INTERVIEW QUESTIONS

For

TIME SERIES FORECASTING

ANALYTIXLABS

Website: www.analytixlabs.co.in

Email: info@analytixlabs.co.in

Time Series Analysis – Forecasting

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

Time Series forecasting & modeling plays an important role in data analysis. Time series analysis is a specialized branch of statistics used extensively in fields such as Econometrics & Operation Research. This skill test was conducted to test your knowledge of time series concepts.

Time Series: A time series is a data series consisting of several values over a time interval. e.g. daily BSE Sensex closing point, weekly sales and monthly profit of a company etc.

Typically, in a time series it is assumed that value at any given point of time is a result of its historical values. This assumption is the basis of performing a time series analysis.

Example: Suppose Mr. X starts his job in year 2010 and his starting salary was $5,000 per month. Every years he is appraised and salary reached to a level of $20,000 per month in year 2014. His annual salary can be considered a time series and it is clear that every year's salary is function of previous year's salary (here function is appraisal rating).

Time Series Modeling 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.

Components of a Time Series:
1. **Trend:** Series could be **constantly increasing or decreasing** or **first decreasing for a considerable time period and then decreasing**. This trend is identified and then removed from the time series in ARIMA forecasting process.

2. **Seasonality:** Repeating pattern with fixed period.
Example - Sales in festive seasons. Sales of Candies and sales of Chocolates peaks in every October Month and December month respectively every year in US. It is because of Halloween and Christmas falling in those months. The time-series should be de-seasonalized in ARIMA forecasting process.

3. **Random Variation (Irregular Component)**
This is the unexplained variation in the time-series which is totally random. Erratic movements that are not predictable because they do not follow a pattern.

Example - Earthquake

Terminologies related to Time Series

1. Stationary Series: A stationary series should have mean and variance of the series is constant over time. The series has to be stationary before building a time series with ARIMA. Most of the time series are non-stationary. If series is non-stationary, we need to make it stationary with detrending, differencing etc.

Why Stationary?
To calculate the expected value, we generally take a mean across time intervals. The mean across many time intervals makes sense only when the expected value is the same across those time periods. If the mean and population variance can vary, there is no point estimating by taking an average across time.

2. White Noise: A white noise process is one with a constant mean of zero, a constant variance and no correlation between its values at different times. White noise series exhibit a very erratic, jumpy, unpredictable behavior. Since values are uncorrelated, previous values do not help us to forecast future values.
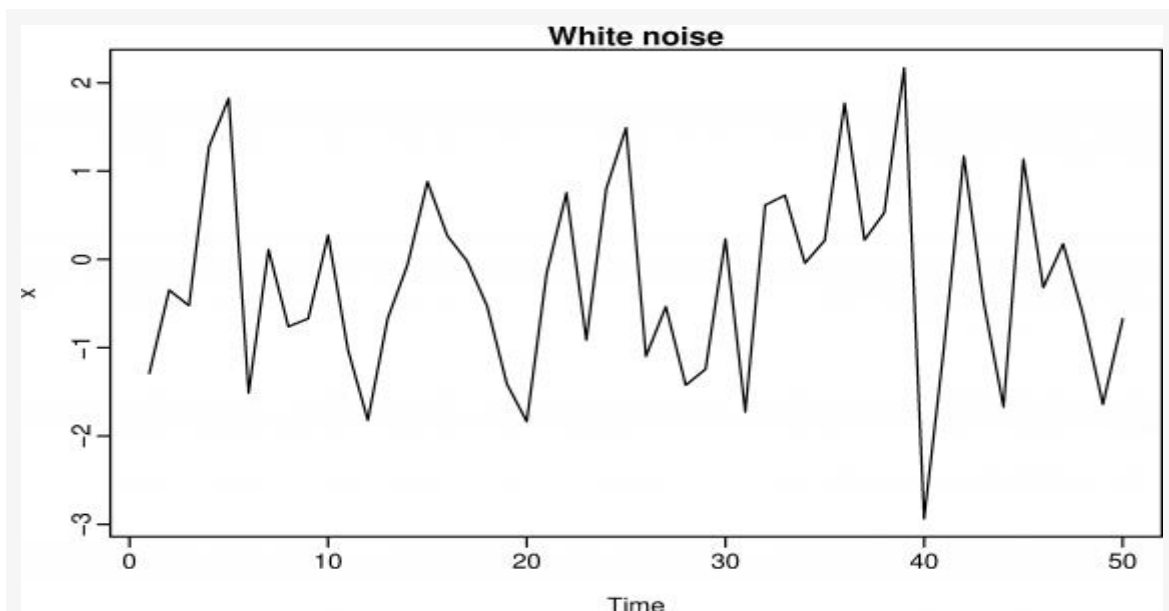

3. Autocorrelation: Autocorrelation refers to the correlation of a time series with its own past and future values. Autocorrelation is also sometimes called "lagged correlation" or "serial correlation".

4. Random Walk: In layman's term, it means past data provides no information about the direction of future movements.
It is called **random-walk-without-drift model**: it assumes that, at each point in time, the series merely takes a random step away from its last recorded position, with steps whose mean value is zero.

If the mean step size is some non-zero value α, the process is said to be a **random-walk-with -drift** (slow steady change) whose prediction equation is **Ŷt = Yt-1 + α**

A random walk process is non-stationary as its mean and variance increases with t.

Time Series : Random Walk
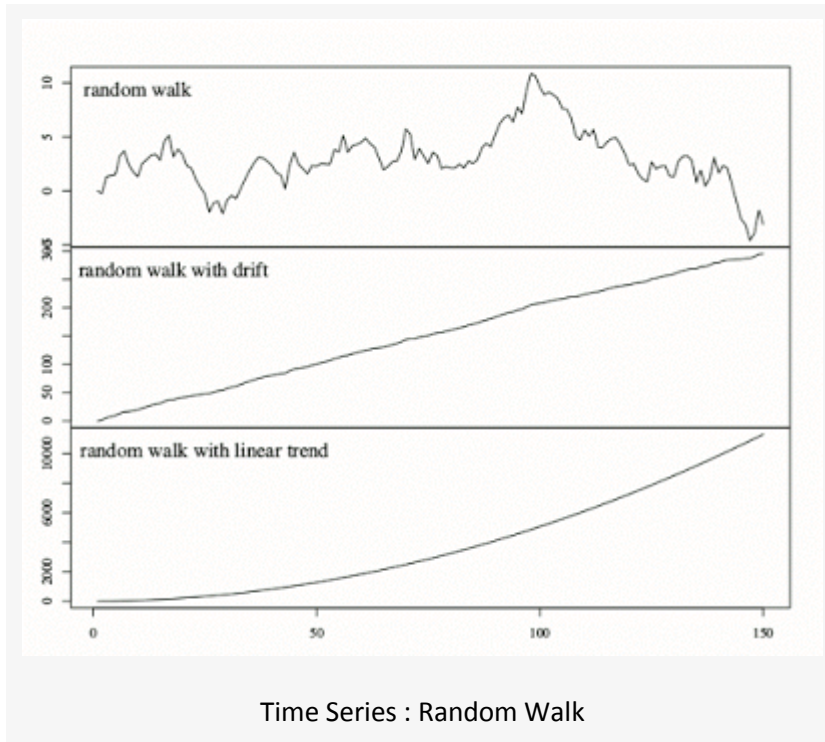
What is ARIMA (Box-Jenkins Approach) process?
ARIMA stands for Auto-Regressive Integrated Moving Average. It is also known as **Box-Jenkins approach**. It is one of the most popular techniques used for time series analysis and forecasting purpose.

We would cover ARIMA in a series of blogs starting from introduction, theory and finally the process of performing ARIMA.

Well, coming back to ARIMA, as its full form indicates that it involves two components:
1. **Auto-regressive component**
2. **Moving average component**
We would first understand these components one by one.

1. Auto-regressive Component

It implies relationship of a value of a series at a point of time with its own previous values. Such relationship can exist with any order of lag.
Lag: Lag is basically value at a previous point of time. It can have various orders as shown in the table below. It hints toward a pointed relationship.

Month	Sales	Lag(Sales)	Lag2(Sales)
Jan-15	100	-	-
Feb-15	103	100	-
Mar-15	98	103	100
Apr-15	116	98	103
May-15	120	116	98
Jun-15	100	120	116
Jul-15	130	100	120
Aug-15	133	130	100
Sep-15	104	133	130
Oct-15	137	104	133
Nov-15	143	137	104
Dec-15	105	143	137

Time Series : Lag

2. Moving average components

It implies the current deviation from mean depends on previous deviations. Such relationship can exist with any number of lags which decides the order of moving average.

Moving Average -

Moving Average is average of consecutive values at various time periods. It can have various orders as shown in the table below. It hints toward a distributed relationship as moving itself is derivative of various lags.

Month	Sales	Movave2(Sales)	Movave3(Sales)
Jan-15	100	100	100
Feb-15	103	101.5	101.5
Mar-15	98	100.5	100.33
Apr-15	116	107	105.67
May-15	120	118	111.33
Jun-15	100	110	112.00
Jul-15	130	115	116.67
Aug-15	133	131.5	121.00
Sep-15	104	118.5	122.33
Oct-15	137	120.5	124.67
Nov-15	143	140	128.00
Dec-15	105	124	128.33

Moving Average Explanation

Moving average is itself considered as one of the most rudimentary methods of forecasting. So if you drag the average formula in excel further (beyond Dec-15), it would give you forecast for next month.
Both Auto-regressive (lag based) and moving average components in conjunction are used by ARIMA technique for forecasting a time series.

ARIMA Modeling Steps:

1. Plot the time series data
2. Check volatility - Run Box-Cox transformation to stabilize the variance
3. Check whether data contains seasonality. If yes, two options - either take seasonal differencing or fit seasonal arima model.
4. If the data are non-stationary: take first differences of the data until the data are stationary
5. Identify orders of p, d and q by examining the ACF/PACF
6. Try your chosen models, and use the AICC/BIC to search for a better model.
7. Check the residuals from your chosen model by plotting the ACF of the residuals, and doing a portmanteau test of the residuals. If they do not look like white noise, try a modified model.
8. Check whether residuals are normally distributed with mean zero and constant variance
9. Once step 7 and 8 are completed, calculate forecasts

Many of the simple time series models are special cases of ARIMA Model

1. Simple Exponential Smoothing ARIMA(0,1,1)
2. Holt's Exponential Smoothing ARIMA(0,2,2)
3. White noise ARIMA(0,0,0)
4. Random walk ARIMA(0,1,0) with no constant
5. Random walk with drift ARIMA(0,1,0) with a constant
6. Auto regression ARIMA(p,0,0)
7. Moving average ARIMA(0,0,q)

STEP BY STEP PROCESS OF ARIMA MODEL BUILDING IN SAS:

Data File: *SASHELP.AIR*

Data Preparation Steps For ARIMA Modeling

1. Check if there is variance that changes with time - **Volatility.** For ARIMA, the volatility should not be very high.
2. If the **volatility is very high**, we need to make it non-volatile.
3. Check for **Stationary** - a series should be stationary before performing ARIMA.
4. If data is **non-stationary**, we need to make it stationary.
5. Check for **Seasonality** in the data

