

x

-

(<http://play.google.com/store/apps/details?id=com.analyticsvidhya.android>)



(<https://www.analyticsvidhya.com/blog/>)



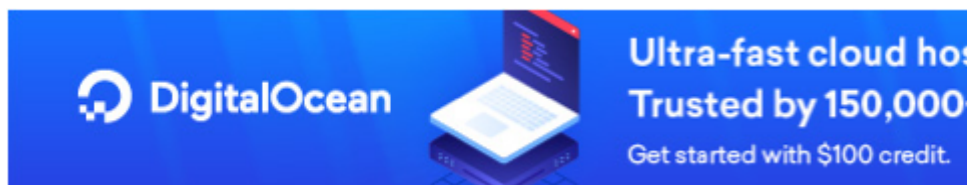
(https://datahack.analyticsvidhya.com/contest/american-express-amexpert-2018/?utm_source=AVtopblogBanner)

[BUSINESS ANALYTICS \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/BUSINESS-ANALYTICS/\)](https://www.analyticsvidhya.com/blog/category/business-analytics/)

[TIME SERIES \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/TIME-SERIES/\)](https://www.analyticsvidhya.com/blog/category/time-series/)

A Complete Tutorial on Time Series Modeling in R

[TAVISH SRIVASTAVA \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/AUTHOR/TAVISH1/\)](https://www.analyticsvidhya.com/blog/author/tavish1/), DECEMBER 16, 2015



Introduction

‘Time’ is the most important factor which ensures success in a business. It’s difficult to keep up with the pace of time. But, technology has developed some powerful methods using which we can ‘see things’ ahead of time. Don’t worry, I am not talking about Time Machine. Let’s be realistic here!

I’m talking about the methods of prediction & forecasting. One such method, which deals with time based data is **Time Series Modeling**. As the name suggests, it involves working on time (years, days, hours, minutes) based data, to derive hidden insights to make informed decision making.

Time series models are very useful models when you have serially correlated data. Most of business houses work on time series data to analyze sales number for the next year, website traffic, competition position and much more. However, it is also one of the areas, which many analysts do not understand.

So, if you aren't sure about complete process of time series modeling, this guide would introduce you to various levels of time series modeling and its related techniques.



<https://www.analyticsvidhya.com/wp-content/uploads/2015/12/tsm1.jpg>

The following topics are covered in this tutorial as shown below:

Table of Contents

1. Basics – Time Series Modeling
2. Exploration of Time Series Data in R
3. Introduction to ARMA Time Series Modeling
4. Framework and Application of ARIMA Time Series Modeling

Time to get started!

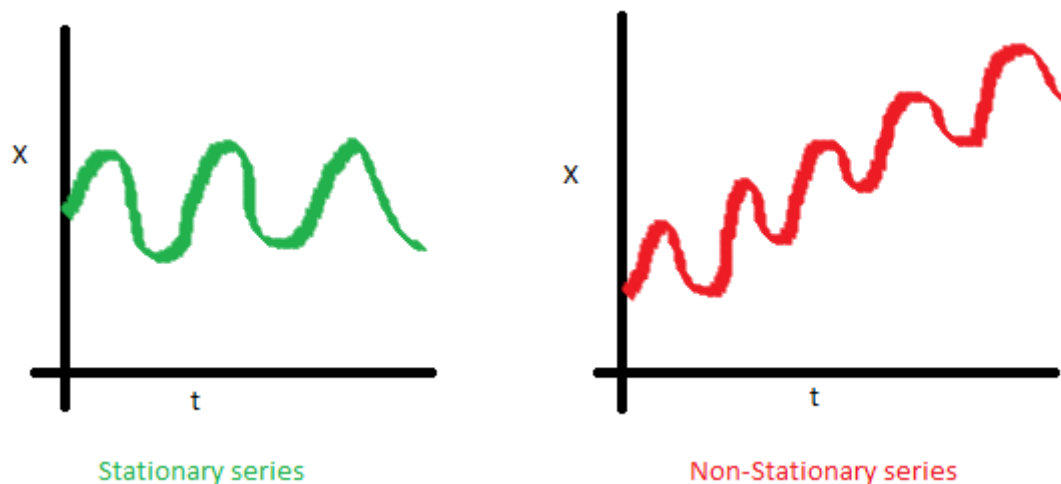
1. Basics – Time Series Modeling

Let's begin from basics. This includes stationary series, random walks, Rho Coefficient, Dickey Fuller Test of Stationarity. If these terms are already scaring you, don't worry – they will become clear in a bit and I bet you will start enjoying the subject as I explain it.

Stationary Series

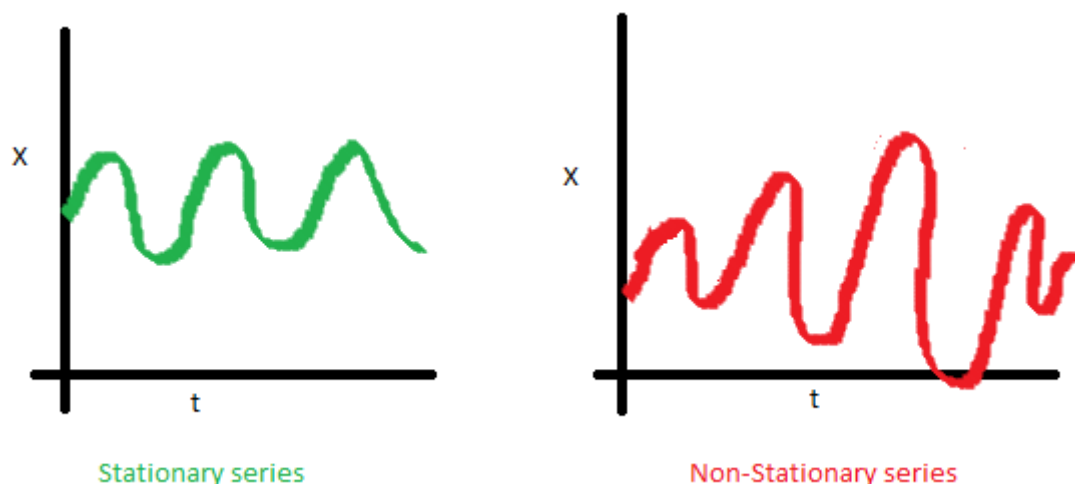
There are three basic criterion for a series to be classified as stationary series :

1. The mean of the series should not be a function of time rather should be a constant. The image below has the left hand graph satisfying the condition whereas the graph in red has a time dependent mean.



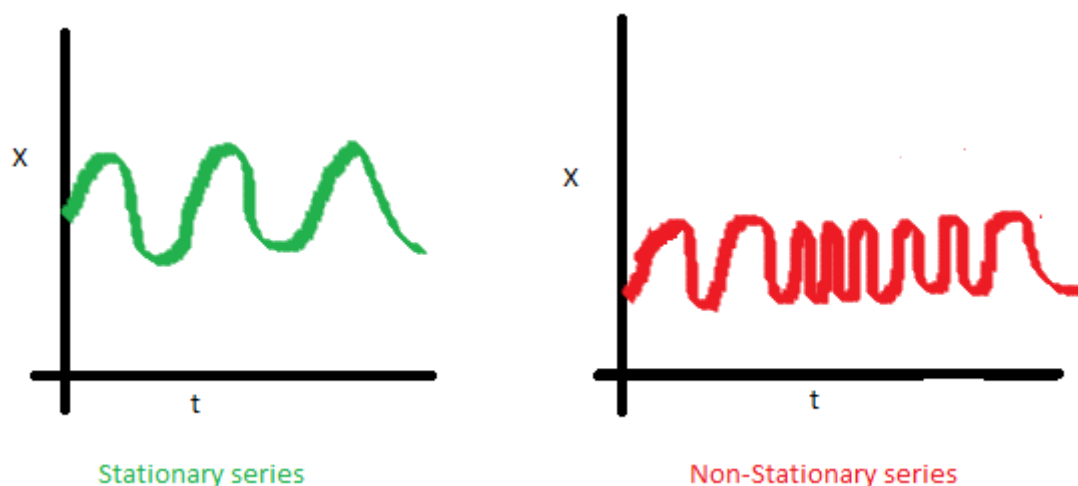
(https://www.analyticsvidhya.com/wp-content/uploads/2015/02/Mean_nonstationary.png)

2. The variance of the series should not be a function of time. This property is known as homoscedasticity. Following graph depicts what is and what is not a stationary series. (Notice the varying spread of distribution in the right hand graph)



(https://www.analyticsvidhya.com/wp-content/uploads/2015/02/Var_nonstationary.png)

3. The covariance of the i th term and the $(i + m)$ th term should not be a function of time. In the following graph, you will notice the spread becomes closer as the time increases. Hence, the covariance is not constant with time for the 'red series'.



(https://www.analyticsvidhya.com/wp-content/uploads/2015/02/Cov_nonstationary.png)

Why do I care about 'stationarity' of a time series?

The reason I took up this section first was that until unless your time series is stationary, you cannot build a time series model. In cases where the stationary criterion are violated, the first requisite becomes to stationarize the time series and then try stochastic models to predict this time series. There are multiple ways of bringing this stationarity. Some of them are Detrending, Differencing etc.

Random Walk

This is the most basic concept of the time series. You might know the concept well. But, I found many people in the industry who interpret random walk as a stationary process. In this section with the help of some mathematics, I will make this concept crystal clear for ever. Let's take an example.

Example: Imagine a girl moving randomly on a giant chess board. In this case, next position of the girl is only dependent on the last position.



(<https://www.analyticsvidhya.com/wp-content/uploads/2015/02/RandomWalk.gif>)

(Source: <http://scifun.chem.wisc.edu/WOP/RandomWalk.html>)

Now imagine, you are sitting in another room and are not able to see the girl. You want to predict the position of the girl with time. How accurate will you be? Of course you will become more and more inaccurate as the position of the girl changes. At $t=0$ you exactly know where the girl is. Next time, she can only move to 8 squares and hence your probability dips to $1/8$ instead of 1 and it keeps on going down. Now let's try to formulate this series :

$$X(t) = X(t-1) + Er(t)$$

where $Er(t)$ is the error at time point t . This is the randomness the girl brings at every point in time.

Now, if we recursively fit in all the X s, we will finally end up to the following equation :

$$X(t) = X(0) + \text{Sum}(Er(1), Er(2), Er(3) \dots Er(t))$$

Now, lets try validating our assumptions of stationary series on this random walk formulation:

1. Is the Mean constant ?

$$E[X(t)] = E[X(0)] + \text{Sum}(E[Er(1)], E[Er(2)], E[Er(3)] \dots E[Er(t)])$$

We know that Expectation of any Error will be zero as it is random.

Hence we get $E[X(t)] = E[X(0)] = \text{Constant}$.

2. Is the Variance constant?

$$\text{Var}[X(t)] = \text{Var}[X(0)] + \text{Sum}(\text{Var}[Er(1)], \text{Var}[Er(2)], \text{Var}[Er(3)] \dots \text{Var}[Er(t)])$$

$$\text{Var}[X(t)] = t * \text{Var}(\text{Error}) = \text{Time dependent}.$$

Hence, we infer that the random walk is not a stationary process as it has a time variant variance. Also, if we check the covariance, we see that too is dependent on time.

Let's spice up things a bit,

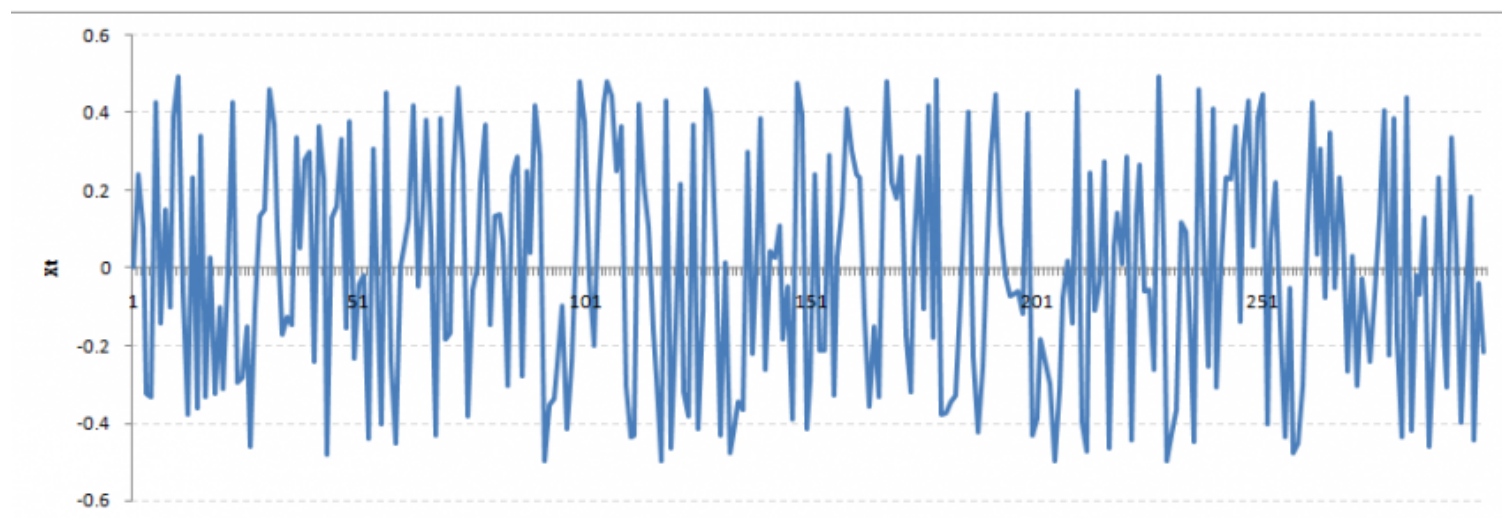
We already know that a random walk is a non-stationary process. Let us introduce a new coefficient in the equation to see if we can make the formulation stationary.

