# Time Series Questions and Answers

## 1. What is 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. ARIMA technique exploits the auto-correlation (Correlation of observation with its lags) for forecasting.

So talking mathematically,

Vt = p(Vt-n) + e

It means Value (V) at time "t" is a function of value at time "n" instance ago with an error (e). Value at time "t" can depend on one or various lags of various order.

## 2. What are the 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

## 3. What are the Terminologies related to Time Series?

- **Stationary Series**
- **White Noise**
- **Autocorrelation**
- **Random Walk**

## 4. What is 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.

**5. What is ARIMA?**

**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 on SAS. Well, coming back to ARIMA, as its full form indicates that it involves two components:

1. **Auto-regressive component**
2. **Moving average component**

**6. What are the 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

**7. How to check the series?**

As a matter of practice, we first plot the time series and have a cursory look upon it. It can be done directly in SAS using following code:

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

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

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.

**8. What are the 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:

eck Stationary

**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.*

**9. What is 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.

**10. What are the conditions are satisfied then a time series is stationary?**

1. Mean is constant and does not depend on time
2. Autocovariance function depends on s and t only through their difference |s-t| (where t and s are moments in time)
3. 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.

**11. Smoothing parameter close to one gives more weight or influence to recent observations over the forecast?**

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:

$\hat{y}_{T+1|T}=\alpha y_T+\alpha(1-\alpha)y_T-1+\alpha(1-\alpha)2y_{T-2}+\cdots,(7.1)$

where $0\leq\alpha\leq 10\leq\alpha\leq 1$ is the smoothing parameter. The one-step-ahead forecast for time T+1T+1 is a weighted average of all the observations in the series $y_1,\ldots,y_T$. The rate at which the weights decrease is controlled by the parameter αα.

**12. Explain Autoregressive models?**

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.

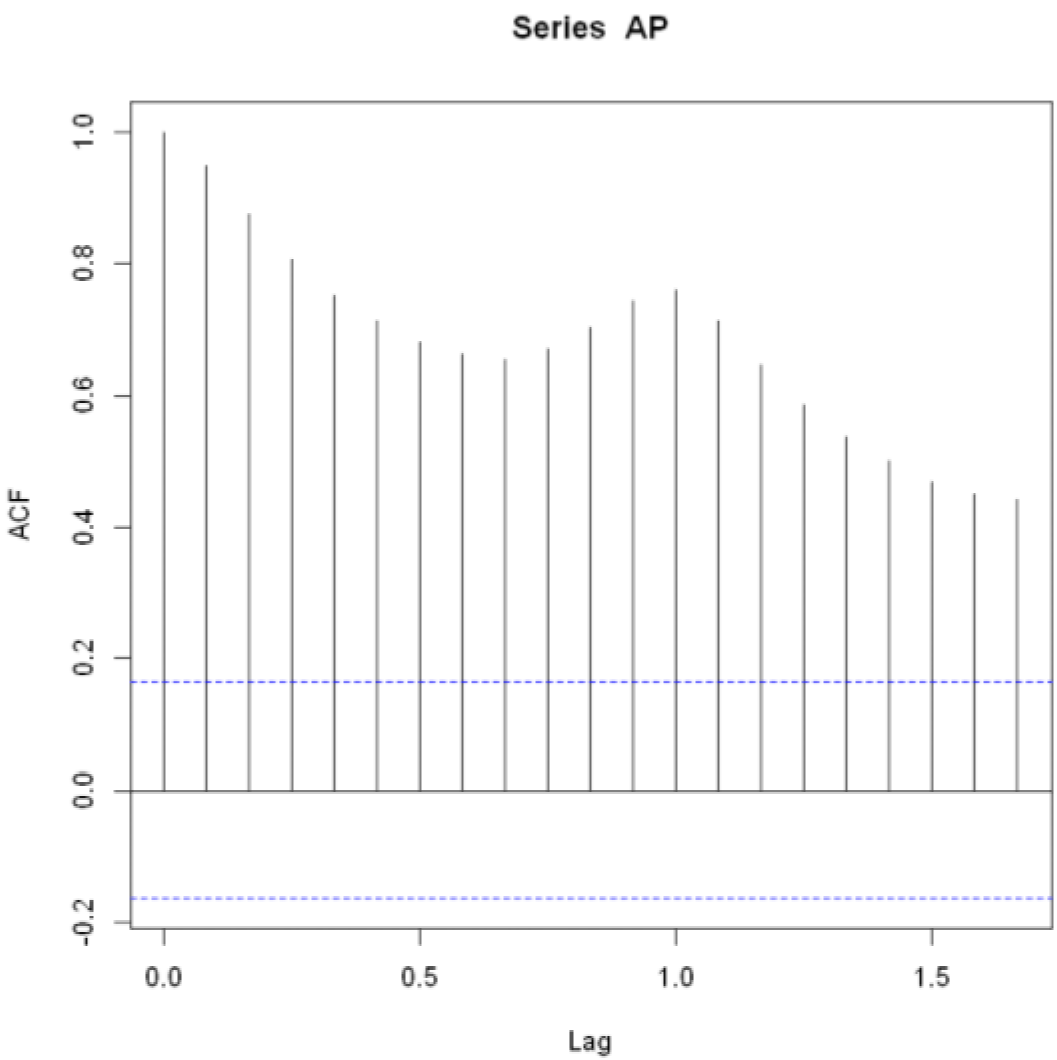
The example can be extended to include multiple series analogous to multivariate linear regression.

**13. Explain Partial Autocorrelation Function (PACF)?**

PACF is same as ACF just that the intermediate lags between t and t-p is removed i.e. correlation between Y(t) and Y(t-p) with p-1 lags excluded.