Step 1 : Check the series: we first plot the time series and have a cursory look upon it. It can be done in SAS using following code:

```
proc sgplot data = sashelp.AIR;
series x = date Y = AIR;
run;
```

It would give you the following plot in the result window:

SAS : Time Series Modeling
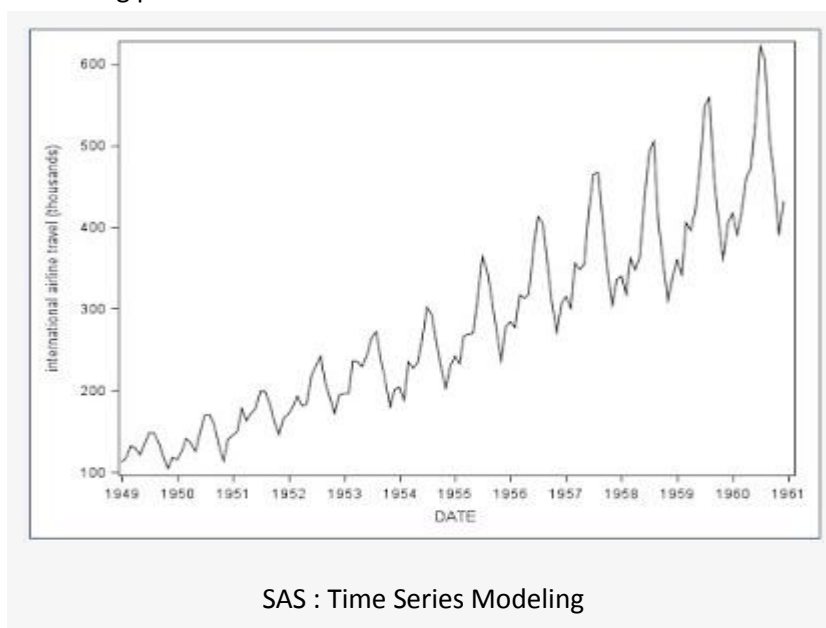
It is clear from the chart above that the series of AIR is having an **increasing trend and consistent pattern over time**. The peaks are at a constant time interval which is indicative of presence of seasonality in the series.

*This is a **non-stationary series** for sure and hence we need to make it stationary first.*

Practically, ARIMA works well in case of such types of series with a clear trend and seasonality. We first separate and capture the trend and seasonality component off the time-series and we are left with a series i.e.

stationary. This stationary series is forecasted using ARIMA and then final forecasting incorporates the pre-captured trend and seasonality.

Step 2: Check the volatility of the series

Volatility is the degree of variation of a time-series over time. For ARIMA, the volatility should not be very high. For checking the volatility of time-series, we do a scatter plot using the following SAS code:

Proc gplot data=SAShelp.AIR;
plot Date * AIR;
Run;

It would give you the following plot in the result window:

Check the volatility of Series
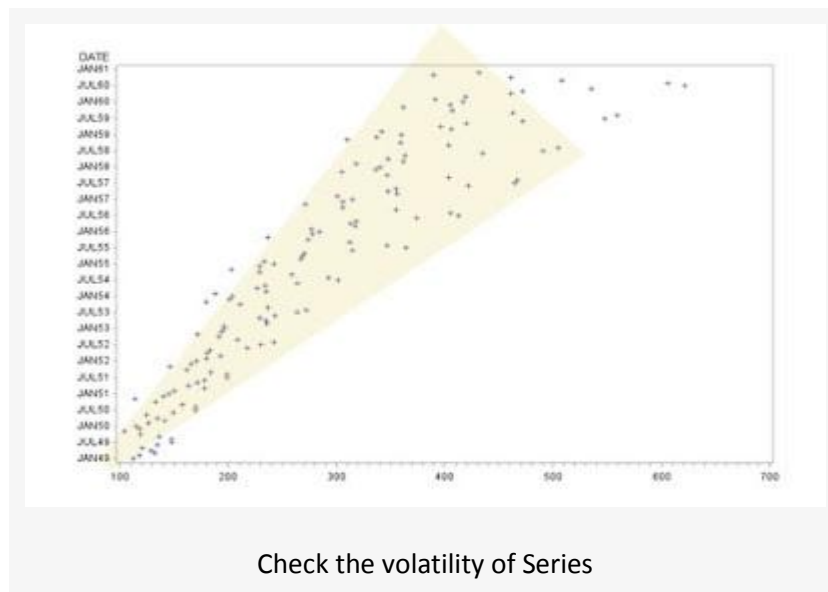
The highlighted area is showing the diverging pattern (Fan shaped) of the scatter plot and hence depicting that the data is volatile. Ideally, the highlighted pattern should be parallel for ARIMA modeling.

Step 3: Treatment of Volatile Series

We need to make the series non-volatile and move ahead. We would transform the AIR series and remove volatility. Generally a hit and trail method for transformation is used, but we would suggest to not to waste your time.

Box-Cox Transformation can be used to help you out and recommend the suitable transformation.

Proc Transreg Data = sashelp.AIR;
Model BOXCOX (AIR) = Identity(Date);
Run;

You get following plot along with Lamba value, which is "0" in this case.

Box-Cox Analysis for AIR
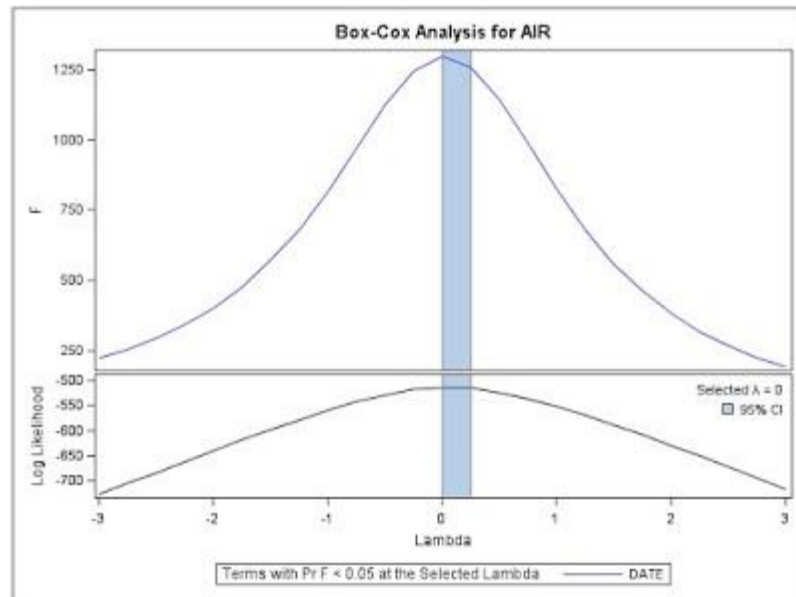
Now based on this Lambda value, you can decide the transformation. Take help from the table provided below.

Common Box-Cox Transformations

Lambda	Suitable Transformation
-2	$Y^{-2} = 1/Y^2$
-1	$Y^{-1} = 1/Y^1$
-0.5	$Y^{-0.5} = 1/(Sqrt(Y))$
0	log(Y)
0.5	$Y^{0.5} = Sqrt(Y)$
1	$Y^1 = Y$
2	Y^2

Box cox Transformation

In our case, it is suggesting a log transformation, so we do the same. We create a new variable (Log_AIR).

```
Data Masterdata;
Set  SAShelp.AIR;
Log_AIR = log(AIR);
Run;
```

We can check the volatility again of the transformed series, just to be sure, using scatter plot as elaborated above.

Step 4: Check For Non-Stationarity

Now on the transformed series, we check whether the series is stationary or non-stationary.

For performing ARIMA, a series should be stationary, however if the series is non-stationary, we make it stationary.

Rather than identifying the series's stationarity visually, we can use **Augmented Dickey-Fuller Unit Ratio Test** for the same.

Unit Root - Homogeneous Non-Stationarity Data

Dickey-Fuller test: *The Dickey-Fuller test is used to test the null hypothesis that the time series exhibits a lag d unit root against the alternative of stationarity.*

Null Hypothesis: Non-Stationary
Alternative Hypothesis: Stationary

There are three types by which you can calculate test statistics of dickey-fuller test.

1. **Zero Mean - No Intercept.** Series is a random walk **without drift**.
2. **Single Mean - Includes Intercept.** Series is a random walk **with drift.**
3. **Trend - Includes Intercept and Trend.** Series is a random walk **with linear trend**.

All the above test statistics are computed from the OLS regression model

Drawback of ADF Test

Uncertainty about what test version to use, i.e. about including the intercept and time trend terms.

Inappropriate exclusion or inclusion of these terms substantially affects test reliability.

Using of prior knowledge (for instance, as result of visual inspection of a given time series) about whether the intercept and time trend should be included is the mostly recommended way to overcome the difficulty mentioned.

We run **Proc ARIMA** with **Stationarity = (ADF)** option to do so:

```
PROC ARIMA DATA= Masterdata;
IDENTIFY VAR = log_Air STATIONARITY= (ADF);
RUN;
QUIT;
```

There are many outputs of the above code, a part of which is used for checking stationarity:

Important Note: *Check **Tau Statistics (Pr < Tau)** in **ADF Unit Root Tests** table. It should be less than 0.05 to say data is stationary at 5% level of significance.*

Step 5: Make Non-Stationary Data Stationary

ARIMA : Check Stationary
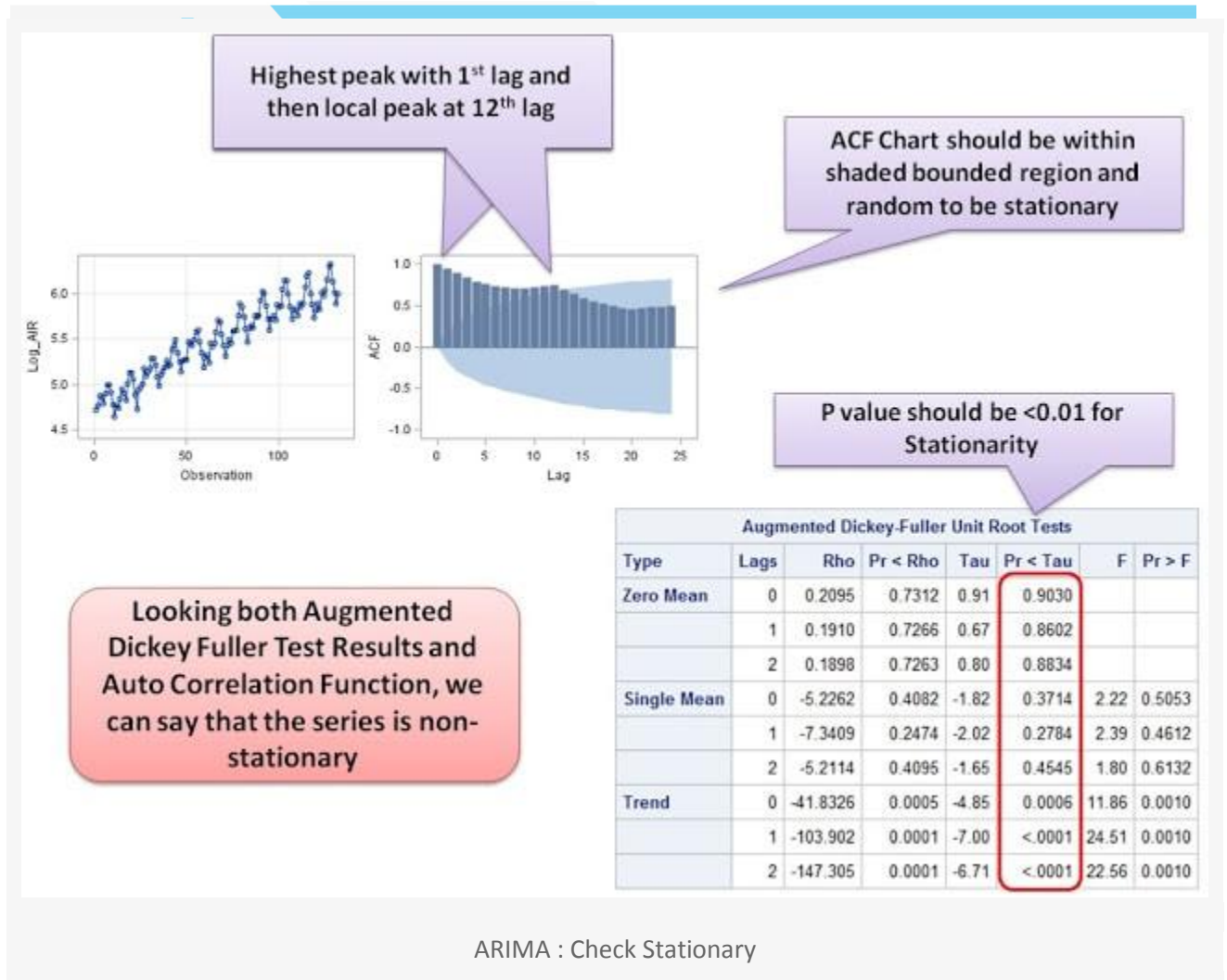
Post establishing the non-stationarity of the series, we need to make the series stationary. Differencing process is used for making the series stationary.