Introduced coefficient : Rho

$$X(t) = \text{Rho} * X(t-1) + Er(t)$$

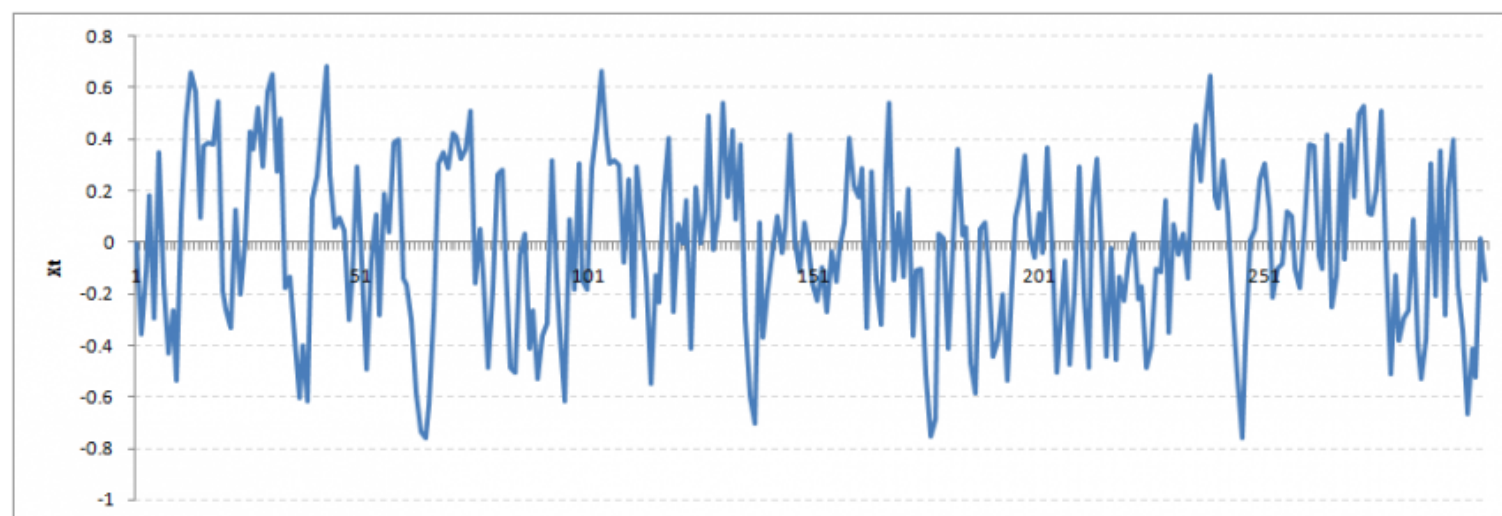
Now, we will vary the value of Rho to see if we can make the series stationary. Here we will interpret the scatter visually and not do any test to check stationarity.

Let's start with a perfectly stationary series with $\text{Rho} = 0$. Here is the plot for the time series :



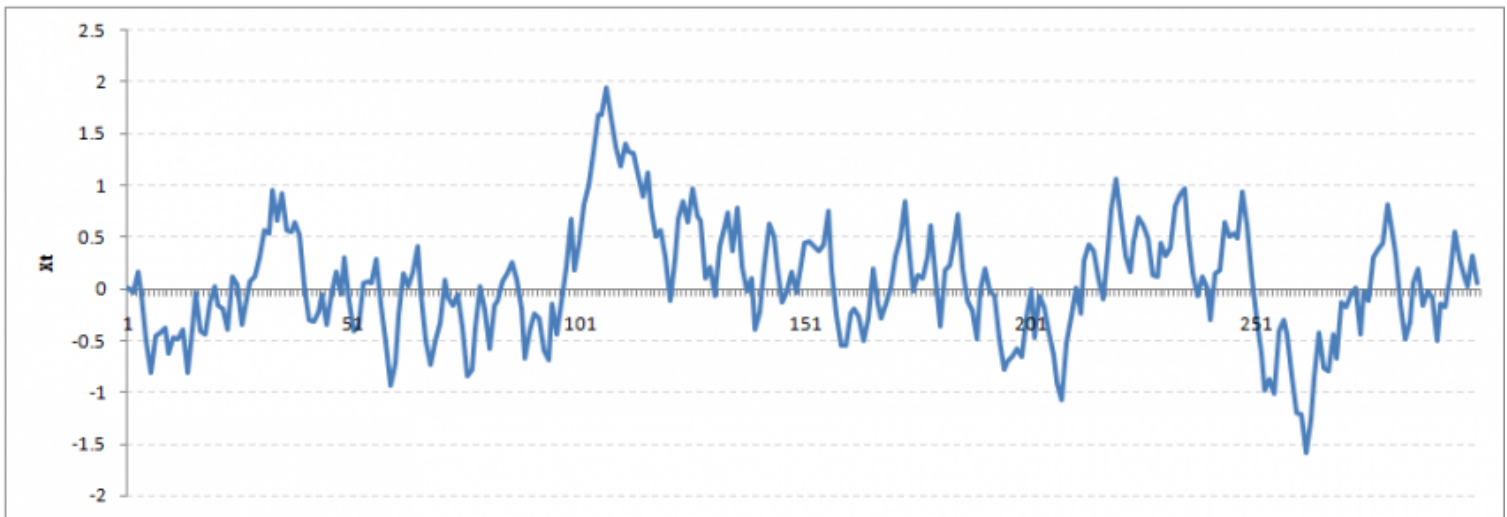
(<https://www.analyticsvidhya.com/wp-content/uploads/2015/02/rho0.png>)

Increase the value of Rho to 0.5 gives us following graph :



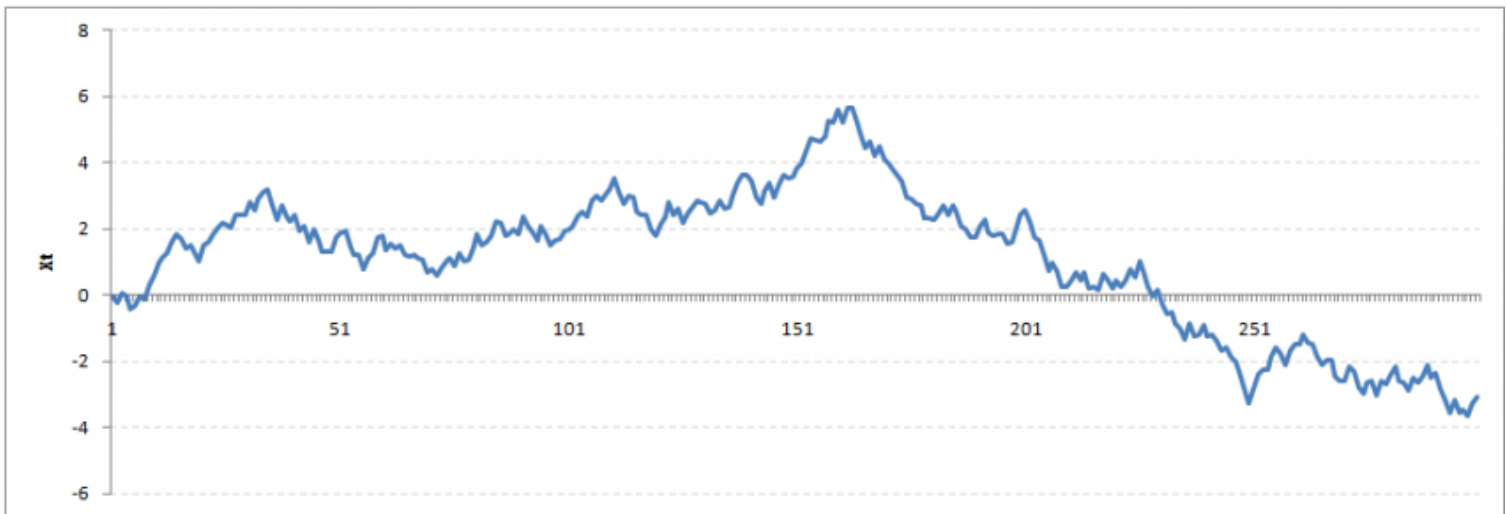
(<https://www.analyticsvidhya.com/wp-content/uploads/2015/02/rho5.png>)

You might notice that our cycles have become broader but essentially there does not seem to be a serious violation of stationary assumptions. Let's now take a more extreme case of $\text{Rho} = 0.9$



(<https://www.analyticsvidhya.com/wp-content/uploads/2015/02/rho9.png>)

We still see that the X returns back from extreme values to zero after some intervals. This series also is not violating non-stationarity significantly. Now, let's take a look at the random walk with $\rho = 1$.



(<https://www.analyticsvidhya.com/wp-content/uploads/2015/02/rho1.png>)

This obviously is a violation to stationary conditions. What makes $\rho = 1$ a special case which comes out badly in stationary test? We will find the mathematical reason to this.

Let's take expectation on each side of the equation " $X(t) = \rho * X(t-1) + \epsilon(t)$ "

$$E[X(t)] = \rho * E[X(t-1)]$$

This equation is very insightful. The next X (or at time point t) is being pulled down to $\rho * \text{Last value of } X$.

For instance, if $X(t - 1) = 1$, $E[X(t)] = 0.5$ (for $\text{Rho} = 0.5$). Now, if X moves to any direction from zero, it is pulled back to zero in next step. The only component which can drive it even further is the error term. Error term is equally probable to go in either direction. What happens when the Rho becomes 1? No force can pull the X down in the next step.

Dickey Fuller Test of Stationarity

What you just learnt in the last section is formally known as Dickey Fuller test. Here is a small tweak which is made for our equation to convert it to a Dickey Fuller test:

$$X(t) = \text{Rho} * X(t-1) + \text{Er}(t)$$

$$\Rightarrow X(t) - X(t-1) = (\text{Rho} - 1) X(t-1) + \text{Er}(t)$$

We have to test if $\text{Rho} - 1$ is significantly different than zero or not. If the null hypothesis gets rejected, we'll get a stationary time series.

Stationary testing and converting a series into a stationary series are the most critical processes in a time series modelling. You need to memorize each and every detail of this concept to move on to the next step of time series modelling.

Let's now consider an example to show you what a time series looks like.

2. Exploration of Time Series Data in R

Here we'll learn to handle time series data on R. Our scope will be restricted to data exploring in a time series type of data set and not go to building time series models.

I have used an inbuilt data set of R called `AirPassengers`. The dataset consists of monthly totals of international airline passengers, 1949 to 1960.