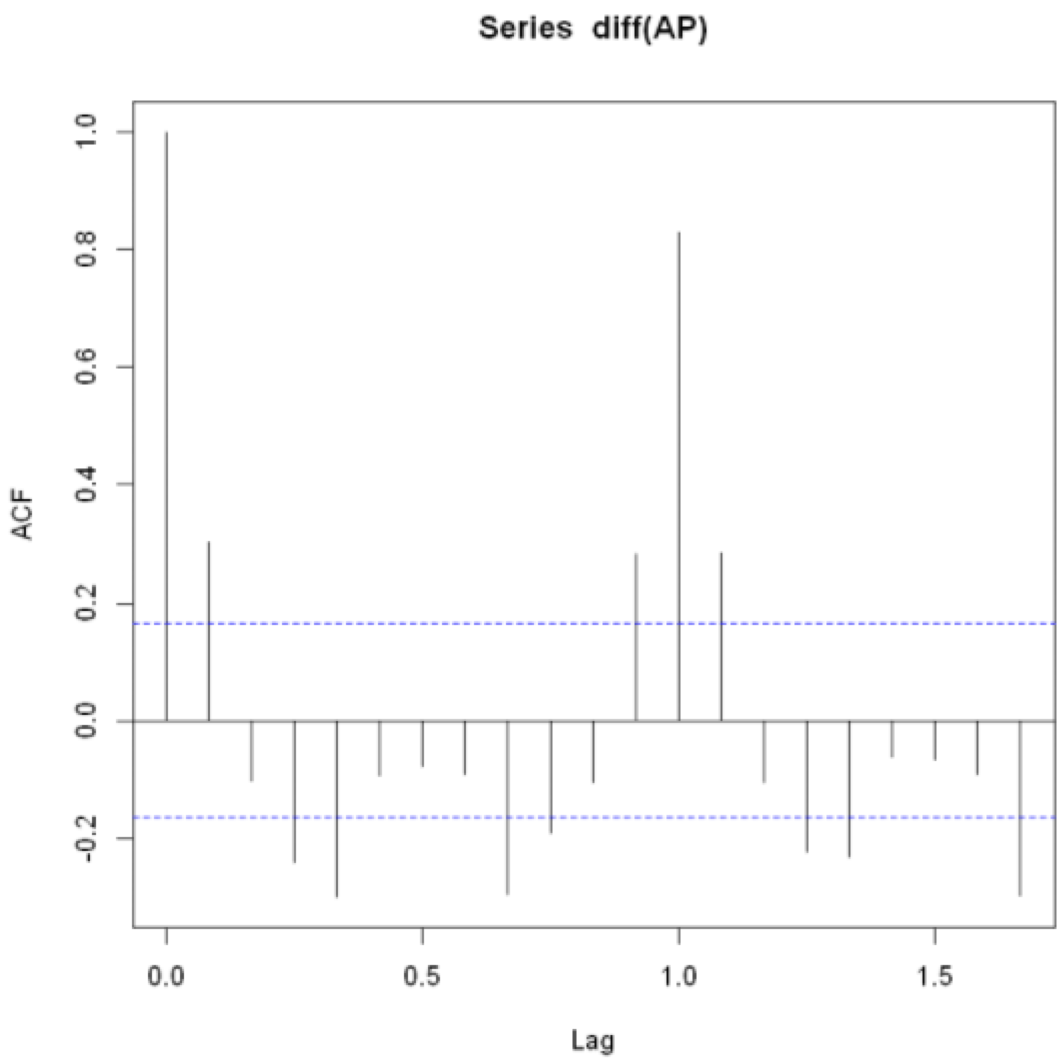
We’ll plot ACF as well as PACF in R

```
acf(AP,lag.max = 20)
```




Here, we can see that the ACF is not well within the confidence interval band. Hence the series is not stationary. We'll differentiate the series once and observe the results.

```
acf(diff(AP),lag.max = 20)
```




Now, our series is stationary as most of the spikes are within the boundary. Thus, the value for I will be 1.

Also, the value of ACF at lag-1 is out of boundary, thus value of p should be 0.

Now,we can fit ARIMA on the time series data.

arima(AP,order = c(0,1,1))

We can try different combinations for q and check out for AIC, BIC values for each model.

The model with the lowest AIC is the correct model.

To automatically choose best values for p,q and d, we can use auto.arima.

auto.arima(AP)

After fitting the best model, we check for randomness. tsdiag() produces a diagnostic plot of a fitted time.

tsdiag(auto.arima(AP))

Next, we'll predict the values for next 8 time intervals, i.e. next 8 months.

predict(auto.arima(AP),n.ahead=8)

The output of predict will contain two lists. Predicted values ($pred) and standard errors of prediction ($se)

**14. Explain Stationarity Check?**

In order to check if a series is stationary or not we use Ljung-Box test or Augmented Dickey-Fuller Test.

```
Box.test(AP, lag=20,type="Ljung-Box")
```

```
        Box-Ljung test

data:  AP
X-squared = 1434.1, df = 20, p-value < 2.2e-16
```

*Ljung–Box Test for stationarity check*

Here, we observe that the p-value < 0.05. Hence, we can say that the series is stationary.

Now, If data is not stationary then we have to make the data as stationary by differentiating.

For this kind of data, we use Autoregressive Integrated Moving Average(ARIMA).

**15. Explain Cyclic Variation?**

- The variation of observations in a time series occurring generally in business and economics where the rises and falls in the data are not of fixed period is known as Cyclic Variation.
- The duration of these cycles is more than a year.

For Eg: Sensex Price