Differencing: Transformation of the series to a new time series where the values are the differences between consecutive values

Differencing Procedure may be applied consecutively more than once, giving rise to the **"first differences"**, "**second differences"**, etc.

Differencing Orders:
1st order   : $\nabla xt = xt - xt\text{-}1$. **For eg. Sales - lag1 (Sales)**
2nd order : $\nabla 2xt = (\nabla xt - \nabla xt\text{-}1) = xt - 2xt\text{-}1 + xt\text{-}2$

It is unlikely that more than two differencing orders would ever be required.
Note: If there is a physical explanation for a trend or seasonal cycle: use **regression** to make series stationary.

While we have run the code above, we have got "Autocorrelation Check for White Noise" along with "Augmented Dickey-Fuller Unit Root Tests".

Looking at "Autocorrelation Check for White Noise", we decide the order(s) of differencing required.

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	638.37	6	<.0001	0.954	0.899	0.851	0.808	0.779	0.756
12	1157.62	12	<.0001	0.738	0.727	0.734	0.744	0.758	0.762
18	1521.94	18	<.0001	0.717	0.663	0.618	0.576	0.544	0.519
24	1785.32	24	<.0001	0.501	0.490	0.498	0.506	0.517	0.520

Stationary : Order of Differencing

The first row of the above autocorrelation matrix shows correlation of time-series with 1st to 6th lags, second row show the same for 7th to 12th lags...and so on...
We can see that in above matrix the highest auto-correlation exists with 1st lag, it starts decreasing but again increases to attain a local peak at 12th lag.

Step 6: Check Seasonality
Highest Correlation with 1st Lag indicates towards the presence of trend and that with **12th lag indicates an annual seasonality.** Hence we need to do differencing at first and Twelfths orders.

We perform differencing and check the stationarity again.
                PROC ARIMA DATA= masterdata ;
                IDENTIFY VAR = Log_Air (1,12) STATIONARITY= (ADF) ;
                RUN;
We have used 1 and 12 in bracket to define the 1st and 12th order of differencing.

Check whether data is stationary
Check Tau Statistics (Pr < Tau) in ADF Unit Root Tests table again and see if the value <0.05 to say data is stationary at 5% level of significance.

How this differencing actually worked:
1. First order (1) Differencing removes the trend, but Seasonality still exists.
2. Second Order (12) Differencing removes the seasonality.

Step 7: Split Data into Training and Validation
Now we can break the data into **Training and Validation** samples. We cannot use **random sampling like we do in regression models** to split the data. Instead, we can use recent data for validation and remaining data be used to train the model. We would develop ARIMA model and forecast on Training part and would check the results on Validation part.
                Data Training Validation;

Set Masterdata;
If date >= '01Jan1960'd then output Validation;
Else output Training;
Run;

Step -8 **: Model Identification**

The order of an ARIMA (autoregressive integrated moving-average) model is usually denoted by the notation ARIMA(p,d,q) or it can be read as AR(p) , I(d), MA(q)

1. **p =** Order of Autoregression (Individual values of time series can be described by linear models based on preceding observations. For instance: x(t) = 3 x(t-1) - 4 x(t-2))
2. **d =** Order of differencing (No. of times data to be differenced to become stationary)
3. **q =** Order of Moving Average (Number of lagged forecast errors in the prediction equation. Past estimation or forecasting errors are taken into account when estimating the next time series value. The difference between the estimation x(t) and the actually observed value x(t) is denoted ε(t). For instance: x(t) = 3 ε(t-1) - 4 ε(t-2).)

Many of the simple time series models are special cases of ARIMA Model

1. Simple Exponential Smoothing ARIMA(0,1,1)
2. Holt's Exponential Smoothing ARIMA(0,2,2)
3. White noise ARIMA(0,0,0)
4. Random walk ARIMA(0,1,0) with no constant
5. Random walk with drift ARIMA(0,1,0) with a constant
6. Autoregression ARIMA(p,0,0)
7. Moving average ARIMA(0,0,q)

We can do the model identification in two ways:

1. Using ACF and PACF Functions

2. Using Minimum Information Criteria Matrix

Autocorrelation Function (ACF): Autocorrelation is a correlation coefficient. However, instead of correlation between two different variables, the correlation is between two values of the same variable at times Xt and Xt-h. Correlation between two or more lags.

Partial Autocorrelation Function (PACF): For a time series, the partial autocorrelation between xt and xt-h is defined as the conditional correlation between xt and xt-h, conditional on xt-h+1, ... , xt-1, the set of observations that come between the time points t and t–h.

ARIMA Procedure
identify var=VariableY(PeriodsOfDifferencing);
estimate p=OrderOfAutoregression q=OrderOfMovingAverage;

where VariableY is modeled as ARIMA(p,d,q) with p = OrderOfAutoregression, d = the order of differencing (determined from PeriodsOfDifferencing), and q = OrderOfMovingAverage.

In first method we check ACF and PACF plots.

P = Partial Auto-correlation Function

Q = Auto-correlation Function

As at first lag Bar is crossing bound, hence Q = 1 should be considered. It acts as Moving Average Component.
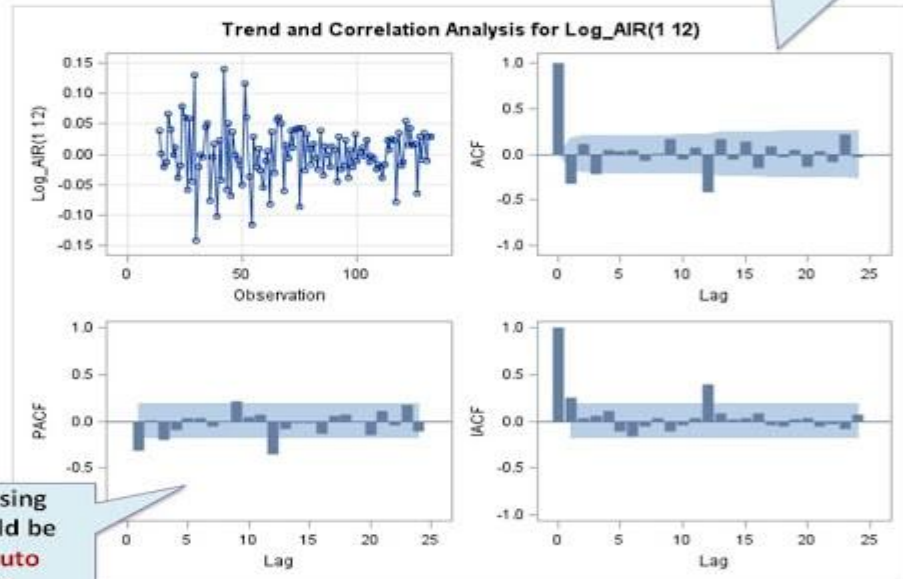
As at first lag Bar is crossing bound, hence P = 1 should be considered. It acts as Auto Regressive Component.

ARIMA - ACF PACF

Using these identified p and q values, we run ARIMA model.

```
PROC ARIMA DATA= Training ;
IDENTIFY VAR = Log_Air(1,12) ;
ESTIMATE P =1 Q =1 OUTSTAT= stats ;
Forecast lead=12 interval = month id = date
out = result;
RUN;
```

Minimum Information Criteria Matrix approach:

A **MINIC** table is then constructed using BIC(m,j) where m=pmin,.......pmax and j=qmin....qmax.

	MA (q)			
AR (p)	0	1	2	3
0	$BIC(0,0)$	$BIC(0,1)$	$BIC(0,2)$	$BIC(0,3)$
1	$BIC(1,0)$	$BIC(1,1)$	$BIC(1,2)$	$BIC(1,3)$
2	$BIC(2,0)$	$BIC(2,1)$	$BIC(2,2)$	$BIC(2,3)$
3	$BIC(3,0)$	$BIC(3,1)$	$BIC(3,2)$	$BIC(3,3)$

ARIMA Orders

We run following code first to get **MINIC**:
PROC ARIMA DATA= Training;
IDENTIFY VAR = Log_Air(1,12) MINIC;
RUN;
It would give you the matrix given below. Find the **minimum value** (largest negative) point in the matrix.

Minimum Information Criterion						
Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	-6.24946	-6.32141	-6.30667	-6.32826	-6.29985	-6.27302
AR 1	-6.33466	-6.29602	-6.28026	-6.29538	-6.26474	-6.23915
AR 2	-6.32028	-6.28278	-6.25595	-6.25609	-6.22534	-6.2063
AR 3	-6.3503	-6.3141	-6.27621	-6.24523	-6.24194	-6.22247
AR 4	-6.33057	-6.29054	-6.25187	-6.25848	-6.21998	-6.1958
AR 5	-6.30796	-6.26784	-6.22782	-6.23527	-6.19898	-6.165

Error series model: AR(9)
Minimum Table Value: BIC(3,0) = -6.3503

ARIMA : MIC

Now we consider the maximum of P(3) and Q(0) suggested by MINIC which is max(3,0) = 3 in this case. And then we iterate ARIMA model for P = 0 to 3 to Q = 0 to 3 (Except 0,0).

