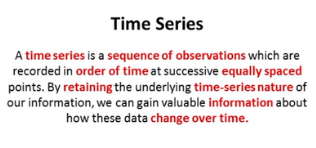
**MODULE – Time Series Forecasting**

Forecasting refers to the exercise of estimating what may occur in the future, taking past and present events into consideration.

**What is time series?**

A time series is a sequence of observations which are recorded in order of time at successive equally spaced points. By retaining the underlying time-series nature of our information, we can gain valuable information about how these data change over time.



**Types of data with time element**

* ***Cross sectional data***: Collection of observations of various features at a single instance of time. It is like taking a snapshot of the information.
* ***Time series data***: Collection of observations of a single feature at different instances of time. The time intervals should be equally spaced.
* ***Panel data***: Combination of cross sectional and time series data. It records information for various features at different instances of time.

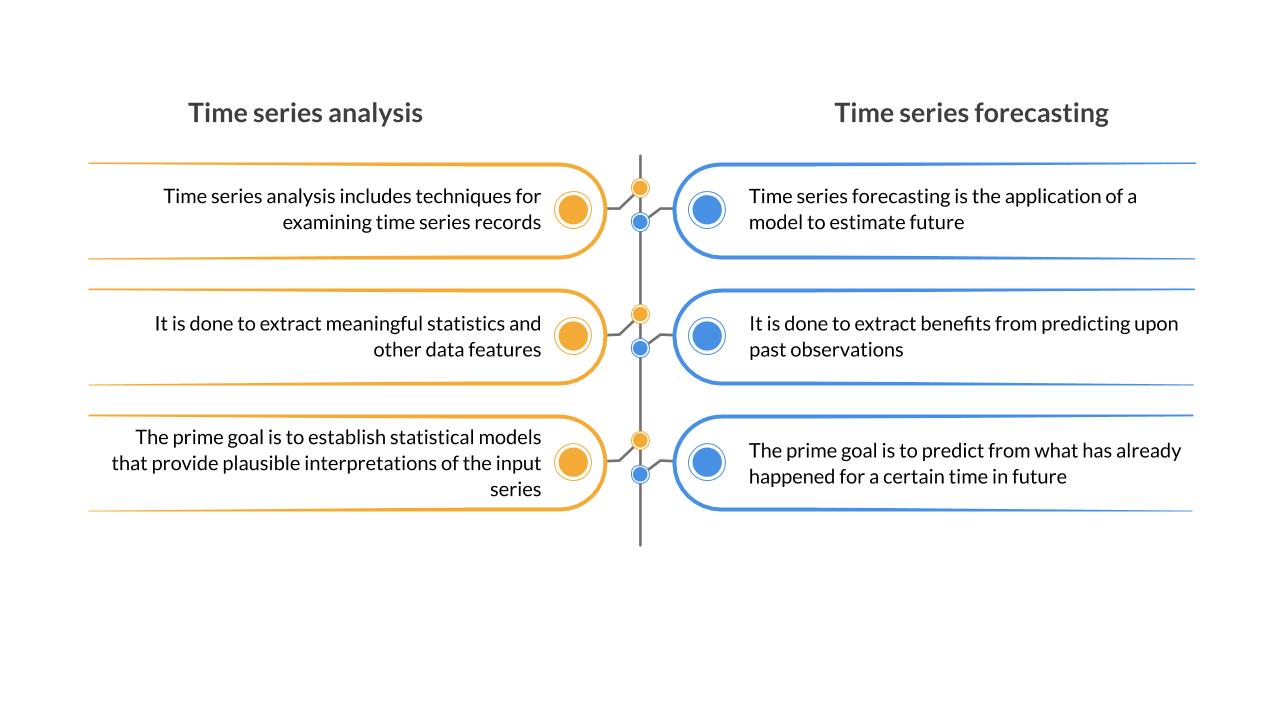
Properties of time series data

* ***Periodically new***: You have a new single entry at a fixed time interval.
* ***Ordered data***: Your entries contain a time sequence (last entry the most recent).
* ***Time vs Data***: Plotting a curve, time would be the primary (x) axis.

Applications of time series

* ***Understand***: Understand the past behaviour of an event.
* ***Predict***: Predict patterns using past trends and depict a clear picture of rise or fall.
* ***Analyse***: Evaluate current events for market analysis and policymaking.
* ***Adjust***: Make important seasonal adjustments to earn more in those seasons.