Loading the Data Set

Following is the code which will help you load the data set and spill out a few top level metrics.

```
> data(AirPassengers)
> class(AirPassengers)
[1] "ts"
```

```
#This tells you that the data series is in a time series format
> start(AirPassengers)
[1] 1949 1
```

```
#This is the start of the time series
```

```
> end(AirPassengers)
[1] 1960 12
```

```
#This is the end of the time series
```

```
> frequency(AirPassengers)
[1] 12
```

```
#The cycle of this time series is 12months in a year
> summary(AirPassengers)
Min. 1st Qu. Median Mean 3rd Qu. Max.
104.0 180.0 265.5 280.3 360.5 622.0
```

Detailed Metrics

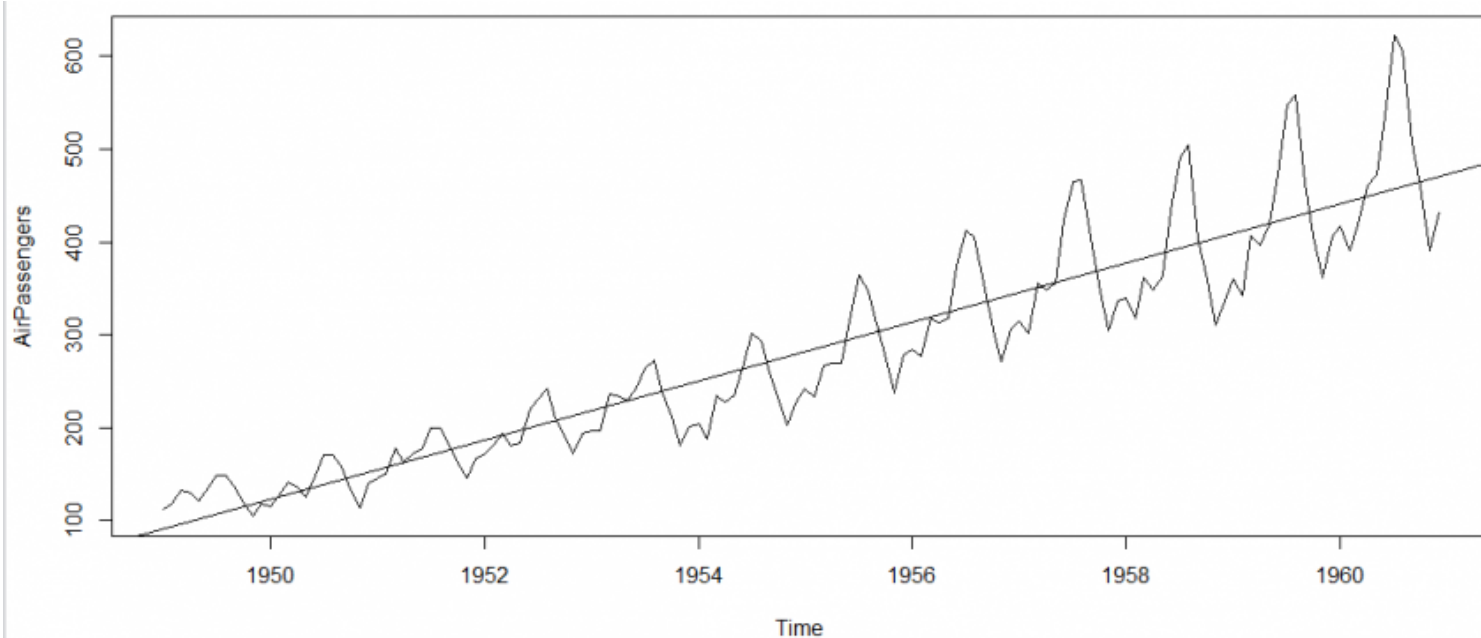
```
#The number of passengers are distributed across the spectrum
```

```
> plot(AirPassengers)
```

```
#This will plot the time series
```

```
> abline(reg=lm(AirPassengers~time(AirPassengers)))
```

```
# This will fit in a line
```



(https://www.analyticsvidhya.com/wp-content/uploads/2015/02/plot_AP1.png)

Here are a few more operations you can do:

```
> cycle(AirPassengers)
```

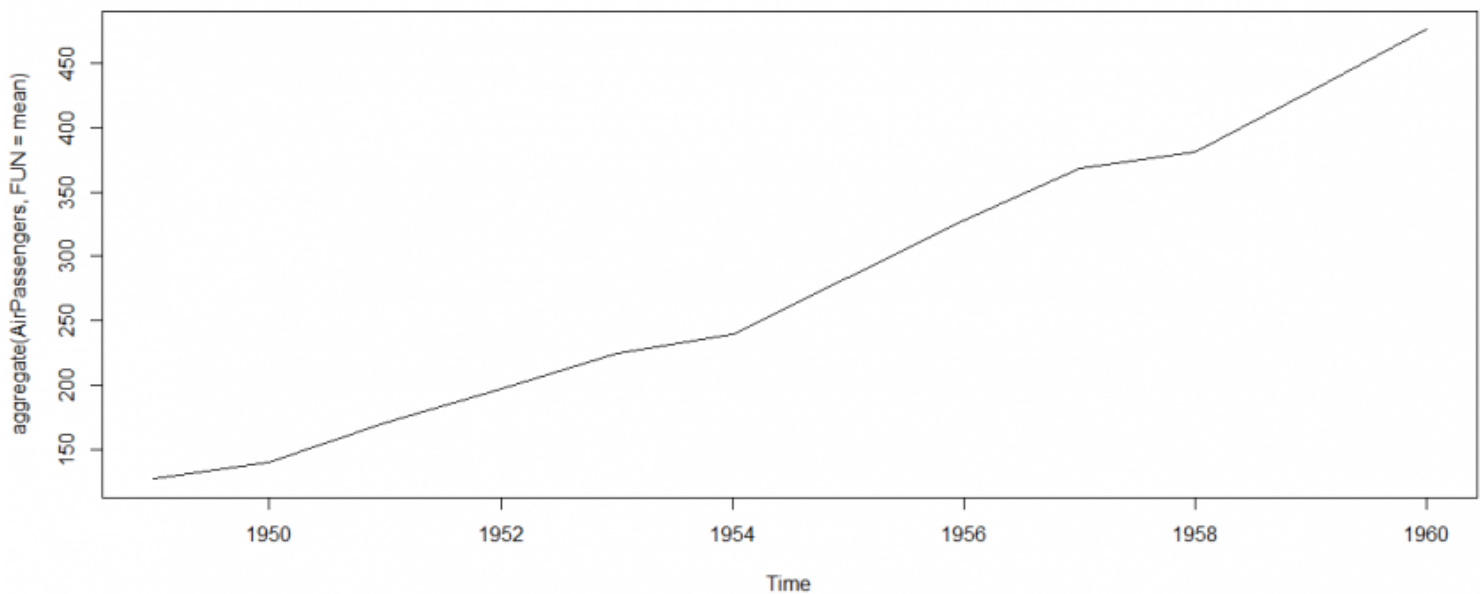
```
#This will print the cycle across years.
```

```
>plot(aggregate(AirPassengers,FUN=mean))
```

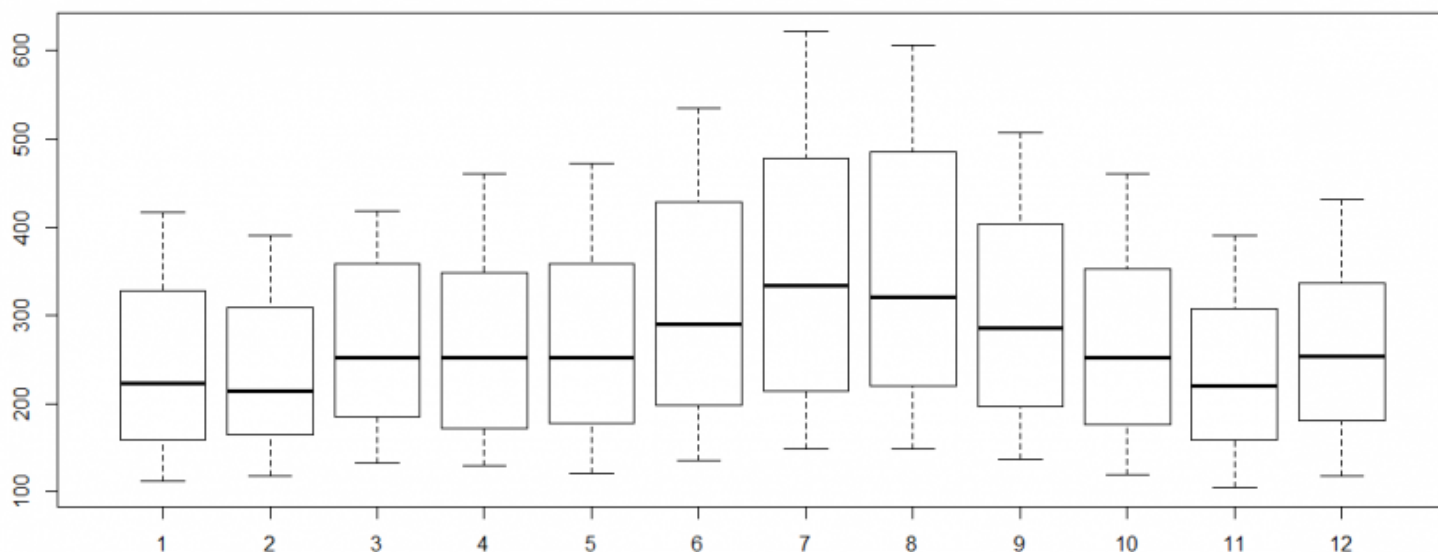
```
#This will aggregate the cycles and display a year on year trend
```

```
> boxplot(AirPassengers~cycle(AirPassengers))
```

```
#Box plot across months will give us a sense on seasonal effect
```



(https://www.analyticsvidhya.com/wp-content/uploads/2015/02/plot_aggregate.png).



(https://www.analyticsvidhya.com/wp-content/uploads/2015/02/plot_month_wise.png)

Important Inferences

1. The year on year trend clearly shows that the #passengers have been increasing without fail.
2. The variance and the mean value in July and August is much higher than rest of the months.
3. Even though the mean value of each month is quite different their variance is small. Hence, we have strong seasonal effect with a cycle of 12 months or less.

Exploring data becomes most important in a time series model – without this exploration, you will not know whether a series is stationary or not. As in this case we already know many details about the kind of model we are looking out for.

Let's now take up a few time series models and their characteristics. We will also take this problem forward and make a few predictions.

3. Introduction to ARMA Time Series Modeling

ARMA models are commonly used in time series modeling. In ARMA model, AR stands for auto-regression and MA stands for moving average. If these words sound intimidating to you, worry not – I'll simplify these concepts in next few minutes for you!

We will now develop a knack for these terms and understand the characteristics associated with these models.

But before we start, you should remember, AR or MA are not applicable on non-stationary series.

In case you get a non stationary series, you first need to stationarize the series (by taking difference / transformation) and then choose from the available time series models.

First, I'll explain each of these two models (AR & MA) individually. Next, we will look at the characteristics of these models.

Auto-Regressive Time Series Model

Let's understanding AR models using the case below:

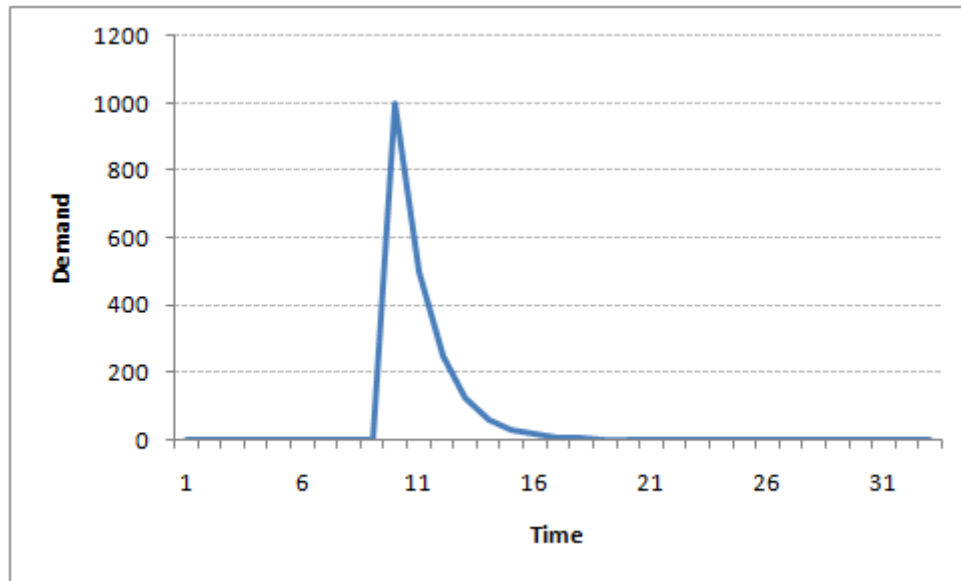
The current GDP of a country say $x(t)$ is dependent on the last year's GDP i.e. $x(t - 1)$. The hypothesis being that the total cost of production of products & services in a country in a fiscal year (known as GDP) is dependent on the set up of manufacturing plants / services in the previous year and the newly set up industries / plants / services in the current year. But the primary component of the GDP is the former one.

Hence, we can formally write the equation of GDP as:

$$x(t) = \alpha * x(t - 1) + \text{error}(t)$$

This equation is known as *AR(1) formulation*. The numeral one (1) denotes that the next instance is solely dependent on the previous instance. The alpha is a coefficient which we seek so as to minimize the error function. Notice that $x(t-1)$ is indeed linked to $x(t-2)$ in the same fashion. Hence, any shock to $x(t)$ will gradually fade off in future.

For instance, let's say $x(t)$ is the number of juice bottles sold in a city on a particular day. During winters, very few vendors purchased juice bottles. Suddenly, on a particular day, the temperature rose and the demand of juice bottles soared to 1000. However, after a few days, the climate became cold again. But, knowing that the people got used to drinking juice during the hot days, there were 50% of the people still drinking juice during the cold days. In following days, the proportion went down to 25% (50% of 50%) and then gradually to a small number after significant number of days. The following graph explains the inertia property of AR series:



(<https://www.analyticsvidhya.com/wp-content/uploads/2015/02/AR1.png>).