%Macro top_models;

%do p = 0 %to 3 ;
%do q = 0 % to 3 ;

PROC ARIMA DATA= test ;
IDENTIFY VAR = Log_Air(1,12) ;
ESTIMATE P = &p. Q =&q. OUTSTAT= stats_&p._&q. ;
Forecast lead=12 interval = month id = date
out = result_&p._&q.;
RUN;
Quit;

data stats_&p._&q.;
set   stats_&p._&q.;
p = &p.;
q = &q.;
Run;

```
data result_&p._&q.;
set   result_&p._&q.;
p = &p.;
q = &q.;
Run;

%end;
%end;

Data final_stats ;
set %do p = 0  %to 3 ;
%do q = 0 % to 3 ;
stats_&p._&q.
%end;
%end;;
Run;
Data final_results ;
set %do p = 0  %to 3 ;
%do q = 0 % to 3 ;
result_&p._&q.
%end;
%end;;
Run;

%Mend;
%top_models
```

/ Then to calculate the mean of AIC and SBC*/*
proc sql;
create table final_stats_1 as select p,q, sum(_VALUE_)/2 as mean_aic_sbc from final_stats
where _STAT_ in ('AIC','SBC')
group by p,q
order by mean_aic_sbc;
quit;
Save AIC and SBC values of all the iterations and choose top 5-7 models with minimum mean(AIC,SBC) values.

Now for all these selected models selected using AIC and SBC average, we calculate MAPE on validation data. We run the ARIMA on validation data with all selected P and Q.

Mean Squared Percentage Error (MAPE) for each model: MAPE = Abs(Actual – Predicted) / Actual *100

Use the following code to calculate MAPE :

```
Proc SQL;
create table final_results_1 as select a.p, a.q, a.date,a.forecast, b.log_air
from final_results as a join validation as b
on a.date = b.date;
quit;

Data Mape;
set final_results_1 ;
Ind_Mape = abs(log_air - forecast)/ log_air;
Run;

Proc Sql;
create table mape as select p, q, mean(ind_mape) as mape from mape
group by p, q
order by mape ;
quit;
```

Results:

p	q	mean_aic_sbc
2	3	-414.7
3	3	-414.4
0	1	-407.3
1	0	-407.0
0	3	-405.4
3	0	-405.1
0	2	-404.0
3	2	-403.9
1	1	-403.8
2	0	-403.6
1	3	-402.2
2	1	-401.8
1	2	-401.6
2	2	-400.5
3	1	-399.7
0	0	-397.8

Models selected on the basis of mean of AIC and SBC

| p	q	mape
0	3	1.0%
1	3	1.1%
2	2	1.1%
0	2	1.2%
0	1	1.2%
2	3	1.2%
3	0	1.2%
1	2	1.3%
3	1	1.3%
1	0	1.3%
3	3	1.3%
2	0	1.3%
3	2	1.3%
1	1	1.3%
2	1	1.4%
0	0	1.4%

ARIMA : MAPE

Model with least MAPE is finally your climax model which is **p= 0, q=3;**

STEP BY STEP PROCESS OF ARIMA MODEL BUILDING IN R:

Data Set Description: *Manufacturer's stocks of evaporated and sweetened condensed milk (case goods), Jan 1971 – Dec 1980*

```
library(forecast)
library(fpp)
# Plot time series data
tsdisplay(condmilk)
```


Time Series Plot
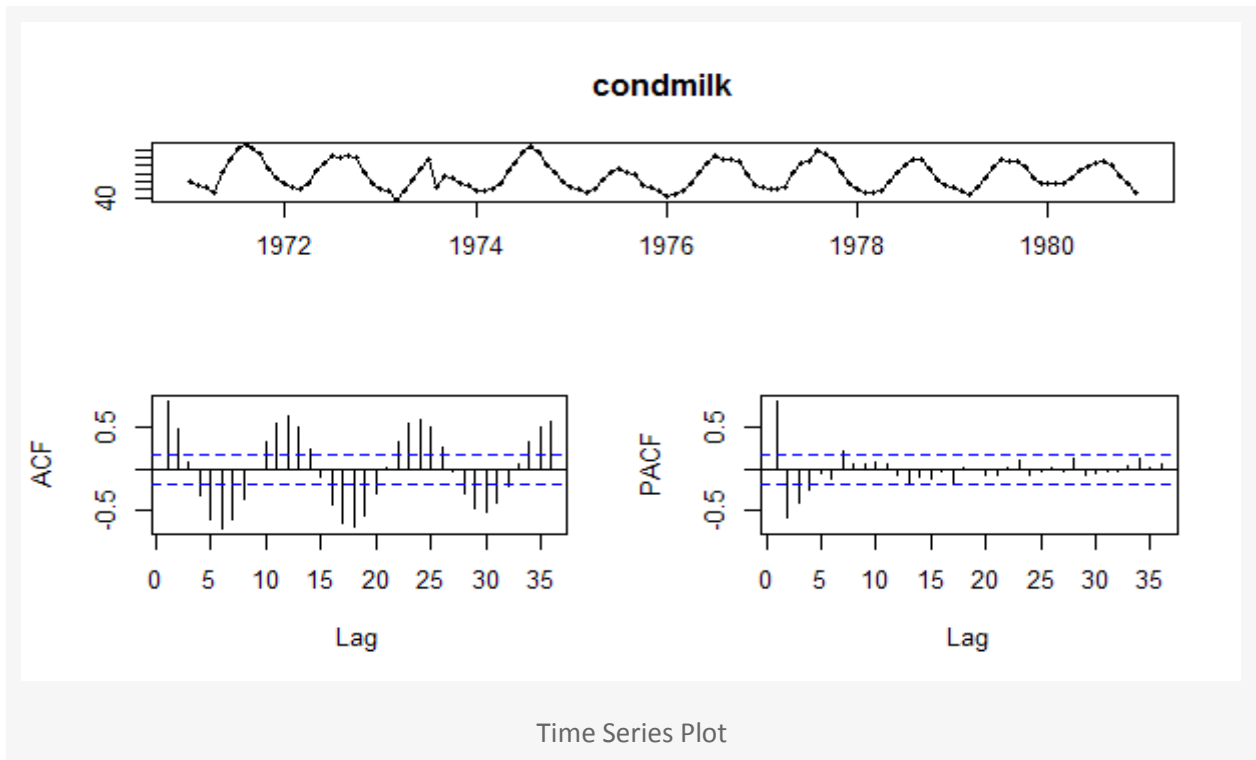
Step I: Check Volatility: If the data show different variation at different levels of the series, then a transformation can be beneficial. Apply box cox transformation to find the best transformation technique to stabilize the variance.

Lambda values:

$\lambda = 1$ (No substantive transformation)

$\lambda = 0.5$ (Square root plus linear transformation)

$\lambda = 0$ (Natural logarithm)

$\lambda = -1$ (Inverse plus 1)

Note : InvBoxCox() function reverses the transformation

R Code : Check Volatility

```
lambda = BoxCox.lambda(condmilk)
tsdata2 = BoxCox(condmilk, lambda=lambda)
```

tsdisplay(tsdata2)

Step 2: How to detect Seasonality
Seasonality usually causes the series to be non stationary because the average values at some particular times within the seasonal span (months, for example) may be different than the average values at other times.
R Code: Detect Seasonality

seasonplot(condmilk)
monthplot(condmilk)

How to treat Seasonality
1. Seasonal differencing: It is defined as a difference between a value and a value with lag that is a multiple of S. With S = 4, which may occur with quarterly data, a seasonal difference is (1-B4)xt = xt - xt-4.

2. Differencing for Trend and Seasonality: When both trend and seasonality are present, we may need to apply both a non-seasonal first difference and a seasonal difference.

3. Fit Seasonal ARIMA Model: The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. One shorthand notation for the model is
ARIMA(p, d, q) × (P, D, Q)S

with p = non-seasonal AR order, d = non-seasonal differencing, q = non-seasonal MA order, P = seasonal AR order, D = seasonal differencing, Q = seasonal MA order, and S = time span of repeating seasonal pattern.

Step 3: Detect Non-Stationary Data
The stationarity of the data can be known by applying Unit Root Tests - Augmented Dickey–Fuller test (ADF), Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test.

Augmented Dickey–Fuller test (ADF): *The null-hypothesis for an ADF test is that the data are non-stationary. So p-value greater than 0.05 indicates non-stationarity, and p-values less than 0.05 suggest stationarity.*

KPSS Test: *In this case, the null-hypothesis is that the data are stationary. In this case, p-value less than 0.05 indicates non-stationary series and p-value greater than 0.05 indicates stationary series.*

R Code: Detect Non-Stationary Data

```
 # Unit Ratio Tests
library(tseries)
adf = adf.test(tsdata2)
kpss = kpss.test(tsdata2)
adf
kpss
```
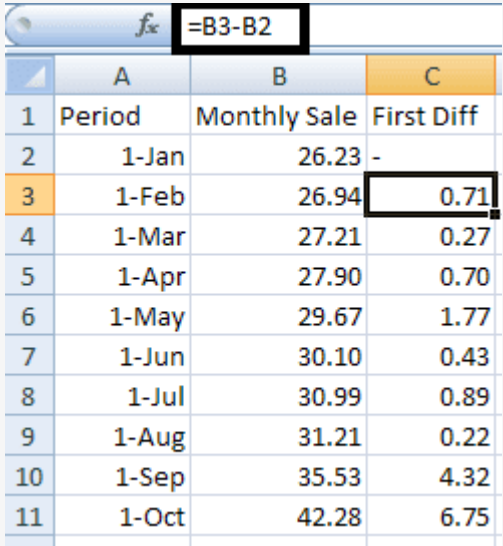*Since p-value of KPSS (0.1) is greater than 0.05, it indicates series is stationary. The p-value of ADF also indicates stationary series.*

Q. How to treat Non-Stationary Data
In this example, the series is already stationary so we don't need to make any treatment. If the data is non-stationary, then we use **Differencing** - computing the differences between consecutive observations.

First difference of a time series
It is the difference of the current value from the lagged value. The first difference of Y at period t is equal to Yt-Yt-1. The calculation is shown below in the image.

	A	B	C
		fx =B3-B2	
1	Period	Monthly Sale	First Diff
2	1-Jan	26.23	-
3	1-Feb	26.94	0.71
4	1-Mar	27.21	0.27
5	1-Apr	27.90	0.70
6	1-May	29.67	1.77
7	1-Jun	30.10	0.43
8	1-Jul	30.99	0.89
9	1-Aug	31.21	0.22
10	1-Sep	35.53	4.32
11	1-Oct	42.28	6.75

First Differencing Calculation

Use ndiffs(), diff() functions to find the number of times differencing needed for the data & to difference the data respectively.

R Code : Treat Non- Stationary Data
> *# Number of Difference Required to make data stationary*
> *ndiffs(tsdata2)*

In this example, series is stationary. Hence, we don't need to perform treatment to make it stationary.