Time series analysis vs Time series forecasting



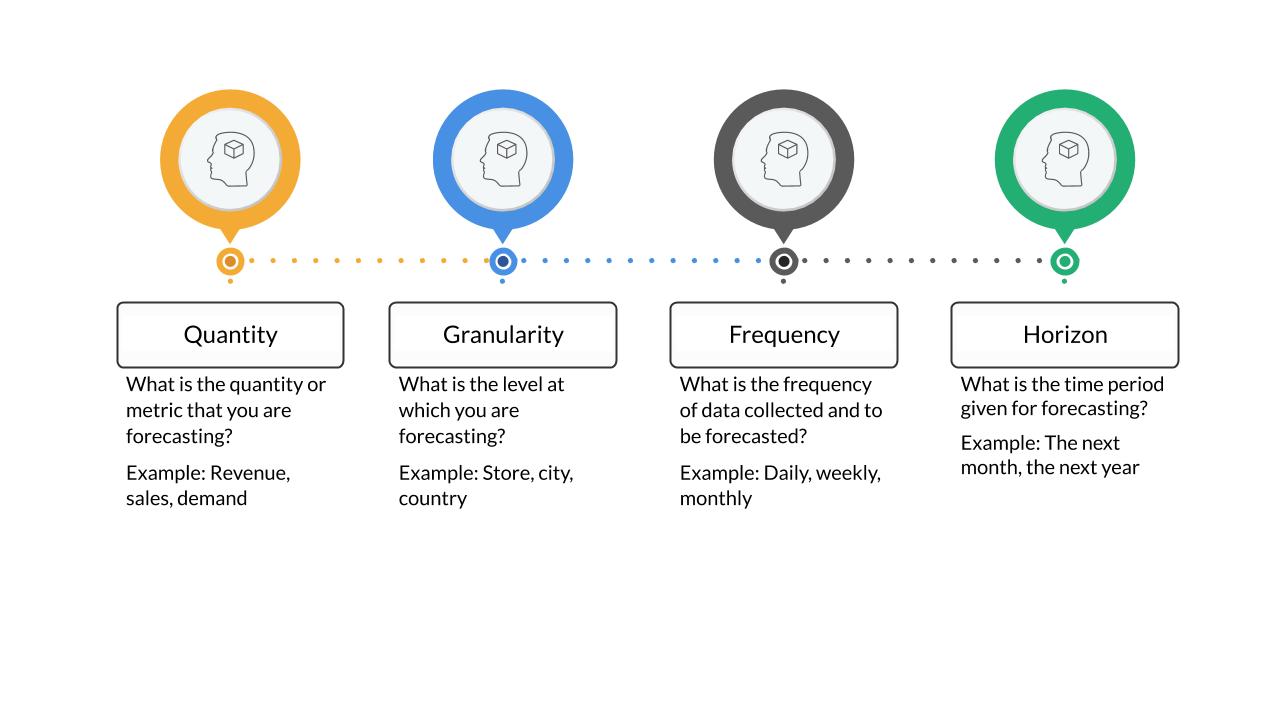
Regression vs Time series

|  |  |
| --- | --- |
| **Regression** | **Time-Series Forecasting** |
| A normal ML data set is a set of observations. Time is of no significance in normal ML data sets. | In a time-series data set, the data points have an ordered dependence on each other because of the time dimension. |
| The time array, if seen in normal data sets, is for indexing purpose and is not related to instance-based information. | The time arrays in time series data sets are elements of primary importance for establishing an ordered relation among the data points. |

**Steps for time series forecasting**

1. Define the problem.
2. Collect the data.
3. Analyse the data.
4. Build the forecast model.
5. Evaluate the forecast model.

**Step 1) Defining the problem – components**



**Step 2) Data Collection**

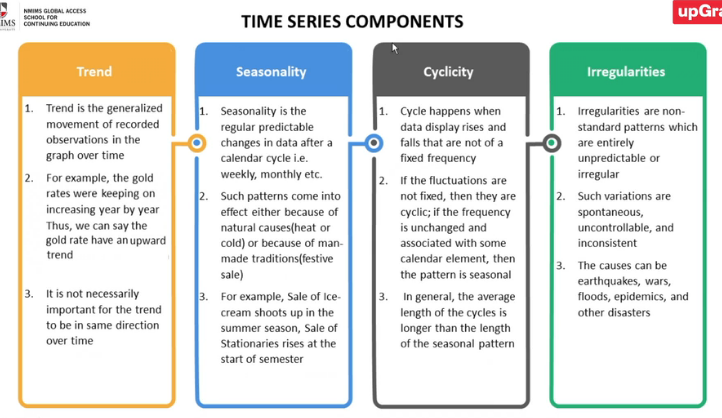
Major aspects of the data collection process are:

* **Data characteristics**
  + ***Relevant***: The time-series data should be relevant for the set objective that we want to achieve
  + ***Accurate***: The data should be accurate in terms of capturing the timestamps and capturing the observation correctly
  + ***Long enough:*** The data should be long enough to forecast. This is because it is important to identify all the patterns in the past and forecast which patterns repeat in the future
* **Data sources**
  + ***Private enterprise data***: e.g., Financial information about the quarterly results of any private organisation
  + ***Public data***: e.g., Government publishes the economic indicators
  + ***System/Sensor data***: e.g., Logs generated by the servers during their 24/7 working hours

**Step 3) Analyse the data**

The components of the time series include:

* Trend
* Seasonality
* Cyclicity
* Irregularities

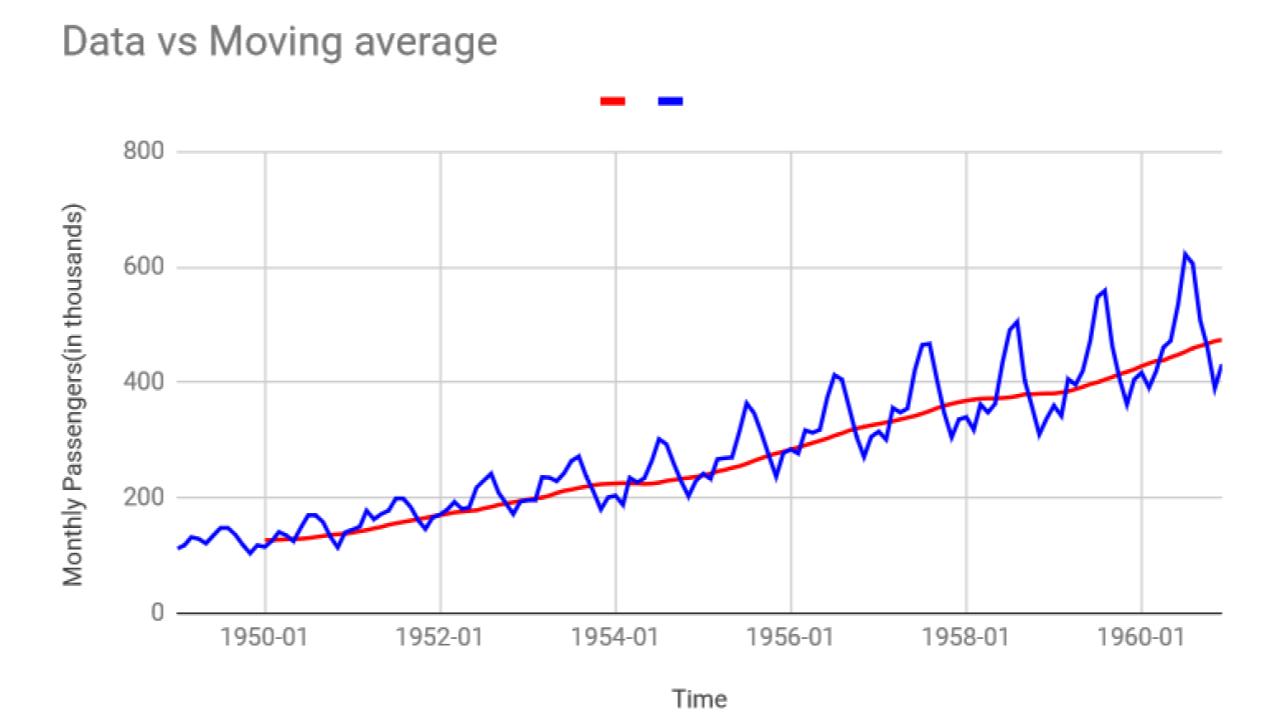


**Step 4) Build the forecast model**

**Forecast Model – Simple moving average**

It is based on the principle that the near past is the great estimator of the future.

In the simple moving average technique, the prediction for a timestamp is the average of its previous ‘n’ timestamp observations. ‘n’ is called the moving window of the method.

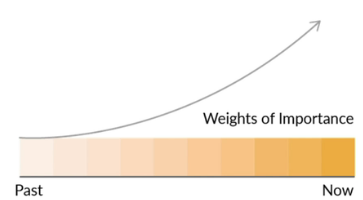


Here are **some important inferences from the simple moving average** method:

* The moving average **captures the trend** of the data automatically.
* It does **not** capture **seasonal effects** within the year.