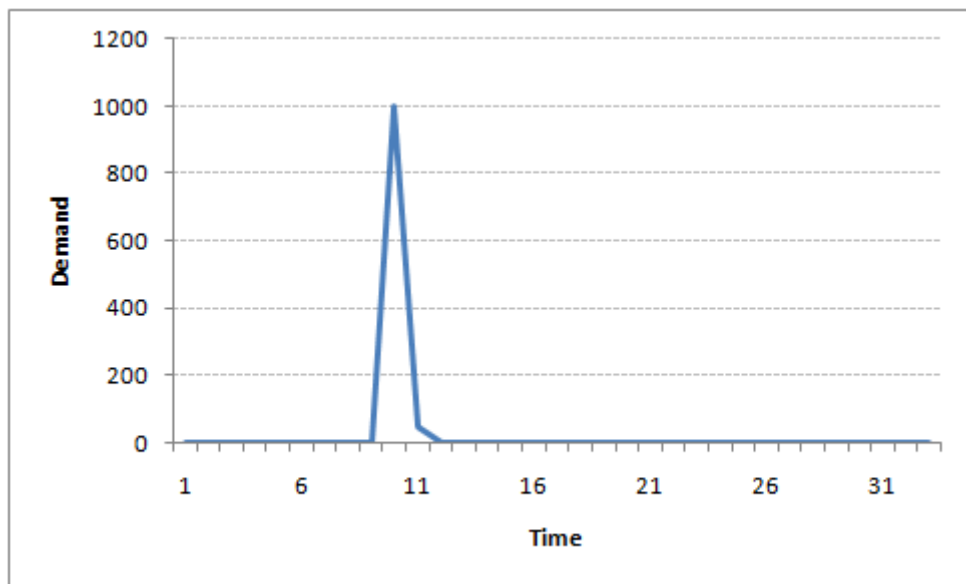
Moving Average Time Series Model

Let's take another case to understand Moving average time series model.

A manufacturer produces a certain type of bag, which was readily available in the market. Being a competitive market, the sale of the bag stood at zero for many days. So, one day he did some experiment with the design and produced a different type of bag. This type of bag was not available anywhere in the market. Thus, he was able to sell the entire stock of 1000 bags (lets call this as $x(t)$). The demand got so high that the bag ran out of stock. As a result, some 100 odd customers couldn't purchase this bag. Lets call this gap as the error at that time point. With time, the bag had lost its woo factor. But still few customers were left who went empty handed the previous day. Following is a simple formulation to depict the scenario :

$$x(t) = \text{beta} * \text{error}(t-1) + \text{error}(t)$$

If we try plotting this graph, it will look something like this :



(<https://www.analyticsvidhya.com/wp-content/uploads/2015/02/MA1.png>)

Did you notice the difference between MA and AR model? In MA model, noise / shock quickly vanishes with time. The AR model has a much lasting effect of the shock.

Difference between AR and MA models

The primary difference between an AR and MA model is based on the correlation between time series objects at different time points. The correlation between $x(t)$ and $x(t-n)$ for $n > \text{order of MA}$ is always zero. This directly flows from the fact that covariance between $x(t)$ and $x(t-n)$ is zero for MA models (something which we refer from the example taken in the previous section). However, the correlation of $x(t)$ and $x(t-n)$ gradually declines with n becoming larger in the AR model. This difference gets exploited irrespective of having the AR model or MA model. The correlation plot can give us the order of MA model.

Exploiting ACF and PACF plots

Once we have got the stationary time series, we must answer two primary questions:

Q1. Is it an AR or MA process?

Q2. What order of AR or MA process do we need to use?

The trick to solve these questions is available in the previous section. Didn't you notice?

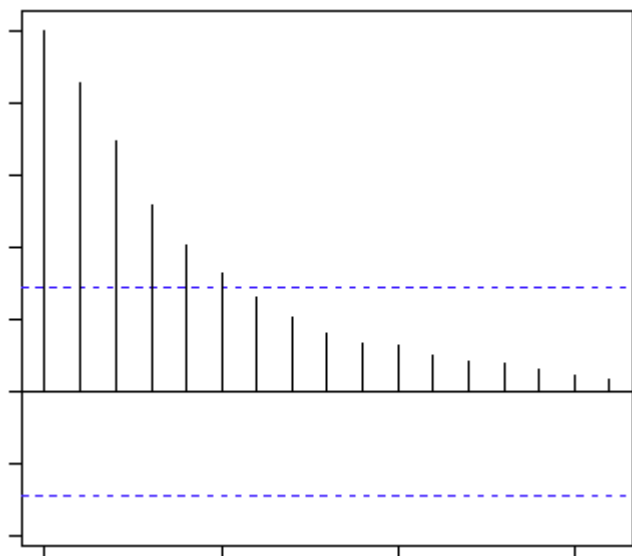
The first question can be answered using Total Correlation Chart (also known as Auto – correlation Function / ACF). ACF is a plot of total correlation between different lag functions. For instance, in GDP problem, the GDP at time point t is $x(t)$. We are interested in the correlation of $x(t)$ with $x(t-1)$, $x(t-2)$ and so on. Now let's reflect on what we have learnt above.

In a moving average series of lag n , we will not get any correlation between $x(t)$ and $x(t - n - 1)$. Hence, the total correlation chart cuts off at n th lag. So it becomes simple to find the lag for a MA series. For an AR series this correlation will gradually go down without any cut off value. So what do we do if it is an AR series?

Here is the second trick. If we find out the partial correlation of each lag, it will cut off after the degree of AR series. For instance, if we have a $AR(1)$ series, if we exclude the effect of 1st lag ($x(t-1)$), our 2nd lag ($x(t-2)$) is independent of $x(t)$. Hence, the partial correlation function (PACF) will drop sharply after the 1st lag. Following are the examples which will clarify any doubts you have on this concept :

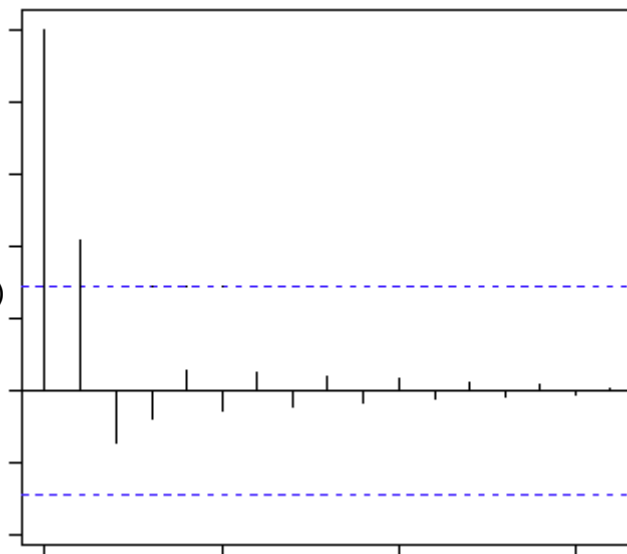
ACF

PACF



([https://www.analyticsvidhya.com/wp-](https://www.analyticsvidhya.com/wp-content/uploads/2015/02/Gradual-decline.gif)

[content/uploads/2015/02/Gradual-decline.gif](https://www.analyticsvidhya.com/wp-content/uploads/2015/02/Gradual-decline.gif))

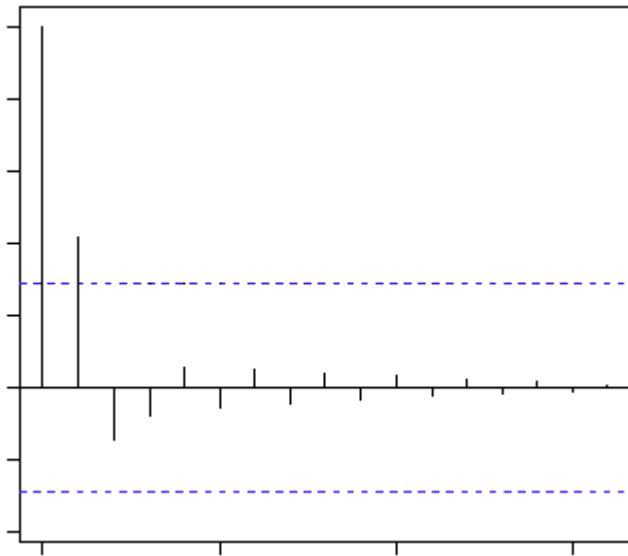


(<https://www.analyticsvidhya.com/wp-content/uploads/2015/02/cut-off.gif>)

The blue line above shows significantly different values than zero. Clearly, the graph above has a cut off on PACF curve after 2nd lag which means this is mostly an AR(2) process.

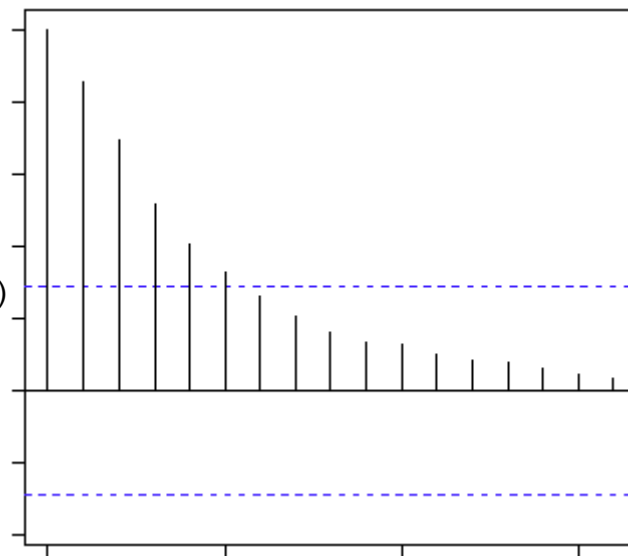
ACF

PACF



([https://www.analyticsvidhya.com/wp-](https://www.analyticsvidhya.com/wp-content/uploads/2015/02/cut-off.gif)

[content/uploads/2015/02/cut-off.gif](https://www.analyticsvidhya.com/wp-content/uploads/2015/02/cut-off.gif))



(<https://www.analyticsvidhya.com/wp-content/uploads/2015/02/Gradual-decline.gif>)

Clearly, the graph above has a cut off on ACF curve after 2nd lag which means this is mostly a MA(2) process.

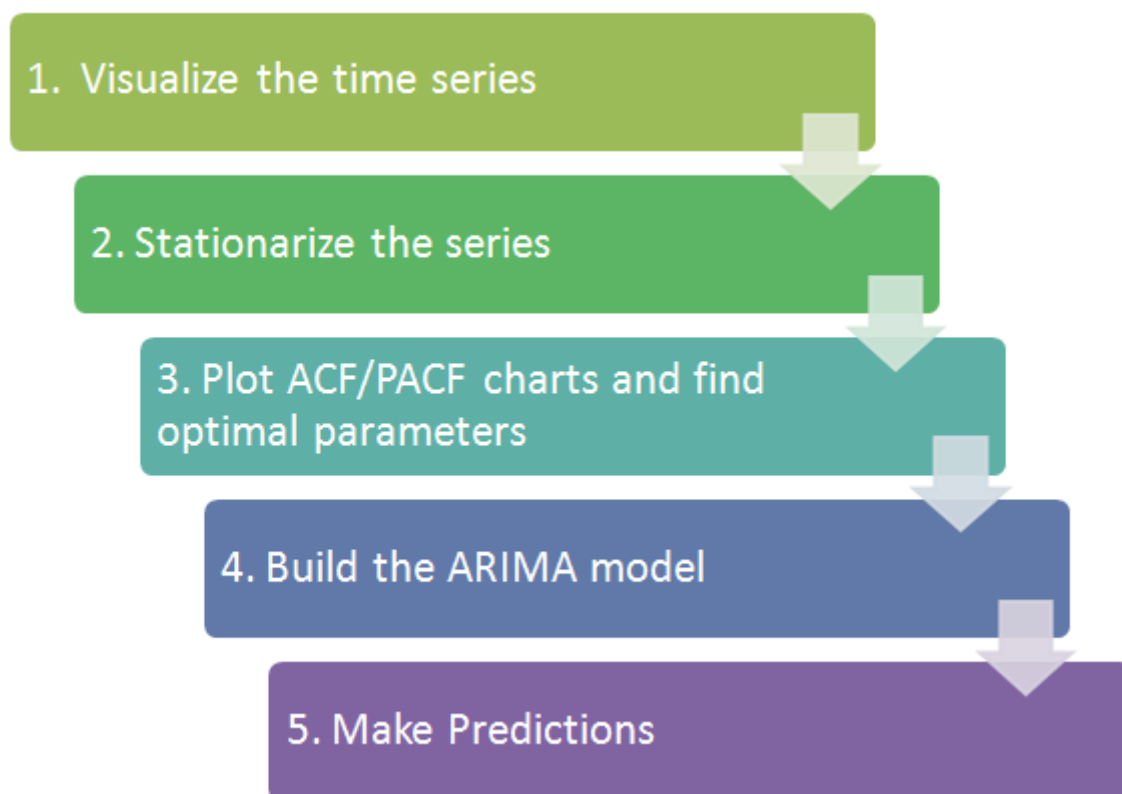
Till now, we have covered on how to identify the type of stationary series using ACF & PACF plots. Now, I'll introduce you to a comprehensive framework to build a time series model. In addition, we'll also discuss about the practical applications of time series modelling.