The code below is just for demonstration. It does not hold for this example.
> *tsdata3 = diff(tsdata2, differences = 1)*
> *plot.ts(tsdata3)*

Step 4 : Model Identification and Estimation
We can do the model identification in two ways:
1. Using ACF and PACF Functions
2. Using Minimum Information Criteria Matrix

Method I: ACF and PACF Functions

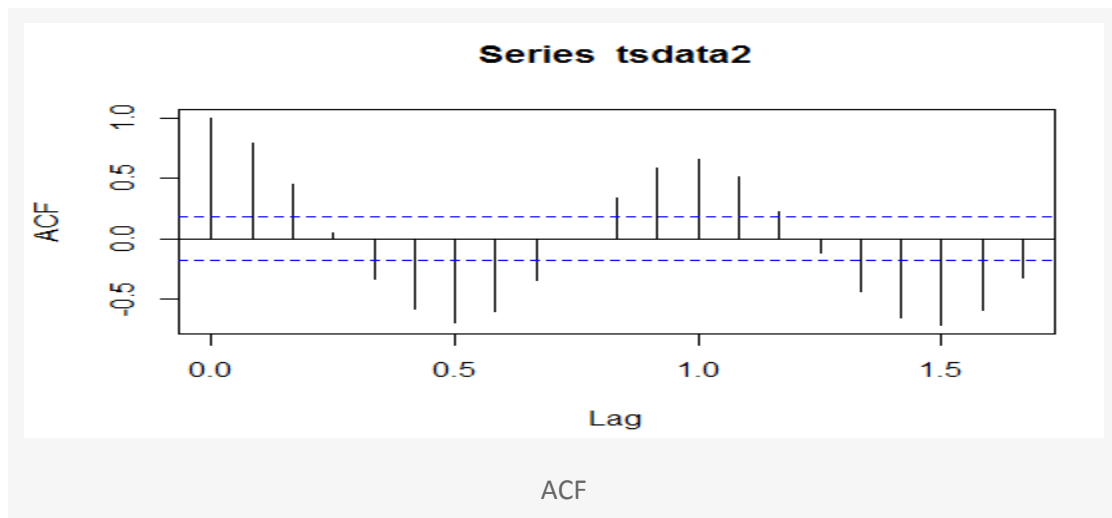**Autocorrelation Function (ACF)**

Autocorrelation is a correlation coefficient. However, instead of correlation between two different variables, the correlation is between two values of the same variable at times Xt and Xt-h. Correlation between two or more lags.

*If the autocorrelation at lag 1 exceeds the significance bounds, **set q = 1***

If the time series is a moving average of order 1, called a MA(1), we should see only one significant autocorrelation coefficient at lag 1. This is because a MA(1) process has a memory of only one period. If the time series is a MA(2), we should see only two significant autocorrelation coefficients, at lag 1 and 2, because a MA(2) process has a memory of only two periods.
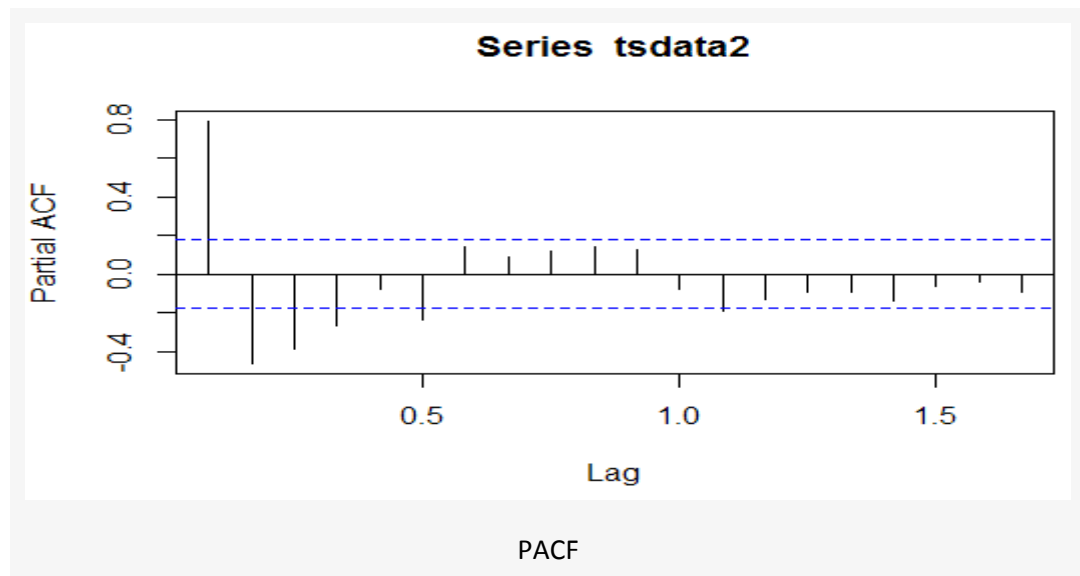
R Code : ACF

acf(tsdata2, lag.max = 20)

ACF

Partial Autocorrelation Function (PACF): For a time series, the partial autocorrelation between xt and xt-h is defined as the conditional correlation between xt and xt-h, conditional on xt-h+1, ... , xt-1, the set of observations that come between the time points t and t–h.

*If the partial autocorrelation at lag 1 exceeds the significance bounds, **set p = 1***

If the time-series has an autoregressive order of 1, called AR(1), then we should see only the first partial autocorrelation coefficient as significant. If it has an AR(2), then we should see only the first and second partial autocorrelation coefficients as significant.

R Code : PACF

pacf(tsdata2, lag.max = 20)

PACF

Method II : Minimum AIC / BIC Criteria

Fit a series of ARIMA models with combinations of p, d and q and select the model having minimum AIC / BIC.

R Code: Automatic Selection Algorithm

```
#Automatic Selection Algorithm - Fast
auto.arima(tsdata2, trace= TRUE, ic ="aicc", approximation = FALSE)
#Auto Algorithm - Slow but more accurate
auto.arima(tsdata2, trace= TRUE, ic ="aicc", approximation = FALSE, stepwise = FALSE)
```

Final Model

finalmodel = arima(tsdata2, order = c(0, 0, 3), seasonal = list(order = c(2,0,0), period = 12))
summary(finalmodel)

Compare Multiple Models

AIC(arima(tsdata2, order = c(1, 0, 0), seasonal = list(order = c(2,0,0), period = 12)),
  arima(tsdata2, order = c(2, 0, 0), seasonal = list(order = c(2,0,0), period = 12)),
  arima(tsdata2, order = c(0, 0, 3), seasonal = list(order = c(2,0,0), period = 12)))

Residual Diagnostics

#1. Residuals are Uncorrelated (White Noise)
#2. Residuals are normally distributed with mean zero
#3. Residuals have constant Variance

R Code :
```
# Check whether the residuals look like white noise (Independent)
# p>0.05 then the residuals are independent (white noise)
tsdisplay(residuals(finalmodel))
Box.test(finalmodel$residuals, lag = 20, type = "Ljung-Box")
# p-values shown for the Ljung-Box statistic plot are incorrect so calculate
#critical chi squared value
# Chi-squared 20 d.f. and critical value at the 0.05
qchisq(0.05, 20, lower.tail = F)
# Observed Chi-squared 13.584 < 31.41 so we don't reject null hypothesis
# It means residuals are independent or uncorrelated (white noise) at lags 1-20.
# whether the forecast errors are normally distributed
qqnorm(finalmodel$residuals); qqline(finalmodel$residuals) # Normality Plot
```

How to choose the number of lags for the Ljung-Box test
For non-seasonal time series,

Number of lags to test = minimum (10, length of time series / 5)
or simply take 10

For seasonal time series,
Number of lags to test = minimum (2m, length of time series / 5)
where, m = period of seasonality
or simply take 2m

Forecasting
```
        # predict the next 5 periods
        Forecastmodel = forecast.Arima(finalmodel, h = 5, lambda = lambda)
```
Note : If lambda specified, forecasts back-transformed via an inverse Box-Cox transformation.

If you have a fitted arima model, you can use it to forecast other time series.
```
inpt = arima(newdata, model=Forecastmodel)
```

Notes:
How auto.arima function works?
```
        auto.arima(kingsts, approximation=FALSE, start.p=1, start.q=1, trace=TRUE, seasonal=TRUE)
```

1. The number of differences d is determined using repeated KPSS tests.
2. The values of p and q are then chosen by minimizing the AIC after differencing the data d times. Rather than considering every possible combination of p and q, the algorithm uses a stepwise search to traverse the model space.

(a) The best model (with smallest AICc) is selected from the following four:
ARIMA(2,d,2),
ARIMA(0,d,0),
ARIMA(1,d,0),
ARIMA(0,d,1).

If d=0 then the constant c is included; if d≥1 then the constant c is set to zero. This is called the "current model".

(b) Variations on the current model are considered:
vary p and/or q from the current model by ±1;
include/exclude c from the current model.

The best model considered so far (either the current model, or one of these variations) becomes the new current model.

(c) Repeat Step 2(b) until no lower AICc can be found.

Objective Type Questions:

Q. Which of the following is an example of time series problem?

1. Estimating number of hotel rooms booking in next 6 months.

2. Estimating the total sales in next 3 years of an insurance company.

3. Estimating the number of calls for the next one week.

A) Only 3

B) 1 and 2

C) 2 and 3

D) 1 and 3

E) 1,2 and 3

Solution: (E)

All the above options have a time component associated.

Q. Which of the following is not an example of a time series model?

A) Naive approach

B) Exponential smoothing

C) Moving Average

D) None of the above

Solution: (D)

Naïve approach: Estimating technique in which the last period's actuals are used as this period's forecast, without adjusting them or attempting to establish causal factors. It is used only for comparison with the forecasts generated by the better (sophisticated) techniques.

In exponential smoothing, older data is given progressively-less relative importance whereas newer data is given progressively-greater importance.

In time series analysis, the moving-average (MA) model is a common approach for modeling univariate time series. The moving-average model specifies that the output variable depends linearly on the current and various past values of a stochastic (imperfectly predictable) term.

**Q. Which of the following can't be a component for a time series plot?**

A) Seasonality

B) Trend

C) Cyclical

D) Noise

E) None of the above

Solution: (E)

A seasonal pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week). Seasonality is always of a fixed and known period. Hence, seasonal time series are sometimes called periodic time series

Seasonality is always of a fixed and known period. A cyclic pattern exists when data exhibit rises and falls that are not of fixed period.

Trend is defined as the 'long term' movement in a time series without calendar related and irregular effects, and is a reflection of the underlying level. It is the result of influences such as population growth, price inflation and general economic changes. The following graph depicts a series in which there is an obvious upward trend over time.
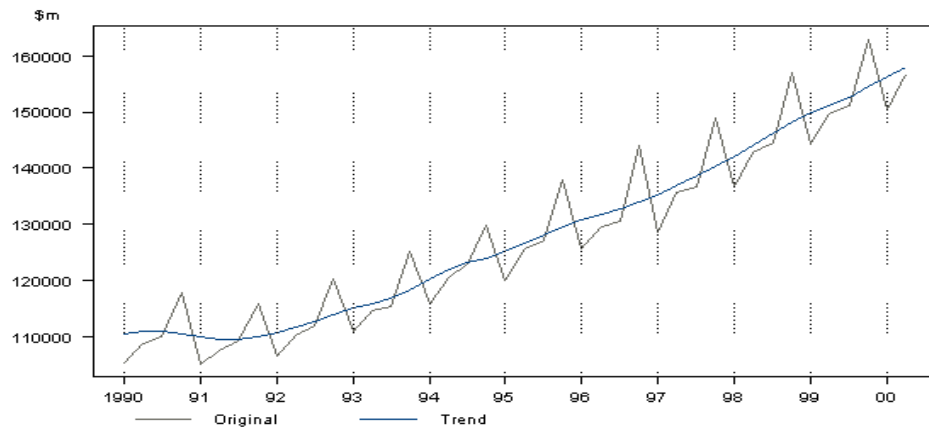
Quarterly Gross Domestic Product

Noise: In discrete time, white noise is a discrete signal whose samples are regarded as a sequence of serially uncorrelated random variables with zero mean and finite variance.
Thus all of the above mentioned are components of a time series.

