

Feedback

Alright, let's come back to topic.

Get started in ML the right way

Hi! I am Selva, and I am glad you are reading this!

With ML+ University's sequential courses, you now have a clear path to become ML expert. No shifting between multiple books. Get started with the most direct, effective way to becoming ML expert. This is the real deal. (Includes video lessons + jupyter notebooks + portfolio projects + exercises + doubt clearing support)

Enroll to the Foundations of ML Course (FREE)

Want the whole package? Start the [Complete ML Mastery Path](#)

3. Dickey-Fuller Test

Before going into ADF test, let's first understand what is the Dickey-Fuller test.

A Dickey-Fuller test is a unit root test that tests the null hypothesis that $\alpha=1$ in the following model equation. α is the coefficient of the first lag on Y .

Null Hypothesis (H_0): $\alpha=1$

$$y_t = c + \beta t + \alpha y_{t-1} + \phi \Delta Y_{t-1} + e_t$$

where,

- $y(t-1)$ = lag 1 of time series

Feedback

- $\Delta Y(t-1)$ = first difference of the series at time (t-1)

Fundamentally, it has a similar null hypothesis as the unit root test. That is, the coefficient of $Y(t-1)$ is 1, implying the presence of a unit root. If not rejected, the series is taken to be non-stationary.

The Augmented Dickey-Fuller test evolved based on the above equation and is one of the most common form of Unit Root test.

4. How does Augmented Dickey Fuller (ADF) Test work?

As the name suggest, the ADF test is an 'augmented' version of the Dickey Fuller test.

The ADF test expands the Dickey-Fuller test equation to include high order regressive process in the model.

$$y_t = c + \beta t + \alpha y_{t-1} + \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} \dots + \phi_p \Delta Y_{t-p} + e_t$$

If you notice, we have only added more differencing terms, while the rest of the equation remains the same. This adds more thoroughness to the test.

The null hypothesis however is still the same as the Dickey Fuller test.

A key point to remember here is: Since the null hypothesis assumes the presence of unit root, that is $\alpha=1$, the p-value obtained should be less than the significance level (say 0.05) in order to reject the null hypothesis. Thereby, inferring that the series is stationary.

However, this is a very common mistake analysts commit with this test. That is, if the p-value is less than significance level, people mistakenly take the series to be non-stationary.

5. ADF Test in Python

So, how to perform a Augmented Dickey-Fuller test in Python?

The `statsmodel` package provides a reliable implementation of the ADF test via the `adfuller()` function in `statsmodels.tsa.stattools`. It returns the following outputs:

1. The p-value
2. The value of the test statistic
3. Number of lags considered for the test
4. The critical value cutoffs.

When the test statistic is lower than the critical value shown, you reject the null hypothesis and infer that the time series is stationary.

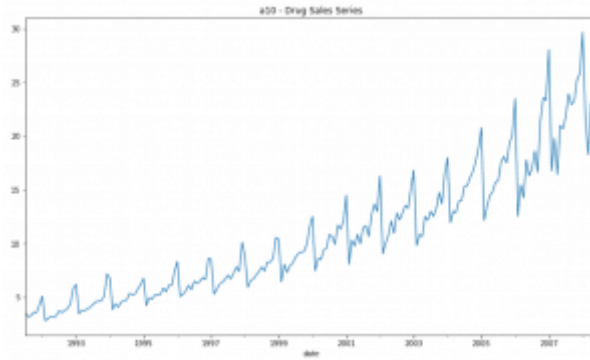
Alright, let's run the ADF test on the `a10` dataset from the `fpp` package from R. This dataset counts the total monthly scripts for pharmaceutical products falling under ATC code A10. The original source of this dataset is the Australian Health Insurance Commission.

As see earlier, the null hypothesis of the test is the presence of unit root, that is, the series is non-stationary.

```
# Setup and Import data
from statsmodels.tsa.stattools import adfuller
import pandas as pd
import numpy as np
%matplotlib inline

url = 'https://raw.githubusercontent.com/selva86/datasets/master/a1
df = pd.read_csv(url, parse_dates=['date'], index_col='date')
series = df.loc[:, 'value'].values
df.plot(figsize=(14,8), legend=None, title='a10 - Drug Sales Series
```

Feedback



The packages and the data is loaded, we have everything needed to perform the test using `adfuller()` .

An optional argument the `adfuller()` accepts is the number of lags you want to consider while performing the **OLS regression**.

By default, this value is $12 * (\text{nobs}/100)^{\{1/4\}}$, where `nobs` is the number of observations in the series. But, optionally you can specify either the maximum number of lags with `maxlags` parameter or let the algorithm compute the optimal number iteratively.

This can be done by setting the `autolag='AIC'` . By doing so, the `adfuller` will choose a the number of lags that yields the lowest AIC. This is usually a good option to follow.

```
# ADF Test
result = adfuller(series, autolag='AIC')
print(f'ADF Statistic: {result[0]}')
print(f'n_lags: {result[1]}')
print(f'p-value: {result[1]}')
for key, value in result[4].items():
    print('Critical Values:')
    print(f'    {key}, {value}')
```

Result:

ADF **Statistic:** 3.1451856893067296

n_lags: 1.0

p-value: 1.0

Critical Values:

1%, -3.465620397124192

Critical Values:

5%, -2.8770397560752436

Critical Values:

10%, -2.5750324547306476

The p-value is obtained is greater than significance level of 0.05 and the ADF statistic is higher than any of the critical values.