4. Framework and Application of ARIMA Time Series Modeling

A quick revision, Till here we've learnt basics of time series modeling, time series in R and ARMA modeling. Now is the time to join these pieces and make an interesting story.

Overview of the Framework

This framework(shown below) specifies the step by step approach on '**How to do a Time Series Analysis**':



(<https://www.analyticsvidhya.com/wp-content/uploads/2015/02/flowchart.png>)

As you would be aware, the first three steps have already been discussed above. Nevertheless, the same has been delineated briefly below:

Step 1: Visualize the Time Series

It is essential to analyze the trends prior to building any kind of time series model. The details we are interested in pertain to any kind of trend, seasonality or random behaviour in the series. We have covered this part in the second part of this series.

Step 2: Stationarize the Series

Once we know the patterns, trends, cycles and seasonality, we can check if the series is stationary or not. Dickey – Fuller is one of the popular test to check the same. We have covered this test in the [first part](https://www.analyticsvidhya.com/blog/2015/02/step-step-guide-learn-time-series/) (<https://www.analyticsvidhya.com/blog/2015/02/step-step-guide-learn-time-series/>) of this article series. This doesn't end here! What if the series is found to be non-stationary?

There are three commonly used techniques to make a time series stationary:

1. **Detrending** : Here, we simply remove the trend component from the time series. For instance, the equation of my time series is:

$$x(t) = (\text{mean} + \text{trend} * t) + \text{error}$$

We'll simply remove the part in the parentheses and build model for the rest.

2. **Differencing** : This is the commonly used technique to remove non-stationarity. Here we try to model the differences of the terms and not the actual term. For instance,

$$x(t) - x(t-1) = \text{ARMA}(p, q)$$

This differencing is called as the Integration part in AR(I)MA. Now, we have three parameters

p : AR

d : I

q : MA

3. **Seasonality** : Seasonality can easily be incorporated in the ARIMA model directly. More on this has been discussed in the applications part below.

Step 3: Find Optimal Parameters

The parameters p, d, q can be found using [ACF](https://www.analyticsvidhya.com/blog/2015/03/introduction-auto-regression-moving-average-time-series/) and [PACF](https://www.analyticsvidhya.com/blog/2015/03/introduction-auto-regression-moving-average-time-series/) plots (<https://www.analyticsvidhya.com/blog/2015/03/introduction-auto-regression-moving-average-time-series/>). An addition to this approach is can be, if both ACF and PACF decreases gradually, it indicates that we need to make

the time series stationary and introduce a value to “d”.

Step 4: Build ARIMA Model

With the parameters in hand, we can now try to build ARIMA model. The value found in the previous section might be an approximate estimate and we need to explore more (p,d,q) combinations. The one with the lowest BIC and AIC should be our choice. We can also try some models with a seasonal component. Just in case, we notice any seasonality in ACF/PACF plots.

Step 5: Make Predictions

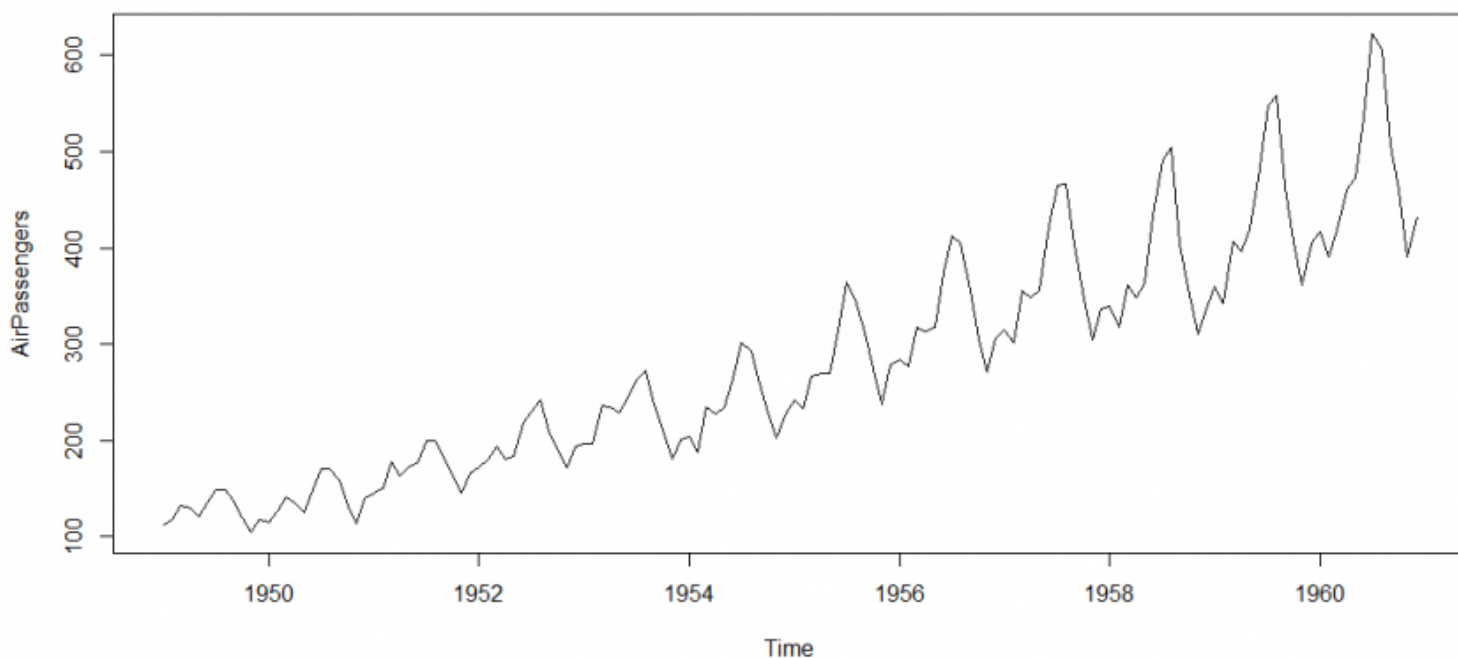
Once we have the final ARIMA model, we are now ready to make predictions on the future time points. We can also visualize the trends to cross validate if the model works fine.

Applications of Time Series Model

Now, we'll use the same example that we have used above. Then, using time series, we'll make future predictions. We recommend you to check out the example before proceeding further.

Where did we start ?

Following is the plot of the number of passengers with years. Try and make observations on this plot before moving further in the article.



(https://www.analyticsvidhya.com/wp-content/uploads/2015/02/plot_AP.png)

Here are my observations :

1. There is a trend component which grows the passenger year by year.
2. There looks to be a seasonal component which has a cycle less than 12 months.
3. The variance in the data keeps on increasing with time.

We know that we need to address two issues before we test stationary series. One, we need to remove unequal variances. We do this using log of the series. Two, we need to address the trend component. We do this by taking difference of the series. Now, let's test the resultant series.

```
adf.test(diff(log(AirPassengers)), alternative="stationary", k=0)
```

Augmented Dickey-Fuller Test

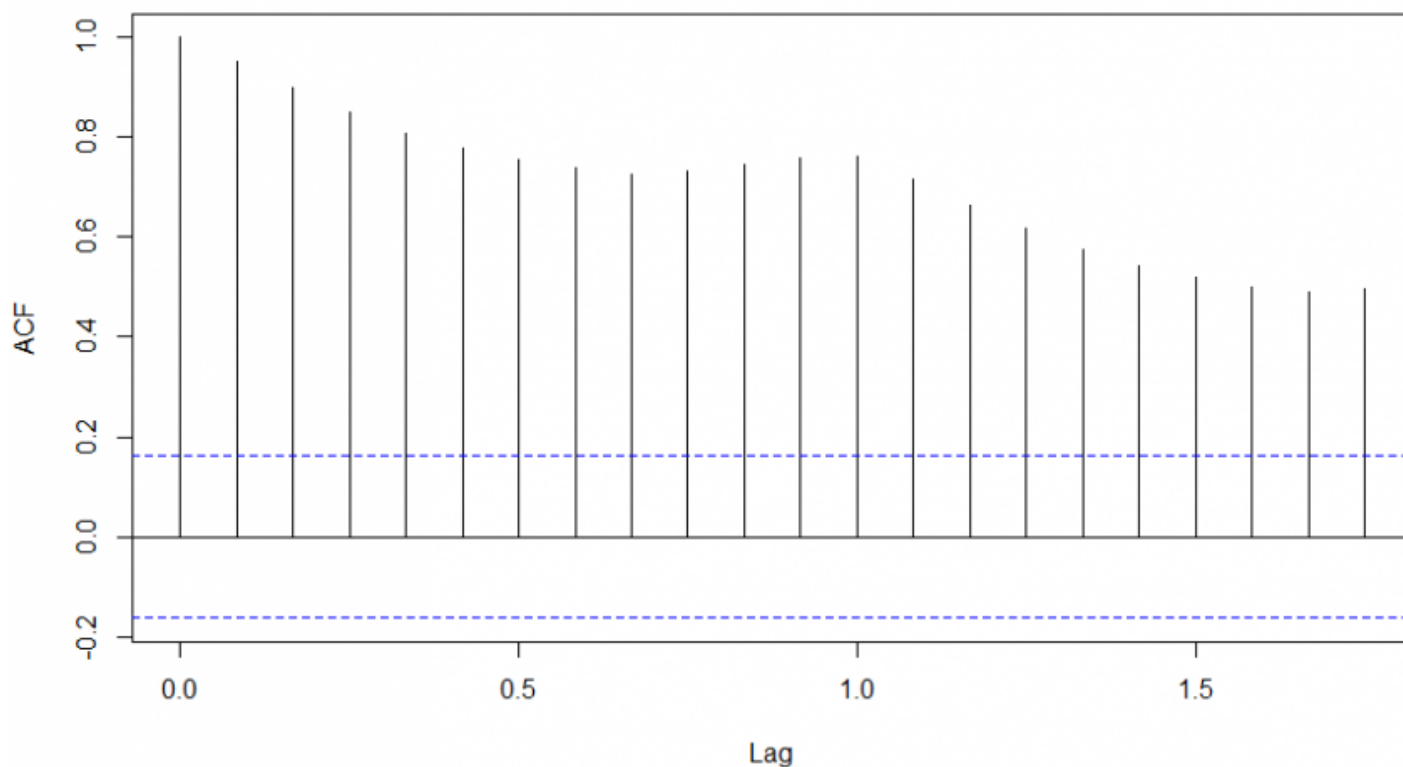
```
data: diff(log(AirPassengers))  
Dickey-Fuller = -9.6003, Lag order = 0,  
p-value = 0.01  
alternative hypothesis: stationary
```

We see that the series is stationary enough to do any kind of time series modelling.