Q. Which of the following is relatively easier to estimate in time series modeling?
A) Seasonality
B) Cyclical
C) No difference between Seasonality and Cyclical

Solution: (A)
As we seen in previous solution, as seasonality exhibits fixed structure; it is easier to estimate.
Q. The below time series plot contains both Cyclical and Seasonality component.
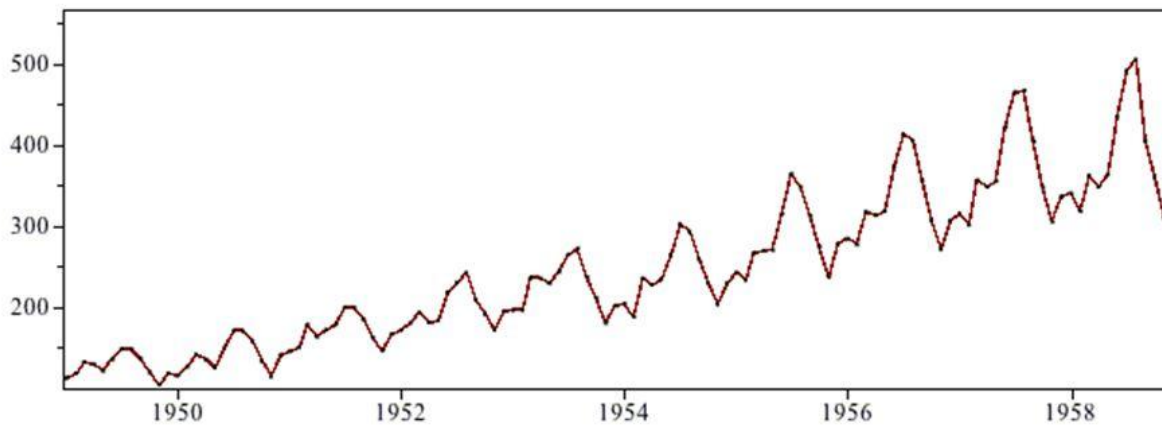
A) TRUE
B) FALSE

Solution: (B)
There is a repeated trend in the plot above at regular intervals of time and is thus only seasonal in nature.

Q. Adjacent observations in time series data (excluding white noise) are independent and identically distributed (IID).
A) TRUE
B) FALSE

Solution: (B)
Clusters of observations are frequently correlated with increasing strength as the time intervals between them become shorter. This needs to be true because in time series forecasting is done based on previous observations and not the currently observed data unlike classification or regression.

Q. Smoothing parameter close to one gives more weight or influence to recent observations over the forecast.
A) TRUE
B) FALSE

Solution: (A)
It may be sensible to attach larger weights to more recent observations than to observations from the distant past. This is exactly the concept behind simple exponential smoothing. Forecasts are calculated using weighted averages where the weights decrease exponentially as observations come from further in the past — the smallest weights are associated with the oldest observations:

$$Y_{t+1} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)(1-\alpha)Y_{t-2} + \cdots$$

where 0≤α≤1 is the smoothing parameter. The one-step-ahead forecast for time t+1 is a weighted average of all the observations in the series Y1,…,Yt. The rate at which the weights decrease is controlled by the parameter α.

Q. Sum of weights in exponential smoothing is _____.

A) <1

B) 1

C) >1

D) None of the above

Solution: (B)

Below table shows the weights attached to observations for four different values of α when forecasting using simple exponential smoothing. Note that the sum of the weights even for a small α will be approximately one for any reasonable sample size.

Observation	α=0.2	α=0.4	α=0.6	α=0.8
yT	0.2	0.4	0.6	0.8
yT−1	0.16	0.24	0.24	0.16
yT−2	0.128	0.144	0.096	0.032
yT−3	0.102	0.0864	0.0384	0.0064
yT−4	(0.2)(0.8)	(0.4)(0.6)	(0.6)(0.4)	(0.8)(0.2)
yT−5	(0.2)(0.8)	(0.4)(0.6)	(0.6)(0.4)	(0.8)(0.2)

Q. The last period's forecast was 70 and demand was 60. What is the simple exponential smoothing forecast with alpha of 0.4 for the next period?

A) 63.8

B) 65

C) 62

D) 66

Solution: (D)

Yt-1= 70, St-1= 60, Alpha = 0.4

Substituting the values we get

0.4*60 + 0.6*70= 24 + 42= 66

Q. What does auto covariance measure?

A) Linear dependence between multiple points on the different series observed at different times

B) Quadratic dependence between two points on the same series observed at different times

C) Linear dependence between two points on different series observed at same time

D) Linear dependence between two points on the same series observed at different times

Solution: (D)

Option D is the definition of auto covariance.

Q. Which of the following is not a necessary condition for weakly stationary time series?

A) Mean is constant and does not depend on time

B) Auto covariance function depends on s and t only through their difference |s-t| (where t and s are moments in time)

C) The time series under considerations is a finite variance process

D) Time series is Gaussian

Solution: (D)

A Gaussian time series implies stationarity is strict stationarity.

Q. Which of the following is not a technique used in smoothing time series?

A) Nearest Neighbor Regression

B) Locally weighted scatter plot smoothing

C) Tree based models like (CART)

D) Smoothing Splines

Solution: (C)

Time series smoothing and filtering can be expressed in terms of local regression models. Polynomials and regression splines also provide important techniques for smoothing. CART based models do not provide an equation to superimpose on time series and thus cannot be used for smoothing. All the other techniques are well documented smoothing techniques.

Q. If the demand is 100 during October 2016, 200 in November 2016, 300 in December 2016, 400 in January 2017. What is the 3-month simple moving average for February 2017?
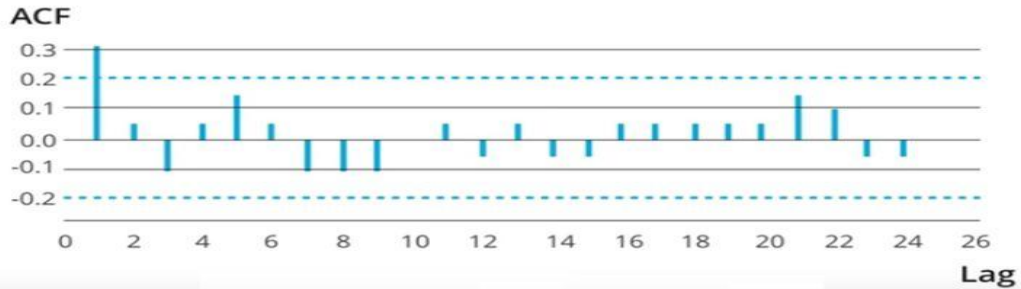
A) 300

B) 350

C) 400

D) Need more information

Solution: (A)

X`= (xt-3 + xt-2 + xt-1) /3

(200+300+400)/ 3 = 900/3 =300

Q. Looking at the below ACF plot, would you suggest to apply AR or MA in ARIMA modeling technique?

A) AR
B) MA
C) Can't Say

Solution: (A)

MA model is considered in the following situation, If the autocorrelation function (ACF) of the differenced series displays a sharp cutoff and/or the lag-1 autocorrelation is negative–i.e., if the series appears slightly "overdifferenced"–then consider adding an MA term to the model. The lag beyond which the ACF cuts off is the indicated number of MA terms.

But as there are no observable sharp cutoffs the AR model must be preferred.

Q. Suppose, you are a data scientist and you observed the views on the website increases during the month of Jan-Mar. whereas the views during Nov-Dec decreases. Does the above statement represent seasonality?
A) TRUE
B) FALSE
C) Can't Say

Solution: (A)

Yes this is a definite seasonal trend as there is a change in the views at particular times.
Remember, Seasonality is a presence of variations at specific periodic intervals.

Q. Which of the following graph can be used to detect seasonality in time series data?
Ans:
1. Multiple box
2. Autocorrelation

A) Only 1
B) Only 2
C) 1 and 2
D) None of these

Solution: (C)

Seasonality is a presence of variations at specific periodic intervals.

The variation of distribution can be observed in multiple box plots. And thus seasonality can be easily spotted. Autocorrelation plot should show spikes at lags equal to the period.

Q. Stationarity is a desirable property for a time series process.
A) TRUE
B) FALSE
Solution: (A)
When the following conditions are satisfied then a time series is stationary.

Mean is constant and does not depend on time
Auto covariance function depends on s and t only through their difference $|s-t|$ (where t and s are moments in time)
The time series under considerations is a finite variance process
These conditions are essential prerequisites for mathematically representing a time series to be used for analysis and forecasting. Thus stationarity is a desirable property.

Q. Suppose you are given a time series dataset which has only 4 columns (id, Time, X, Target).
Ans:

id	Time	X	Target
1	1	100	10
2	2	200	20
3	3	300	30
1	4	400	40
2	5	500	50
3	6	600	60
1	7	500	50
2	8	400	40
3	9	500	30
4	10	700	20

What would be the rolling mean of feature X if you are given the window size 2?
Note: X column represents rolling mean.

Solution:

Quater	Time	X'	Target
1	1	NaN	10
2	2	NaN	20
3	3	150	30
1	4	250	40
2	5	350	50
3	6	450	60
1	7	550	50
2	8	550	40
3	9	450	30
4	10	450	20

X`= xt-2 + xt-1 /2

Based on the above formula: (100 +200) /2 =150; (200+300)/2 = 250 and so on.

Q. Imagine, you are working on a time series dataset. Your manager has asked you to build a highly accurate model. You started to build two types of models which are given below.

Model 1: Decision Tree model
Model 2: Time series regression model

At the end of evaluation of these two models, you found that model 2 is better than model 1. What could be the possible reason for your inference?
A) Model 1 couldn't map the linear relationship as good as Model 2
B) Model 1 will always be better than Model 2
C) You can't compare decision tree with time series regression
D) None of these

Solution: (A)
A time series model is similar to a regression model. So it is good at finding simple linear relationships. While a tree based model though efficient will not be as good at finding and exploiting linear relationships.

Q. What type of analysis could be most effective for predicting temperature on the following type of data?

Date	Temperature	precipitation	temperature/precipitation
12/12/12	7	0.2	35
13/12/12	9	0.123	73.1707317073
14/12/12	9.2	0.34	27.0588235294
15/12/12	10	0.453	22.0750551876
16/12/12	12	0.33	36.3636363636
17/12/12	11	0.8	13.75

A) Time Series Analysis

B) Classification

C) Clustering

D) None of the above

Solution: (A)

The data is obtained on consecutive days and thus the most effective type of analysis will be time series analysis.

Q. What is the first difference of temperature / precipitation variable?

Date	Temperature	precipitation	temperature/precipitation
12/12/12	7	0.2	35
13/12/12	9	0.123	73.1707317073
14/12/12	9.2	0.34	27.0588235294
15/12/12	10	0.453	22.0750551876
16/12/12	12	0.33	36.3636363636
17/12/12	11	0.8	13.75

A) 15, 12.2,-43.2,-23.2, 14.3,-7

B) 38.17,-46.11,-4.98, 14.29,-22.61

C) 35, 38.17,-46.11,-4.98, 14.29,-22.61

D) 36.21,-43.23,-5.43, 17.44,-22.61

Solution: (B)

73.17-35 = 38.17

27.05-73.17 = – 46.11 and so on..