Clearly, there is no reason to reject the null hypothesis. So, the time series is in fact non-stationary.

6. ADF Test on stationary series

Now, let's see another example of performing the test on a series of random numbers which is usually considered as stationary.

Let's use `np.random.randn()` to generate a randomized series.

```
# ADF test on random numbers
series = np.random.randn(100)
result = adfuller(series, autolag='AIC')
print(f'ADF Statistic: {result[0]}')
print(f'p-value: {result[1]}')
for key, value in result[4].items():
    print('Critical Values:')
    print(f'    {key}, {value}')
```

Result:

ADF **Statistic**: -7.4715740767231456

p-**value**: 5.0386184272419386e-11

Critical Values:

1%, -3.4996365338407074

Critical Values:

5%, -2.8918307730370025

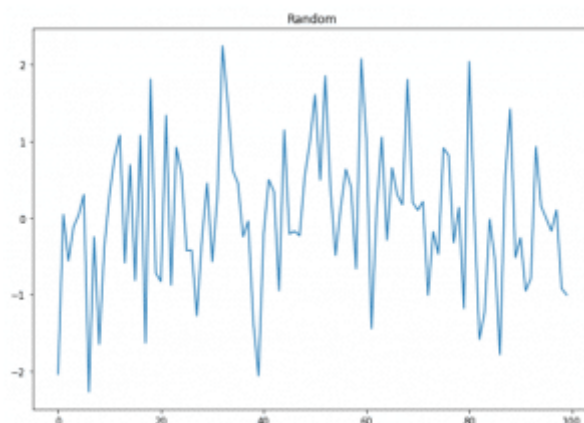
Critical Values:

10%, -2.5829283377617176

The p-value is very less than the significance level of 0.05 and hence we can reject the null hypothesis and take that the series is stationary.

Let's visualise the series as well to confirm.

```
import matplotlib.pyplot as plt
%matplotlib inline
fig, axes = plt.subplots(figsize=(10,7))
plt.plot(series);
plt.title('Random');
```



7. Conclusion

We saw how the Augmented Dickey Fuller Test works and how to perform it using `statsmodels`. Now given any time series, you should be in a position to

perform the ADF Test and make a fair inference on whether the series is stationary or not.

In the next one we'll see how to perform the [KPSS test](#).



Selva Prabhakaran

Selva is the Chief Author and Editor of Machine Learning Plus, with 4 Million+ readership. He has authored courses and books with 100K+ students, and is the Principal Data Scientist of a global firm.

Previous Article

[KPSS Test for Stationarity](#)



Next Article

[data.table in R – The Complete Beginners Guide](#)



ALSO ON MACHINELEARNINGPLUS.COM

Understanding the math with examples (python)

3 years ago • 9 comments

Mahalanobis distance is an effective multivariate distance metric that measures the ...

How to Train Text Classification Model in ...

2 years ago • 2 comments

Text Classification is the process of categorizing texts. SpaCy makes text ...

How to Train spaCy to Autodetect New ...

2 years ago • 13 comments

Named-entity recognition (NER) is the process of automatically identifying ...

Feedback

What do you think?

15 Responses



Upvote



Love

5 Comments

machinelearningplus.com

Disqus' Privacy Policy

Login ▾

Favorite

Tweet

Share

Sort by Newest ▾



Join the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS

**DefaultTempAccount000001** • a year ago

Just a note that I think there's an error in the code in section 5. It should be `print(f'n_lags: {result[2]}')` to print the number of lags.

| • Reply • Share ▸

**Sayed Williams** • a year ago

Wow! This was a quite descriptive article. Thanks.

| • Reply • Share ▸

**Samurai Fox** • 3 years ago

Great Article on ADFuller !

I ran the adfuller and output is as below:

```
`ADF Statistic: -2.457146`
```

```
`p-value: 0.126252`
```

```
`Critical Values: 1%: -3.438, 5%: -2.865, 10%: -2.569`
```

<https://github.com/RSwarnka...>

Is series stationary or non-stationary?

| • Reply • Share ▸

**Nandhakumar G** Samurai Fox • 2 years ago

series is stationary because p-value is 12% so reject the null hypothesis (Non-Stationary), then accept the alternative hypothesis (Stationary)

Feedback

More Articles

Time Series

ARIMA Model – Complete Guide to Time Series Forecasting in Python

Aug 22, 2021

Time Series

Augmented Dickey Fuller Test (ADF Test) – Must Read Guide

Nov 02, 2019

Time Series

KPSS Test for Stationarity

Feedback



Join

Subscribe to Machine Learning Plus for high value data science content



Resources

Blogs

Courses

Project Bluebook

Time Series Template

About us

Terms of Use

Privacy Policy

Contact Us

Refund Policy

© Machinelearningplus. All rights reserved.

Feedback