Next step is to find the right parameters to be used in the ARIMA model. We already know that the 'd' component is 1 as we need 1 difference to make the series stationary. We do this using the Correlation plots. Following are the ACF plots for the series :

#ACF Plots

```
acf(log(AirPassengers))
```

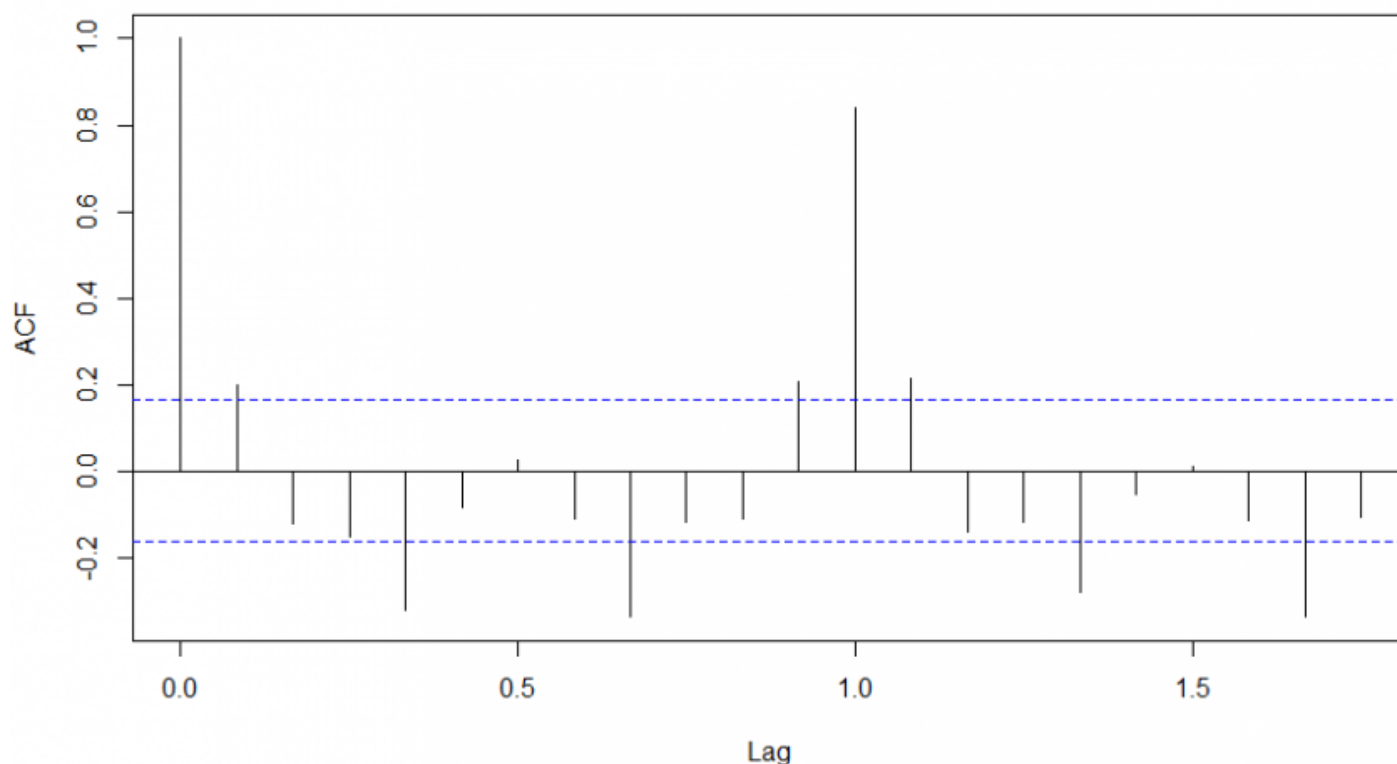


(https://www.analyticsvidhya.com/wp-content/uploads/2015/02/ACF_original.png)

What do you see in the chart shown above?

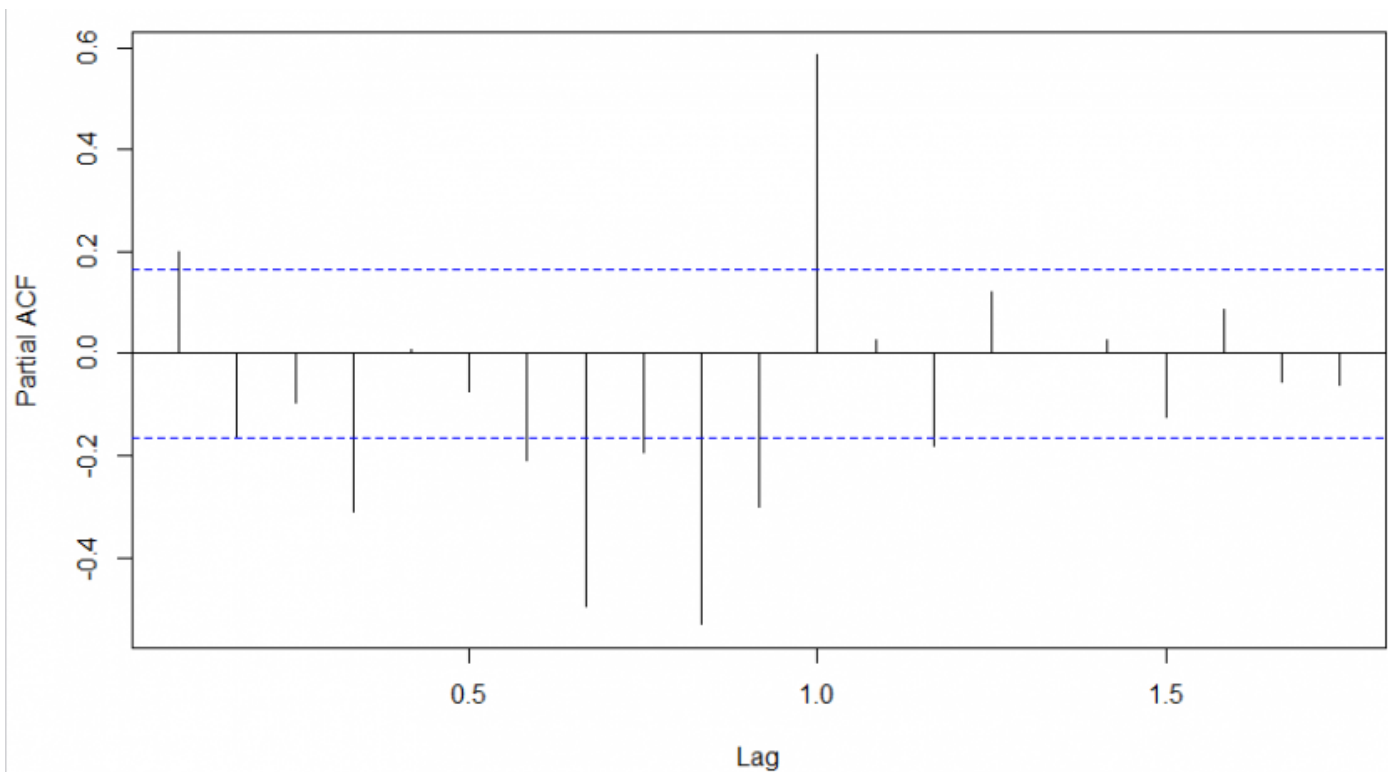
Clearly, the decay of ACF chart is very slow, which means that the population is not stationary. We have already discussed above that we now intend to regress on the difference of logs rather than log directly. Let's see how ACF and PACF curve come out after regressing on the difference.

```
acf(diff(log(AirPassengers)))
```



(<https://www.analyticsvidhya.com/wp-content/uploads/2015/02/ACF-diff.png>)

```
pacf(diff(log(AirPassengers)))
```



(<https://www.analyticsvidhya.com/wp-content/uploads/2015/02/PACF-diff.png>)

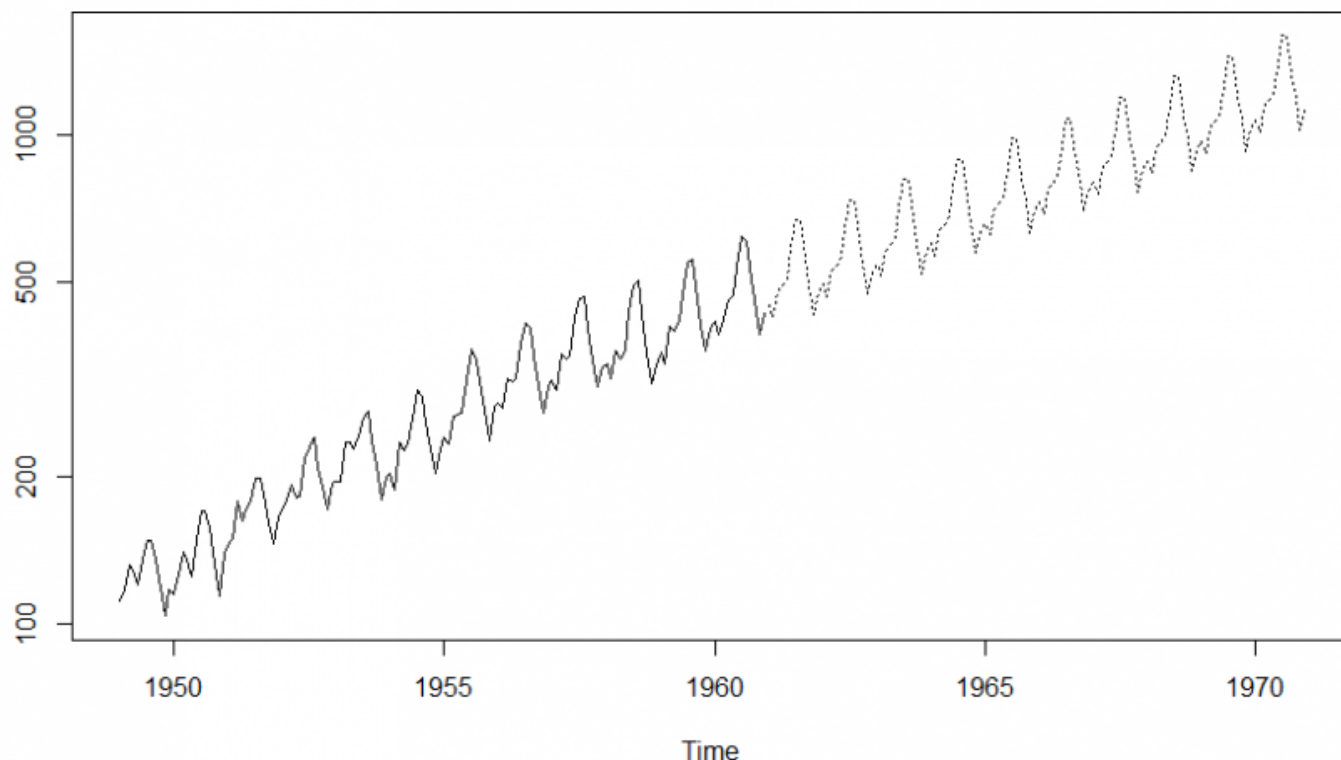
Clearly, ACF plot cuts off after the first lag. Hence, we understood that value of p should be 0 as the ACF is the curve getting a cut off. While value of q should be 1 or 2. After a few iterations, we found that (0,1,1) as (p,d,q) comes out to be the combination with least AIC and BIC.

Let's fit an ARIMA model and predict the future 10 years. Also, we will try fitting in a seasonal component in the ARIMA formulation. Then, we will visualize the prediction along with the training data. You can use the following code to do the same :

```
(fit <- arima(log(AirPassengers), c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12)))
```

```
pred <- predict(fit, n.ahead = 10*12)
```

```
ts.plot(AirPassengers, 2.718^pred$pred, log = "y", lty = c(1,3))
```



(<https://www.analyticsvidhya.com/wp-content/uploads/2015/02/predictions.png>)

End Notes

With this, we come to this end of tutorial on Time Series Modeling. I hope this will help you to improve your knowledge to work on time based data. To reap maximum benefits out of this tutorial, I'd suggest you to practice these R codes side by side and check your progress.

Did you find the article useful? Share with us if you have done similar kind of analysis before. Do let us know your thoughts about this article in the box below.

Note – The discussions of this article are going on at AV's Discuss portal. [Join here](https://discuss.analyticsvidhya.com/t/discussions-for-article-a-complete-tutorial-on-time-series-modeling-in-r/65674?u=jalfaizy) (<https://discuss.analyticsvidhya.com/t/discussions-for-article-a-complete-tutorial-on-time-series-modeling-in-r/65674?u=jalfaizy>)!

If you like what you just read & want to continue your analytics learning, [subscribe to our emails](http://feedburner.google.com/fb/a/mailverify?uri=analyticsvidhya) (<http://feedburner.google.com/fb/a/mailverify?uri=analyticsvidhya>), [follow us on twitter](http://twitter.com/analyticsvidhya) (<http://twitter.com/analyticsvidhya>) or like our [facebook page](#)


<http://facebook.com/analyticsvidhya>).

You can also read this article on Analytics Vidhya's Android APP



https://play.google.com/store/apps/details?id=com.analyticsvidhya.android&utm_source=blog_article&utm_campaign=blog&pcampaignid=MKT-Other-global-all-co-prtnr-py-PartBadge-Mar2515-1

Share this:

 (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/?share=linkedin&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/?share=facebook&nb=1>) 1K+

 (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/?share=google-plus-1&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/?share=twitter&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/?share=pocket&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/?share=reddit&nb=1>)

Like this:

Loading...