Q. Consider the following set of data:

{23.32 32.33 32.88 28.98 33.16 26.33 29.88 32.69 18.98 21.23 26.66 29.89}

What is the lag-one sample autocorrelation of the time series?

A) 0.26
B) 0.52
C) 0.13
D) 0.07

Solution: (C)

$\hat{\rho}1 = (23.32-\bar{x})(32.33-\bar{x})+(32.33-\bar{x})(32.88-\bar{x})+\cdots = 0.130394786$

Where $\bar{x}$ is the mean of the series which is 28.0275

Q. Any stationary time series can be approximately the random superposition of sines and cosines oscillating at various frequencies.
A) TRUE
B) FALSE

Solution: (A)
A weakly stationary time series, xt, is a finite variance process such that
The mean value function, μt, is constant and does not depend on time t, and (ii) the auto covariance function, $\gamma(s,t)$, defined in depends on s and t only through their difference $|s-t|$.
random superposition of sines and cosines oscillating at various frequencies is white noise. white noise is weakly stationary or stationary. If the white noise variates are also normally distributed or Gaussian, the series is also strictly stationary.

Q. Auto covariance function for weakly stationary time series does not depend on _____ ?
A) Separation of xs and xt
B) $h = | s - t |$
C) Location of point at a particular time

Solution: (C)
By definition of weak stationary time series described in previous question.
Q. Two time series are jointly stationary if _____ ?
A) They are each stationary
B) Cross variance function is a function only of lag h
A) Only A
B) Both A and B

Solution: (D)
Joint stationarity is defined based on the above two mentioned conditions.

Q. In autoregressive models _____ ?

A) Current value of dependent variable is influenced by current values of independent variables

B) Current value of dependent variable is influenced by current and past values of independent variables

C) Current value of dependent variable is influenced by past values of both dependent and independent variables

D) None of the above

Solution: (C)

Autoregressive models are based on the idea that the current value of the series, xt, can be explained as a function of p past values, xt−1,xt−2,…,xt−p, where p determines the number of steps into the past needed to forecast the current value. Ex. xt = xt−1 −.90xt−2 + wt,

Where xt-1 and xt-2 are past values of dependent variable and wt the white noise can represent values of independent values.

The example can be extended to include multiple series analogous to multivariate linear regression.

Q. For MA (Moving Average) models the pair σ = 1 and θ = 5 yields the same auto covariance function as the pair σ = 25 and θ = 1/5.

$$x_t = w_t + \frac{1}{5}w_{t-1}, \quad w_t \sim \text{iid } N(0, 25)$$

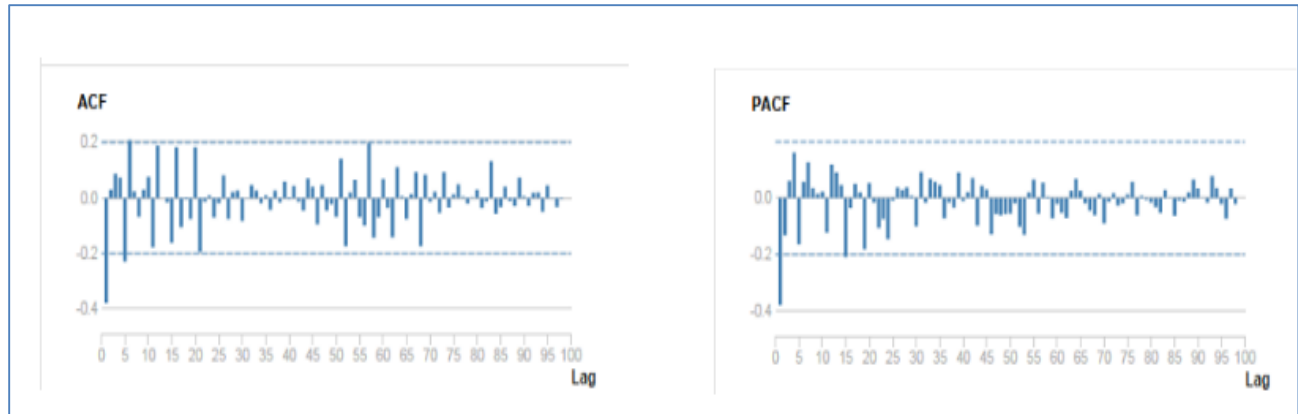$$y_t = v_t + 5v_{t-1}, \quad v_t \sim \text{iid } N(0, 1)$$

A) TRUE
B) FALSE

Solution: (A)

True, because auto covariance is invertible for MA models

Note: that for an MA(1) model, ρ(h) is the same for θ and 1 /θ
The pair σ2 w = 1 and θ = 5 yield the same auto covariance function as the pair σ2 w = 25 and θ = 1/5.

Q. How many AR and MA terms should be included for the time series by looking at the above ACF and PACF plots?

A) AR (1) MA(0)
B) AR(0)MA(1)
C) AR(2)MA(1)
D) AR(1)MA(2)
E) Can't Say

Solution: (B)

Strong negative correlation at lag 1 suggest MA and there is only 1 significant lag

Q. Which of the following is true for white noise?

A) Mean =0
B) Zero auto covariances
C) Zero auto covariances except at lag zero
D) Quadratic Variance

Solution: (C)

A white noise process must have a constant mean, a constant variance and no auto covariance structure (except at lag zero, which is the variance).

Q. For the following MA (3) process Yt = μ + Et + θ1Et-1 + θ2Et-2 + θ3Et-3 , where σt is a zero mean white noise process with variance σ2

A) ACF = 0 at lag 3
B) ACF =0 at lag 5
C) ACF =1 at lag 1
D) ACF =0 at lag 2
E) ACF = 0 at lag 3 and at lag 5

Solution: (B)

Recall that an MA(q) process only has memory of length q. This means that all of the autocorrelation coefficients will have a value of zero beyond lag q. This can be seen by examining the MA equation, and seeing that only the past q disturbance terms enter into the equation, so that if we iterate this equation forward through time by more than q periods, the current value of the disturbance term will no longer affect y. Finally, since the autocorrelation function at lag zero is the correlation of y at time t with y at time t (i.e. the correlation of y_t with itself), it must be one by definition.

Q. Consider the following AR(1) model with the disturbances having zero mean and unit variance.
$y_t = 0.4 + 0.2y_{t-1} + u_t$ **The (unconditional) variance of y will be given by?**

A) 1.5
B) 1.04
C) 0.5
D) 2

Solution: (B)
Variance of the disturbances divided by (1 minus the square of the autoregressive coefficient
Which in this case is : $1/(1-(0.2^2)) = 1/0.96 = 1.041$

Q. The pacf (partial autocorrelation function) is necessary for distinguishing between _____ ?
A) An AR and MA model is solution: False
B) An AR and an ARMA is solution: True
C) An MA and an ARMA is solution: False
D) Different models from within the ARMA family

Solution: (B)

Table 3.1. Behavior of the ACF and PACF for ARMA Models

	$AR(p)$	$MA(q)$	$ARMA(p, q)$
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Q. Second differencing in time series can help to eliminate which trend?
A) Quadratic Trend
B) Linear Trend
C) Both A & B
D) None of the above

Solution: (A)

The first difference eliminates a linear trend. A second difference, can eliminate a quadratic trend, and so on.

Q. Which of the following cross validation techniques is better suited for time series data?

A) k-Fold Cross Validation

B) Leave-one-out Cross Validation

C) Stratified Shuffle Split Cross Validation

D) Forward Chaining Cross Validation

Solution: (D)

Time series is ordered data. So the validation data must be ordered to. Forward chaining ensures this. It works as follows:

fold 1 : training [1], test [2]

fold 2 : training [1 2], test [3]

fold 3 : training [1 2 3], test [4]

fold 4 : training [1 2 3 4], test [5]

fold 5 : training [1 2 3 4 5], test [6]

Q. BIC penalizes complex models more strongly than the AIC.

A) TRUE

B) FALSE

Solution: (A)

AIC = -2*ln(likelihood) + 2*k,
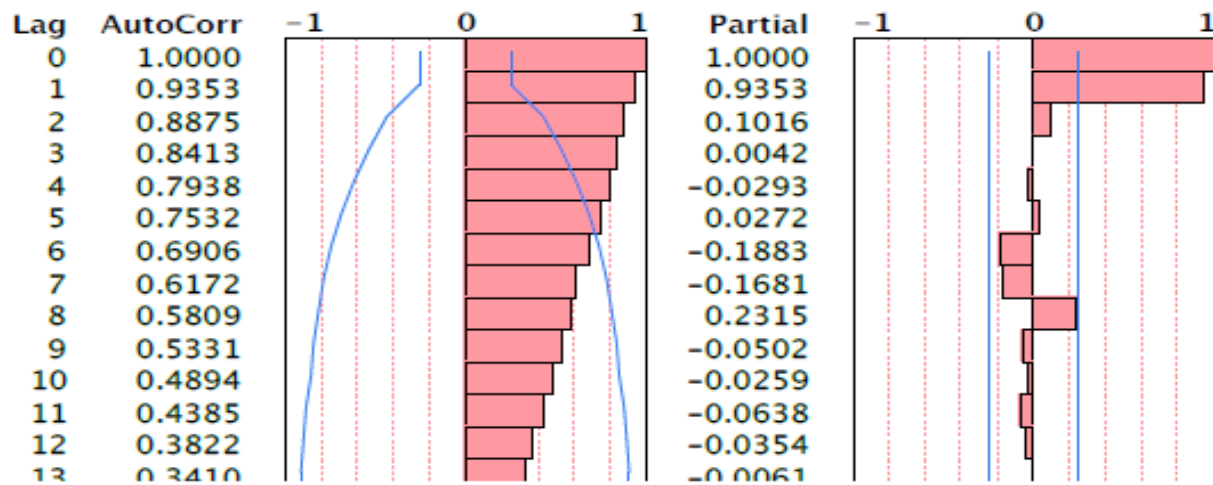
BIC = -2*ln(likelihood) + ln(N)*k,

where:

k = model degrees of freedom

N = number of observations

At relatively low N (7 and less) BIC is more tolerant of free parameters than AIC, but less tolerant at higher N (as the natural log of N overcomes 2).

Q. The figure below shows the estimated autocorrelation and partial autocorrelations of a time series of n = 60 observations. Based on these plots, we should.

Lag	AutoCorr	-1	0	1	Partial	-1	0	1
0	1.0000				1.0000			
1	0.9353				0.9353			
2	0.8875				0.1016			
3	0.8413				0.0042			
4	0.7938				-0.0293			
5	0.7532				0.0272			
6	0.6906				-0.1883			
7	0.6172				-0.1681			
8	0.5809				0.2315			
9	0.5331				-0.0502			
10	0.4894				-0.0259			
11	0.4385				-0.0638			
12	0.3822				-0.0354			
13	0.3410				-0.0061			

A) Transform the data by taking logs

B) Difference the series to obtain stationary data

C) Fit an MA(1) model to the time series

Solution: (B)

The autocorr shows a definite trend and partial autocorrelation shows a choppy trend, in such a scenario taking a log would be of no use. Differencing the series to obtain a stationary series is the only option.

