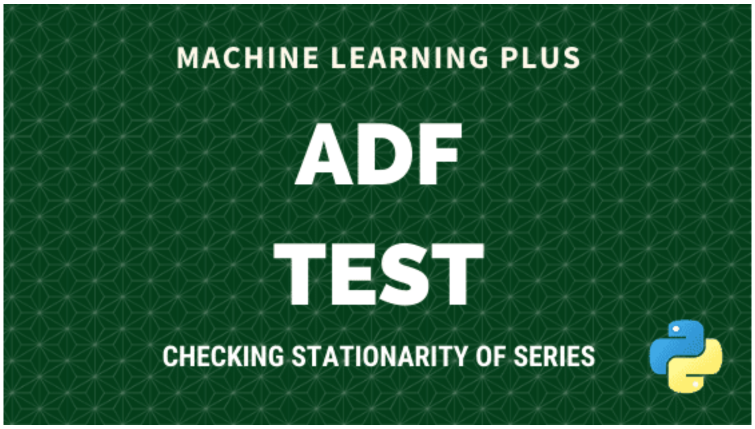
Machine Learning Plus (https://www.machinelearningplus.com/)



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

by Selva Prabhakaran (https://www.machinelearningplus.com/author/selva86/) | Posted on November 2, 2019 (https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/)

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 stationary of a series.

1. Introduction

In **ARIMA time series forecasting (https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/)**, 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 (https://www.machinelearningplus.com/time-series/kpss-test-for-stationarity/)** is something that you need to be fluent in when performing **time series analysis (https://www.machinelearningplus.com/time-series/time-series-analysis-python/)**.

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 and alternate hypothesis and as a result a test statistic is computed and **p-values (https://www.machinelearningplus.com/statistics/p-value/)** 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?

Recent Posts

Principal Components Analysis (PCA) – Better Explained (https://www.machinelearningplus.com/machine-learning/principal-components-analysis-pca-better-explained/)

data.table in R – The Complete Beginners Guide (https://www.machinelearningplus.com/data-manipulation/datatable-in-r-complete-guide/)

Augmented Dickey Fuller Test (ADF Test) – Must Read Guide (https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/)

KPSS Test for Stationarity (https://www.machinelearningplus.com/time-series/kpss-test-for-stationarity/)

101 R data.table Exercises (https://www.machinelearningplus.com/data-manipulation/101-r-data-table-exercises/)

P-Value – Understanding from Scratch (https://www.machinelearningplus.com/statistics/p-value/)

101 Python datatable Exercises (pydatatable) (https://www.machinelearningplus.com/data-manipulation/101-python-datatable-exercises-pydatatable/)

Vector Autoregression (VAR) – Comprehensive Guide with Examples in Python (https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/)

Mahalonobis Distance – Understanding the math with examples (python) (https://www.machinelearningplus.com/statistics/mahalanol distance/)

datetime in Python – Simplified Guide with Clear Examples (https://www.machinelearningplus.com/python/datetime-python-examples/)

Tags

Bigrams (https://www.machinelearningplus.com/tag/bigrams/)
Classification (https://www.machinelearningplus.com/tag/class
Feedback

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.

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

(https://www.machinelearningplus.com/wp-content/uploads/2019/11/equation_1.png)

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.

Alright, let’s come back to topic.

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 mull hypothesis that α=1 in the following model equation. `alpha` is the coefficient of the first lag on Y.

Null Hypothesis (H0): alpha=1

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

(https://www.machinelearningplus.com/wp-content/uploads/2019/11/equation_2.png)

where,

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

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

(https://www.machinelearningplus.com/wp-content/uploads/2019/11/equation_3.png)