MOBIZMO DIY 360 DEGREE ROT

Rs. 1,199.00 ✓ **prin**

TAGS : [ACF PLOTS \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/ACF-PLOTS/\)](https://www.analyticsvidhya.com/blog/tag/acf-plots/), [ARIMA \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/ARIMA/\)](https://www.analyticsvidhya.com/blog/tag/arima/), [ARIMA MODELING \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/ARIMA-MODELING/\)](https://www.analyticsvidhya.com/blog/tag/arima-modeling/), [ARMA MODELING \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/ARMA-MODELING/\)](https://www.analyticsvidhya.com/blog/tag/arma-modeling/), [DICKEY-FULLER TEST \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/DICKEY-FULLER-TEST/\)](https://www.analyticsvidhya.com/blog/tag/dickey-fuller-test/), [FORECASTING \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/FORECASTING/\)](https://www.analyticsvidhya.com/blog/tag/forecasting/), [TIME SERIES \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/TAG/TIME-SERIES/\)](https://www.analyticsvidhya.com/blog/tag/time-series/)

NEXT ARTICLE

Top Business Analytics Programs in India (2015 – 16)[\(https://www.analyticsvidhya.com/blog/2015/12/top-business-analytics-programs-india-2015-2016/\)](https://www.analyticsvidhya.com/blog/2015/12/top-business-analytics-programs-india-2015-2016/)

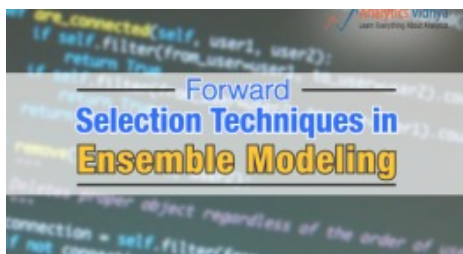
...

PREVIOUS ARTICLE

10 Machine Learning Algorithms Explained to an 'Army Soldier'[\(https://www.analyticsvidhya.com/blog/2015/12/10-machine-learning-algorithms-explained-army-soldier/\)](https://www.analyticsvidhya.com/blog/2015/12/10-machine-learning-algorithms-explained-army-soldier/)[\(https://www.analyticsvidhya.com/blog/author/tavish1/\)](https://www.analyticsvidhya.com/blog/author/tavish1/)**Tavish Srivastava****(<https://www.analyticsvidhya.com/blog/author/tavish1/>)**

Tavish is an IIT post graduate, a results-driven analytics professional and a motivated leader with 7+ years of experience in data science industry. He has led various high performing data scientists teams in financial domain. His work range from creating high level business strategy for customer engagement and acquisition to developing Next-Gen cognitive Deep/Machine Learning capabilities aligned to these high level strategies for multiple domains including Retail Banking, Credit Cards and Insurance. Tavish is fascinated by the idea of artificial intelligence inspired by human intelligence and enjoys every discussion, theory or even movie related to this idea.

RELATED ARTICLES[TAVISH SRIVASTAVA \(HTTPS://WWW.AN...](https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/)[TAVISH SRIVASTAVA \(HTTPS://WWW.AN...](https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/)[AISHWARYA SINGH \(HTTPS://WWW.ANA...](https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/)



(<https://www.analyticsvidhya.com/blog/2015/09/selection-techniques-ensemble-modelling/>)

Learn to use Forward Selection Techniques for Ensemble Modeling
(<https://www.analyticsvidhya.com/blog/2015/09/selection-techniques-ensemble-modelling/>)



(<https://www.analyticsvidhya.com/blog/2015/01/learning-path-sas-beginner-business-analyst/>)

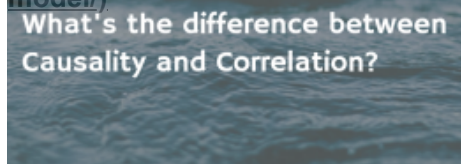
Learning path for SAS – from beginner to a Business Analyst
(<https://www.analyticsvidhya.com/blog/2015/01/learning-path-sas-beginner-business-analyst/>)

This article is quite old and you might not get a prompt response from the author. We request you to post this comment on Analytics Vidhya's **Discussion portal** (<https://discuss.analyticsvidhya.com/>) to get your queries resolved



(<https://www.analyticsvidhya.com/blog/2014/04/solving-survival-model/>)

Simple framework to build a survival analysis model on R
(<https://www.analyticsvidhya.com/blog/2014/04/solving-survival-model/>)



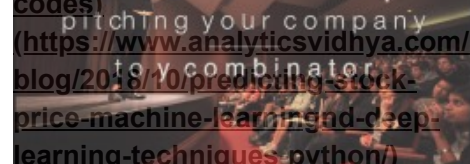
(<https://www.analyticsvidhya.com/blog/2015/06/establish-causality-events/>)

What's the difference between Causality and Correlation?
(<https://www.analyticsvidhya.com/blog/2015/06/establish-causality-events/>)



(<https://www.analyticsvidhya.com/blog/2016/10/predicting-stock-price-machine-learning-and-deep-learning-techniques-python/>)

Stock Prices Prediction Using Machine Learning and Deep Learning Techniques (with Python codes)



(<https://www.analyticsvidhya.com/blog/2016/03/13-machine-learning-data-science-startups-combinator-winter-2016/>)

13 Machine Learning & Data Science Startups from Y Combinator Winter 2016
(<https://www.analyticsvidhya.com/blog/2016/03/13-machine-learning-data-science-startups-combinator-winter-2016/>)

50 COMMENTS



DR.D.K.SAMUEL

[Reply](#)

December 16, 2015 at 5:27 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102002>)

Really useful. Please also write on how to make weather data into a times series for further analysis in R



DR SAHUL BHARTI (HTTP://ALPHABETASTATS.COM)

[Reply](#)

December 16, 2015 at 5:56 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102005>)

Hi

I am a medical specialist (MD Pediatrics) with further training in research and statistics (Panjab University, Chandigarh). In our medical settings, time series data are often seen in ICU and anesthesia related research where patients are continuously monitored for days or even weeks generating such data. Frankly speaking, your article has clearly decoded this arcane process of time series analysis with quite wonderful insight into its practical relevance. Fabulous article Mr Tavish, kindly write more about ARIMA modelling.

Thanks a lot

Dr Sahul Bharti



BASEER

[Reply](#)

December 16, 2015 at 6:37 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102006>)

great article to start with timeseries mod



SHIVAM BANSAL

[Reply](#)

December 16, 2015 at 6:52 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102008>)

Awesome Tutorial.

Big fan of you Tavish, your articles are really great. Explanations in beautiful manner.



ANKUR BHARGAVA

[Reply](#)

December 16, 2015 at 7:19 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102011>)

Please elucidate on PACF part of MA series.

Thanks



TAVISH SRIVASTAVA

[Reply](#)

December 16, 2015 at 9:04 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102014>)

PACF is not really required for MA models, as the degree of MA can be found from ACF directly.

**RASHEED4DEM**[Reply](#)

October 4, 2016 at 2:35 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-116773>)

Thank u it really help a lot

**HUGO**[Reply](#)

December 16, 2015 at 1:02 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102019>)

Hi Tavish.

First off all, congratulations on your work around here. It's been very useful. Thank you
I a doubt and i hope that you can help me

I performed a Dickey-Fuller test on both series ; AirPassengers and diff(log(AirPassengers))

Here the results:

Augmented Dickey-Fuller Test

data: diff(log(AirPassengers))

Dickey-Fuller = -9.6003, Lag order = 0, p-value = 0.01

alternative hypothesis: stationary

and

Augmented Dickey-Fuller Test

data: diff(log(AirPassengers))

Dickey-Fuller = -9.6003, Lag order = 0, p-value = 0.01

alternative hypothesis: stationary

In both tests i got a small p-value that allows me to reject the non stationary hypothesis. Am I right?

If so, the first series is already stationary??

This means that if i had performed a stationary test on the original series had move on to the next step.

Thank you in advance.

**HUGO**[Reply](#)



December 16, 2015 at 1:05 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102020>)

Now with the right results .

Augmented Dickey-Fuller Test

data: AirPassengers

Dickey-Fuller = -4.6392, Lag order = 0, p-value = 0.01

alternative hypothesis: stationary

Augmented Dickey-Fuller Test

data: diff(log(AirPassengers))

Dickey-Fuller = -9.6003, Lag order = 0, p-value = 0.01

alternative hypothesis: stationary



RAM

[Reply](#)

December 18, 2015 at 10:45 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102163>)

@Hugo,

Yes, the `adf.test(AirPassengers)` indicates that the series is stationary. This is a bit misleading.

Reason: This test first does a de-trend on the series, (ie., removes the trend component), then checks for stationarity. Hence it flags the series as stationary.

There is another test in package `fUnitRoots`. Please try this code:

Start

```
install.packages("fUnitRoots") # If you already have installed this package, you can omit this line
```

```
library(fUnitRoots);
```

```
adfTest(AirPassengers);
```

```
adfTest(log(AirPassengers));
```

```
adfTest(diff(AirPassengers));
```

End

Hope this helps..

**ARATI**[Reply](#)

May 3, 2016 at 3:37 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-110379>)

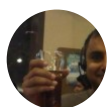
thanks Ram, I had the same question as Hugo and your explanation helped

I just wanted to point out for the benefit of anyone else looking at this that R is cap sensitive, do not forget to capitalize the T in adfTest else your function will not work.

**ANNE**[Reply](#)

March 23, 2018 at 2:44 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-152095>)

If I use diff(AirPassengers) dataset and test it with adfTest it gives stationary

**AJAY (HTTP://DECISIONSTATS.COM)**[Reply](#)

December 16, 2015 at 6:32 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102034>)

Fortunately the auto.arima function allows us to model time series quite nicely though it is quite useful to know the basics. Here is some code I wrote on the same data

<http://rpubs.com/ajaydecis/ts> (<http://rpubs.com/ajaydecis/ts>)

**SRINI**[Reply](#)

December 18, 2015 at 10:21 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102135>)

Awesome Tavish! Short, crisp and absolutely crystal clear 😊

**TOUSIFAHAMED**[Reply](#)

December 20, 2015 at 6:56 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102253>)

Thanks for the post.

Awesome explanation..!

**MIKE RAI**[Reply](#)

December 20, 2015 at 5:54 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102271>)

Rohit,

Please be more specific, and provide the location of the discussion on Inkd, so that Tavish can respond appropriately..

**IAN M ([HTTP://WWW.THEORSOCIETY.COM](http://www.theorsociety.com))**[Reply](#)

December 26, 2015 at 3:54 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102595>)

The pairs of graphs introducing the concepts worked really well. I found the use of english letters for all the formulae clear.

**BA**[Reply](#)

December 28, 2015 at 10:01 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102682>)

Hi, thanks for the tutorial. I have just one comment for the identification of MA order. We have been taught that the length of the first line of the ACF curve is always equal to 1 [because it's $\text{cov}(X_t, X_t)/(\sigma(X_t)*\sigma(X_t) = 1]$. So we don't look at this line, we start counting after this line. If that's the case, your first MA example should be MA(1) instead of MA(2)

**CAROL ZHANG**[Reply](#)

January 1, 2016 at 9:20 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-102903>)

Hi Tavish,

One question about the ADF test.

```
adf.test(diff(log(AirPassengers)), alternative="stationary", k=0)
```

How shall we decide on the value for k? I tried to run another version with no specification of k value. And the default value used is k = 5 (aka. lag order = 5).

Many thanks!



ABDI KENESA

[Reply](#)

January 6, 2016 at 6:41 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-103194>)

Thank you.
Your article is great



STEVEN F CHAPMAN, PH.D.

[Reply](#)

February 18, 2016 at 12:56 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-105876>)

Is there any way we can get a PDF of this? I would like to use it to introduce my staff to trend analysis and some errors to look out for—



SURYA PRAKASH

[Reply](#)

March 11, 2016 at 6:45 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-107077>)

Why did we take d as 1 in this example?



PINTU SENGUPTA

[Reply](#)

September 8, 2016 at 9:44 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-115800>)

We have difference the series once and get to see that the trend is removed. Had the trend been still there we would have difference the series once again. This series did not require to be difference more than once; hence $d=1$.



GURU

[Reply](#)

March 19, 2016 at 1:24 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-107771>)

This article was very helpful

**SIDHRAJ**[Reply](#)

March 26, 2016 at 1:49 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-108335>)

why the author not answer the questions.....
this force us to look for better articles and doubt this one.

**AMY**[Reply](#)

April 18, 2016 at 4:28 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-109633>)

Please explain the parameters to this last line of code
`ts.plot(AirPassengers, 2.718^pred$pred, log = "y", lty = c(1,3))`

**ARATI**[Reply](#)

May 3, 2016 at 5:43 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-110387>)

Hi,
After you run this
`pred <- predict(APmodel, n.ahead=10*12)`

take a look at 'pred'

It is a list of 2 (pred and se – I assume these are predictions and errors.)

I would suggest using a name other than pred in the predict function to avoid confusion , I used the following

```
APforecast <- predict(APmodel, n.ahead=10*12)
```

So APforecast is a list of pred and se and we need to plot the pred values , ie `APforecast$pred`

Also we did the arima on log of AirPassengers, so the forecast we have got is actually log of the true forecast.

Hence we need to find the log inverse of what we have got.

ie. `log(forecast) = APforecast$pred`

so `forecast = e ^ APforecast$pred`

`e = 2.718`

If you find that confusing, I would suggest reading up on natural logarithms and their inverse

the `log = "y"` is to plot on a logarithmic scale – this is not needed, try the function without it and with and observe the results.

The lty bit I have not figured out yet. Drop it and try the ts.plot, it works fine.