Question Context for next two questions:

The remaining questions consider a time series model for annual global temperature. The data for the time series in this analysis begin in 1856 and run through 1997 (n = 142). The measurements give the deviation from typical temperature in degrees Celsius. (Zero would be considered consistent with the long-run average.)
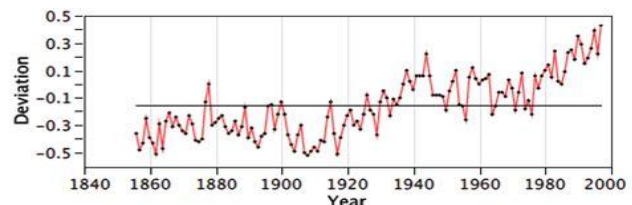
Model Summary

DF	140.0000
Sum of Squared Errors	1.7726
Variance Estimate	0.0127
Standard Deviation	0.1125
Akaike's 'A' Information Criterion	-214.4648
Schwarz's Bayesian Criterion	-211.5160
RSquare	0.7328

	Actual Deviation	Year	Predicted Deviation
133	0.25	1988	0.13245007
134	0.18	1989	0.17909389
135	0.35	1990	0.17945343
136	0.29	1991	0.24712632
137	0.15	1992	0.2641386
138	0.19	1993	0.2188484
139	0.26	1994	0.20740135
140	0.39	1995	0.2282725
141	0.22	1996	0.29244598
142	0.43	1997	0.26369941

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Level Smoothing Weight	0.39680005	0.0900926	4.40	<.0001*

Q. Use the estimated exponential smoothening given above and predict temperature for the next 3 years (1998-2000)

These results summarize the fit of a simple exponential smooth to the time series.

A) 0.2, 0.32, 0.6
B) 0.33, 0.33, 0.33
C) 0.27, 0.27, 0.27
D) 0.4, 0.3, 0.37

Solution: (B)

The predicted value from the exponential smooth is the same for all 3 years, so all we need is the value for next year. The expression for the smooth is

Smooth t = α yt + (1 – α) smooth t-1Hence, for the next point, the next value of the smooth (the prediction for the next observation) is
Smooth n = α yn + (1 – α) smooth n-1
= 0.3968*0.43 + (1 – 0.3968)* 0.3968
= 0.3297

Q. Find 95% prediction intervals for the predictions of temperature in 1999.
These results summarize the fit of a simple exponential smooth to the time series.
A) 0.3297 2 * 0.1125
B) 0.3297 2 * 0.121
C) 0.3297 2 * 0.129
D) 0.3297 2 * 0.22

Solution: (B)
The sd of the prediction errors is
1 period out 0.1125
2 periods out 0.1125 sqrt(1+α2) = 0.1125 * sqrt(1+ 0.39682) ≈ 0.121

Q. Which of the following statement is correct?
1. If autoregressive parameter (p) in an ARIMA model is 1, it means that there is no auto-correlation in the series.
2. If moving average component (q) in an ARIMA model is 1, it means that there is auto-correlation in the series with lag 1.
3. If integrated component (d) in an ARIMA model is 0, it means that the series is not stationary.

A) Only 1   B) Both 1 and 2   C) Only 2   D) All of the statements

Solution: (C)
Autoregressive component: AR stands for autoregressive. Autoregressive parameter is denoted by p. When p =0, it means that there is no auto-correlation in the series. When p=1, it means that the series auto-correlation is till one lag.

Integrated: In ARIMA time series analysis, integrated is denoted by d. Integration is the inverse of differencing. When d=0, it means the series is stationary and we do not need to take the difference of it. When d=1, it means that the series is not stationary and to make it stationary, we need to take the first difference. When d=2, it means that the series has been differenced twice. Usually, more than two time difference is not reliable.

Moving average component: MA stands for moving the average, which is denoted by q. In ARIMA, moving average q=1 means that it is an error term and there is auto-correlation with one lag.

Q. In a time-series forecasting problem, if the seasonal indices for quarters 1, 2, and 3 are 0.80, 0.90, and 0.95 respectively. What can you say about the seasonal index of quarter 4?
A) It will be less than 1
B) It will be greater than 1
C) It will be equal to 1
D) Seasonality does not exist
E) Data is insufficient
Solution: (B)
Q. What are the observations you can make from the below plot?
Ans:
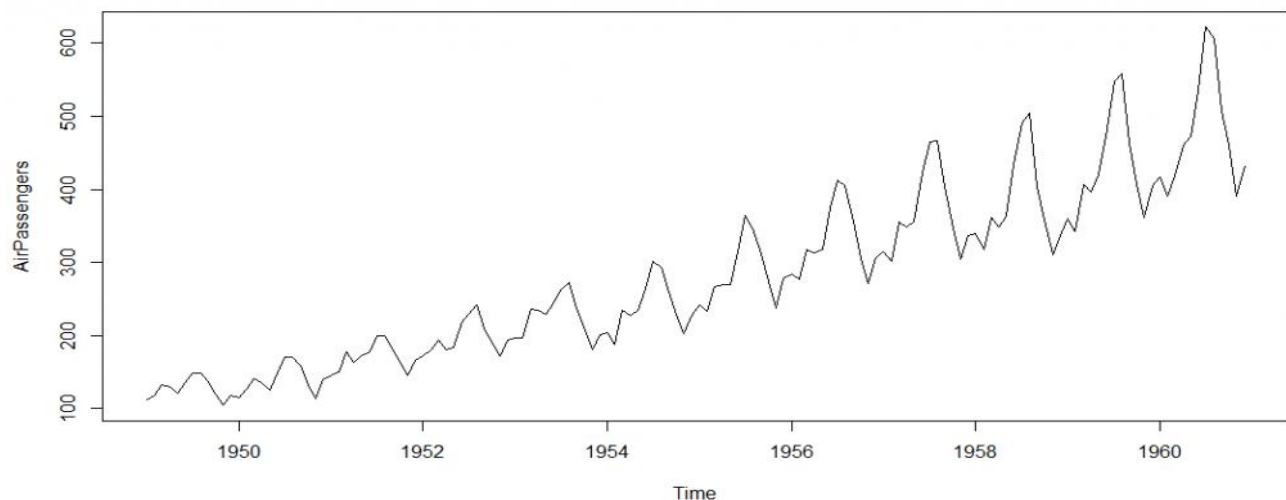
Here are my observations:
1. There is a trend component which grows the passenger year by year.
2. There looks to be a seasonal component which has a cycle less than 12 months.
3. The variance in the data keeps on increasing with time.
We know that we need to address two issues before we test stationary 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.

Q: Can you provide quick overview of ARIMA Time Series Modelling
Ans: Step by step approach on '**How to do a Time Series Analysis**:

1. Visualize the time series

2. Stationarize the series

3. Plot ACF/PACF charts and find optimal parameters

4. Build the ARIMA model

5. Make Predictions

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 pertains 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 of this article series. This doesn't ends here! What if the series is found to be non-stationary?
There are three commonly used technique to make a time series stationary:

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

x(t) = (mean + trend * t) + 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) = 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 and PACF plots. 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 step 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.

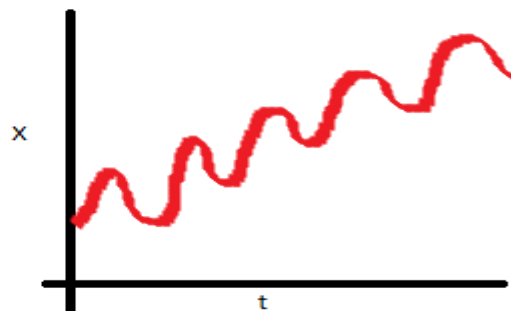### Q. Can you elaborate on Stationarity of time series?

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.
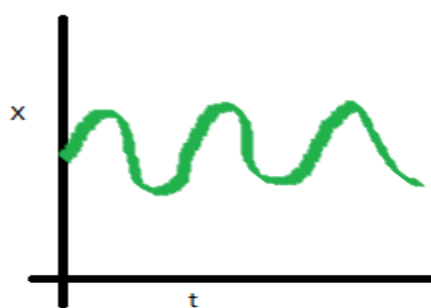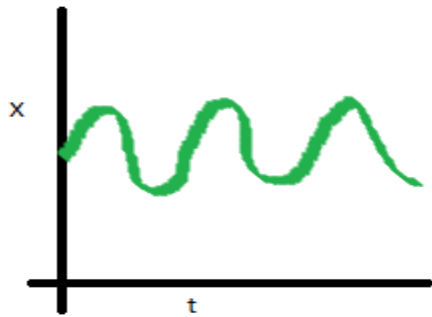

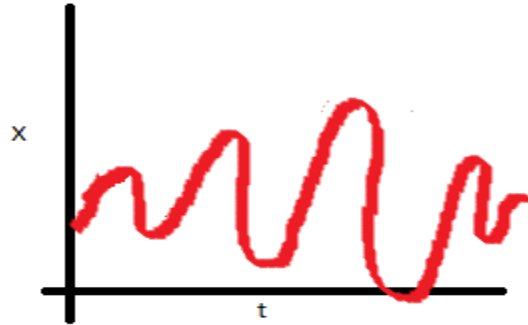Stationary series   Non-Stationary series

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)
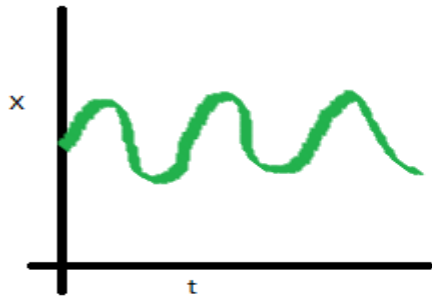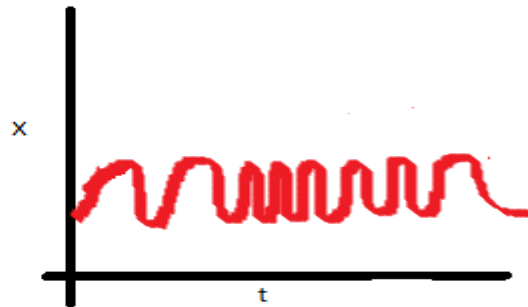
Stationary series

Non-Stationary series

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'.

Stationary series

Non-Stationary series

Q. Why do I care about 'stationarity' of a time series?
Ans:

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.

What is Random Walk?
Ans: This is the most basic concept of the time series. 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.

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 Xs, we will finally end up to the following equation:
$X(t) = X(0) + Sum(Er(1), Er(2), Er(3)\ldots Er(t))$

Q. Is random work stationary series?
Ans: Is the Mean constant?

$E[X(t)] = E[X(0)] + Sum(E[Er(1)], E[Er(2)], E[Er(3)]\ldots E[Er(t)])$
We know that Expectation of any Error will be zero as it is random.
Hence we get $E[X(t)] = E[X(0)] = Constant$.

Is the Variance constant?
$Var[X(t)] = Var[X(0)] + Sum(Var[Er(1)], Var[Er(2)], Var[Er(3)]\ldots Var[Er(t)])$
$Var[X(t)] = t * Var(Error) = Time\ dependent$.

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.