Corpus (https://www.machinelearningplus.com/tag/corpus/) Cosine Similarity (https://www.machinelearningplus.com/tag/cosine-similarity/) data.table (https://www.machinelearningplus.com/tag/data-table/) Data Manipulation (https://www.machinelearningplus.com/tag/d manipulation/) Debugging (https://www.machinelearningplus.com/tag/debugging/) Doc2Vec (https://www.machinelearningplus.com/tag/doc2vec/) Evaluation Metrics (https://www.machinelearningplus.com/tag/evaluation-metrics/) FastText (https://www.machinelearningplus.com/tag/fasttext/) Feature Selection (https://www.machinelearningplus.com/tag/feature-selection/) Gensim (https://www.machinelearningplus.com/tag/gensim/) LDA (https://www.machinelearningplus.com/tag/lda/) Lemmatization (https://www.machinelearningplus.com/tag/lemmatization/) Linear Regression (https://www.machinelearningplus.com/tag/linear-regression/) Logistic (https://www.machinelearningplus.com/tag/logistic/) LSI (https://www.machinelearningplus.com/tag/lsi/) Matplotlib (https://www.machinelearningplus.com/tag/matplotlib Multiprocessing (https://www.machinelearningplus.com/tag/multiprocessing/) NLP (https://www.machinelearningplus.com/tag/nlp/) NLTK (https://www.machinelearningplus.com/tag/nltk/) Numpy (https://www.machinelearningplus.com/tag/numpy/) P-Value (https://www.machinelearningplus.com/tag/p-value/) Pandas (https://www.machinelearningplus.com/tag/pandas/) Parallel Processing (https://www.machinelearningplus.com/tag/parallel-processing/) Phraser (https://www.machinelearningplus.com/tag/phraser/) Practice Exercise (https://www.machinelearningplus.com/tag/practice-exercise/) Python (https://www.machinelearnin R (https://www.machinelearningplus.com/ Regex (https://www.machinelearningplus.com/tag/regex/) Regression (https://www.machinelearningplus.com/tag/regression/) Residual Analysis (https://www.machinelearningplus.com/tag/residual-analysis/) Scikit Learn (https://www.machinelearningplus.com/tag/scikit-learn/) Significance Tests (https://www.machinelearningplus.com/tag/significanc tests/) Soft Cosine Similarity (https://www.machinelearningplus.com/tag/soft-cosine-similarity/) spaCy (https://www.machinelearningplus.com/tag/spacy/) Stationarity (https://www.machinelearningplus.com/tag/stationarit Summarization (https://www.machinelearningplus.com/tag/summarization/) TaggedDocument (https://www.machinelearningplus.com/tag/taggeddocument/) TextBlob (https://www.machinelearningplus.com/tag/textblob/) TFIDF (https://www.machinelearningplus.com/tag/tfidf/) Time Series (https://www.machinelearningplus.com/tag/ti series/) Topic Modeling (https://www.machinelearningplus.com/tag/topic-modeling/) Visualization

Feedback

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 α=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.

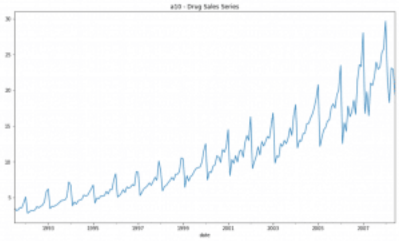
When the test statistic is lower than the critical value shown, you reject the null hypothesis and infer that the time series is stationary.

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/a10.csv'
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');
```



(https://www.machinelearningplus.com/wp-content/uploads/2019/11/a10_series.png)

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

(https://www.machinelearningplus.com/tag/visualizati
Word2Vec (https://www.machinelearningplus.com/tag/word2vec/)

Feedback

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*(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('Critial Values:')
    print(f'   {key}, {value}')
```

Result:

```
ADF Statistic: 3.1451856893067296
n_lags: 1.0
p-value: 1.0
Critial Values:
   1%, -3.465620397124192
Critial Values:
   5%, -2.8770397560752436
Critial 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('Critial Values:')
    print(f'   {key}, {value}')
```

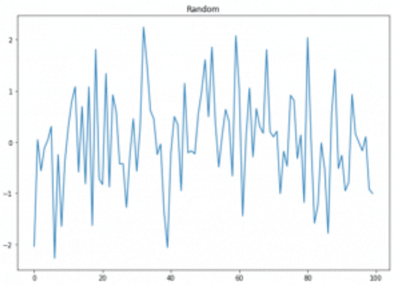
Result:

```
ADF Statistic: -7.4715740767231456
p-value: 5.0386184272419386e-11
Critial Values:
  1%, -3.4996365338407074
Critial Values:
  5%, -2.8918307730370025
Critial 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');
```



(https://www.machinelearningplus.com/wp-content/uploads/2019/11/random_series.png)

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 (https://www.machinelearningplus.com/time-series/kpss-test-for-stationarity/)**.

Feedback

Sponsored

London Apartment Prices Might Surprise You

London Apartments | Search Ads

Jennifer Aniston No Makeup Photo Confirms The Rumors

Trading Blvd

The Cost of Hair Transplant in Dehradun Might Shock You

Hair Transplant | Search Ads

6 Easy Exercises to Lose Thigh Fat Fast

Fitness Engage

Indians born before 1964 are now eligible to test hearing aids

Hearing Aid Trial

17 Cancer Causing Foods You HAVE to Stop Eating

Health & Human Research

What do you think?

2 Responses

👍 Upvote 😍 Love

0 Comments **machinelearningplus.com** **Anurag Bisht**

♡ **Recommend** Tweet Share **Sort by Newest**

Start the discussion…

Be the first to comment.

✉ **Subscribe** **Add Disqus to your siteAdd DisqusAdd** 🔒 **Disqus' Privacy PolicyPrivacy PolicyPrivacy**

Copyright **Machine Learning Plus (https://www.machinelearningplus.com/)**. All rights reserved. | Theme by Superb WordPress Themes (https://superbthemes.com/)

Feedback