**BRANDON**[Reply](#)

September 6, 2016 at 11:54 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-115700>)

Hey Amy, ts.plot() will plot several time series on the same plot. The first two entries are the two time series he's plotting. The last two entries are nice visual parameters (we'll come back to that). Clearly, this plots the AirPassengers time series in a dark, continuous line. The second entry is also a time series, but it is a little more confusing: "2.718^pred\$pred". First, you have to know what pred\$pred is. The function predict() here is a generic function that will work differently for different classes plugged into it (it says so if you type ?predict). The class we're working with is an Arima class. If you type ?predict.Arima you will find a good description of what the function is all about. predict.Arima() spits out something with a "pred" part (for predict) and a "se" part (for standard error). We want the "pred" part, hence pred\$pred. So, pred\$pred is a time series. Now, 2.718^pred\$pred is also. You have to remember that 2.718 is approximately the constant e, and then this makes sense. He's just undoing the log that he placed on the data when he created "fit".

As for the last two parameters, log = "y" sets the y-axis to be on a log scale. And finally, lty = c(1,3) will set the LineType to 1 (for solid) for the original time series and 3 (for dotted) for the predicted time series.

**MUSTAFA**[Reply](#)

April 21, 2016 at 12:45 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-109779>)

Thanks a lot! Very useful article.

**LUCA NICOLI**[Reply](#)

May 4, 2016 at 4:24 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-110439>)

Hi, It is very interesting.

Can you make the same example with Python code?

**REDDAIAH B N**[Reply](#)

May 24, 2016 at 1:23 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-111385>)

Hi Tavish,

Thank you very much for the nice explanation about time series using ARIMA.

However I have the following the queries regarding the analysis.

1.ACF and PACF are to find the p and q values as part of ARIMA? Is only ACF is not enough to find the p and q? If not can you explain the importance of PACF?

Thanks in advance.....:)



MUHAMMAD ARIF

[Reply](#)

June 1, 2016 at 7:01 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-111678>)

if non stationarity is present in data ,can we analyse that data



PARIND

[Reply](#)

June 6, 2016 at 12:00 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-111924>)

Hey Tavish, really enjoyed the content,

Just a small doubt: Can you please elaborate the covariance in stationary terms. I understand the covariance term, but here in time series, it is not coming to my mind. Can you please help me understand the third condition of stationary series i.e "The covariance of the i th term and the $(i + m)$ th term should not be a function of time." Please help me understand from data perspective e.g if i have sales data for each date. how can you explain covariance in real life example with daily sales data.



PARTH GERA

[Reply](#)

July 4, 2016 at 10:52 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-113042>)

Hi Tavish, Thanks a lot .This article was immensely helpful .

I just had one small issue.After the last step, If I want to extract the predicted values from the curve . How do we do that?



RAM

[Reply](#)

July 4, 2016 at 2:26 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-113066>)

@Parth,

You get the predicted values from the variable pred.

pred is a list with two items: pred and se. (prediction and standard error).

To see the predictions, use this command: `print(pred$pred)`



PARTH GERA

[Reply](#)

July 4, 2016 at 5:30 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-113074>)

Hi Ram,

Thanks for your help . Yeah, `print(pred$pred)` would give us log of the predicted values. `print(2.718^pred$pred)` would give us the actual predicted values.

Thanks



RAM

[Reply](#)

July 5, 2016 at 12:18 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-113085>)

Yes, if you use 'log' when creating the model, you will use antilog or exponent to get the predicted values. If you create a model without the log function, you will not use exponent to get the predicted values



MANPREET

[Reply](#)

August 1, 2016 at 10:14 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-114326>)

how to extract the data for the predicted and actual values from R



AKSHAY

[Reply](#)

August 8, 2016 at 9:49 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-114574>)

hello,

the data you used in your tutorial, AirPassengers, is already a time series object.

my question is, HOW can i make/prepare my own time series object?

i currently have a historical currency exchange data set, with first column being date, and the rest 20 columns

are titled by country, and their values are the exchange rate.

after i convert my date column into date object, when i use the same commands used in your tutorial, the results are funny.

for example, `start(data$Date)` will give me a result of:

```
[1] 1 1
```

and `frequency(data$Date)` will return:

```
[1] 1
```

can you please explain HOW to prepare our data accordingly so we can use the functions?

thank you!



BRANDON

[Reply](#)

September 7, 2016 at 12:03 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-115701>)

If you type in `?ts` then you should be on your way. You only need a (single) time series, a frequency, and a start date. The examples at the bottom of the documentation should be very helpful. I'm guessing you'd write something like `ts(your_timeseries_data, frequency = 365, start = c(1980, 153))` for instance if your data started on the 153rd day of 1980.



IMAMHIDAYAT

[Reply](#)

August 30, 2016 at 7:26 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-115316>)

Thank you very much...



RAM

[Reply](#)

September 10, 2016 at 12:11 am (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-115870>)

What is the format of your date value before you converted it ? If you post a few rows from your data, perhaps we can help.



VISHWANATH

[Reply](#)

September 17, 2016 at 11:09 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-116166>)

Thank you, It was very helpful for me 😊

**KEVIN**[Reply](#)

October 2, 2016 at 11:35 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-116697>)

Hi, thanks for the article! I'm still unclear how the parameters $(p,d,q) = (0,1,1)$ were found from the ACF and PCF. I understand d, but not p or q. What do you mean when you say 'cutting off'?

**AMY**[Reply](#)

October 5, 2016 at 5:18 pm (<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-116827>)

Hi Kevin,

ACF plot is a bar chart of the coefficients of correlation between a time series and lags of itself.

PACF plot is a plot of the partial correlation coefficients between the series and lags of itself.

To find p and q you need to look at ACF and PACF plots. The interpretation of ACF and PACF plots to find p and q are as follows:

AR (p) model: If ACF plot tails off* but PACF plot cut off** after p lags

MA(q) model: If PACF plot tails off but ACF plot cut off after q lags

ARMA(p,q) model: If both ACF and PACF plot tail off, you can choose different combinations of p and q , smaller p and q are tried.

ARIMA(p,d,q) model: If it's ARMA with d times differencing to make time series stationary.

Use AIC and BIC to find the most appropriate model. Lower values of AIC and BIC are desirable.

*Tails of mean slow decaying of the plot, i.e. plot has significant spikes at higher lags too.

**Cut off means the bar is significant at lag p and not significant at any higher order lags.

Here is a link that might help you understand the concept further <http://people.duke.edu/~rnau/arimrule.htm>
(<http://people.duke.edu/~rnau/arimrule.htm>)

Hope this helps.

**JOHNNY**[Reply](#)

[October 11, 2016 at 4:16 pm \(https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-117082\)](https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-117082)

Hi. Great article and I am working on a gforce (values + and -) dataset and am having trouble with the log function. NaNs produced and not sure how to go about addressing this.

Any help would be appreciated.



AMIT KUMAR JAIN

[Reply](#)

[November 19, 2016 at 5:56 pm \(https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-118561\)](https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-118561)

GREAT ARTICLE... THANK YOU TAVISH!!!!

One strong suggestion to Analytics Vidya. Please add a link of PDF downloads to these kind of articles (without advertisements) which for a person like me who is creating a repository of awesome articles to learn from will be really helpful!!!!



PARTH GERA

[Reply](#)

[November 19, 2016 at 7:22 pm \(https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-118562\)](https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-118562)

Hi Tavish, Great article. I had one doubt .In the last step , while fitting the arima model , you have used $\log(\text{AirPassengers})$ instead of $\text{diff}(\log(\text{AirPassengers}))$. Why is that so? $\log(\text{Airpassengers})$ isn't a stationary series , right?



DAZ

[Reply](#)

[May 20, 2017 at 3:51 am \(https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-128785\)](https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-128785)

Just an FYI for r-newbies. I don't think its mentioned above by to run `adf.test` you will need to install the `tseries` package.



GAURAV SHARMA

[Reply](#)

[April 29, 2018 at 2:24 am \(https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-152909\)](https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/#comment-152909)

It's handled by defining $c(0, 1, 1)$ while fitting. Here 1st 1 denote to differentiation, which will make series stationary.



(<https://www.analyticsvidhya.com/datahack-summit-2018/?>

utm_source=AVbannerdhslong)



POPULAR POSTS

24 Ultimate Data Science Projects To Boost Your Knowledge and Skills (& can be accessed freely)
(<https://www.analyticsvidhya.com/blog/2018/05/24-ultimate-data-science-projects-to-boost-your-knowledge-and-skills/>)

A Complete Tutorial to Learn Data Science with Python from Scratch
(<https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-scratch-2/>)

Essentials of Machine Learning Algorithms (with Python and R Codes)
(<https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/>)

Understanding Support Vector Machine algorithm from examples (along with code)

(<https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/>)

7 Types of Regression Techniques you should know!

(<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/>)

6 Easy Steps to Learn Naive Bayes Algorithm (with codes in Python and R)

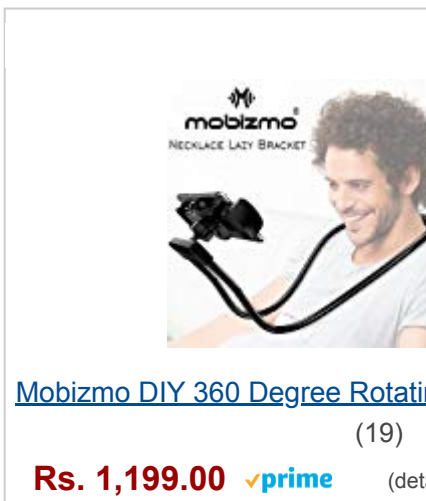
(<https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/>)

A comprehensive beginner's guide to create a Time Series Forecast (with Codes in Python)

(<https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/>)

Introduction to k-Nearest Neighbors: Simplified (with implementation in Python)

(<https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/>)



RECENT POSTS

A Practical Implementation of the Faster R-CNN Algorithm for Object Detection (Part 2 – with Python codes) (<https://www.analyticsvidhya.com/blog/2018/11/implementation-faster-r-cnn-python-object-detection/>)

NOVEMBER 4, 2018

Top 5 Machine Learning GitHub Repositories & Reddit Discussions (October 2018) (<https://www.analyticsvidhya.com/blog/2018/11/best-machine-learning-github-repositories-reddit-threads-october-2018/>)

NOVEMBER 1, 2018

An Introduction to Text Summarization using the TextRank Algorithm (with Python implementation)
(<https://www.analyticsvidhya.com/blog/2018/11/introduction-text-summarization-textrank-python/>)

NOVEMBER 1, 2018

DataHack Radio #13: Data Science and AI in the Oil & Gas Industry with Yogendra Pandey, Ph.D.
(<https://www.analyticsvidhya.com/blog/2018/10/datahack-radio-podcast-oil-gas-ai/>)

OCTOBER 29, 2018



([http://www.edvancer.in/certified-data-scientist-with-python-course?](http://www.edvancer.in/certified-data-scientist-with-python-course?utm_source=AV&utm_medium=AVads&utm_campaign=AVadsnonfc&utm_content=pythonavad)

[utm_source=AV&utm_medium=AVads&utm_campaign=AVadsnonfc&utm_content=pythonavad](http://www.edvancer.in/certified-data-scientist-with-python-course?utm_source=AV&utm_medium=AVads&utm_campaign=AVadsnonfc&utm_content=pythonavad))



([https://trainings.analyticsvidhya.com/courses/course-](https://trainings.analyticsvidhya.com/courses/course-v1:AnalyticsVidhya+DS101+2018T2/about?utm_source=AVStickybanner1)

[v1:AnalyticsVidhya+DS101+2018T2/about?utm_source=AVStickybanner1](https://trainings.analyticsvidhya.com/courses/course-v1:AnalyticsVidhya+DS101+2018T2/about?utm_source=AVStickybanner1))



(https://trainings.analyticsvidhya.com/courses/course-

v1:AnalyticsVidhya+CVDL101+CVDL101_T1/about?
utm_source=CV101AVBlogBanner&utm_medium=Stickybanner2utm_campaign=CV101Banner2

DATA SCIENTISTS

Post Jobs
(https://www.analyticsvidhya.com/corporate/)

Hackathon
(https://trainings.analyticsvidhya.com)

Our Team
(https://datahack.analyticsvidhya.com/)

Discussions
(https://discuss.analyticsvidhya.com)

Career
(https://www.analyticsvidhya.com/career/)

Contact Us
(https://www.analyticsvidhya.com/contact/)

Write for us
(https://www.analyticsvidhya.com/about-me/write/)

COMPANIES

Advertising
(https://www.analyticsvidhya.com/contact/)

Reach Us
(https://www.analyticsvidhya.com/contact/)

JOIN OUR COMMUNITY :

f
(https://www.facebook.com/analyticsvidhya/) 18524 Followers

in
(https://plus.google.com/Analyticsindia) 7513 Followers

G+
(https://plus.google.com/Analyticsindia) 3017 Followers

>