**Forecast Model – Simple exponential smoothing**

* Forecasts generated with exponential smoothing are **weighted averages** of previous measurements.
* As the observations get older, the weights decrease exponentially, i.e., the more recent the finding, the higher its weight.



This *weight* is typically denoted by α.

So, in simple exponential smoothing, the prediction is just the weighted number of results from before. It includes a single parameter α, called the smoothing factor.

What is α, the smoothing factor?

This parameter controls the rate of exponential decay of the earlier measures.

α is always set between 0 to 1.

How α defines the weight to data?

Large values mean that the model focuses primarily on the recent data. Smaller values mean that the older datapoints are given importance.

Actually,

Yt+1 = Yt + α (At – Yt), where

Yt+1 is the forecasted value.

Yt is the last forecasted value.

α is the smoothing factor.

At is the last actual value.

This means that the forecast for the current period is obtained by adding the forecast in the last period to a fraction of the error in the last period, where error is the difference between actual value and the forecasted value in the last period.

The above formula can be re-written as:

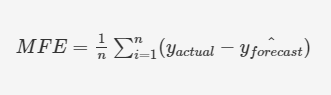
Yt+1 = αAt + (1- α) Yt

which means forecast value is α times last actual value + (1- α) times last forecast value.

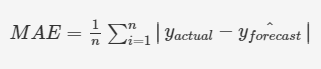
**Model Accuracy**

In forecasting models, we evaluate errors in models via different techniques as a measure of accuracy of the model.

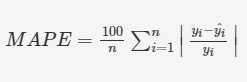
**Mean Forecast Error (MFE)**: The net model error is the average of differences of actual and forecast values.



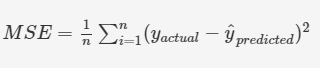
**Mean Absolute Error (MAE):** The net model error is the average of absolute differences of actual and forecast values. This is useful as MFE might cancel out several overestimated and underestimated forecasts.



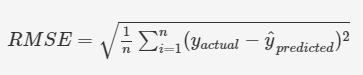
**Mean Absolute Percentage Error (MAPE)**: The net model error is the average of percentages of the absolute difference to the actual values. The problem with MAE is that we don’t’ have anything to compare against error values.



**Mean Squared Error (MSE)**: The net model error is the average of squares of difference of the values. The idea behind MSE is the same as the underlying idea for MAE – you want to capture the absolute deviations so that the negative and positive deviations do not cancel each other out.



**Root Mean Squared Error (RMSE)**: The net model error is the square root of MSE value. Since the error term you obtained from MSE is not in the same dimension as the target variable, you deploy the RMSE where in you take the square root of the MSE value obtained.



Other than the earlier discussed methods (SMA and Simple Exponential Smoothing), there are other methods which are slightly more sophisticated, called as Auto-regressive models.

**Auto -Regressive Models**

Auto-regression implies that it is a regression of the parameter against itself. It means that something is reverting to itself, meaning today’s value is reverting to previous values. In normal regression, you’ll have a y and x and you regress y on the values of x. However, in auto-regression, you will have Yt and you will regress it on values of Yt-k.

It estimates the parameter of concern using linear combinations of the results of the variable.

To predict the Yt, future forecast of a variable, we need the past observations of Yt i.e., Yt-1, Yt-2 etc.

Auto-regressive modelling constraints

In regression, we assume that errors are independently distributed, i.e., it doesn’t depend on the value of the independent variables (*homoscedasticity*).

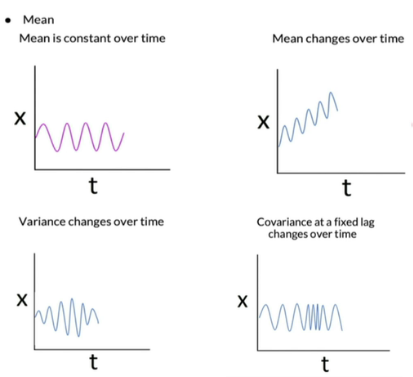
In auto-regression, Yt is dependent on Yt-1, hence the errors are not independent. This leads to “Order Constraint” – the time element creates an order amongst the data. Trying to fit statistical model requires to take care of consistency assumptions.

For auto-regressive modelling, the data fails to abide basic regression assumptions like:

* Have a constant variance.
* Have a constant mean.
* Time independent covariance.

**Trend and Seasonality Accounting**

Out of the four components of time series data – trends, seasonality, cyclicity and irregularity, most time series data get impacted because of **trend** and **seasonality** which results into some of the regression like assumptions getting violated.



Addressing the issue

Trends can result in a varying mean over time, whereas seasonality can result in a changing variance over time.

Resolving the issue

We decompose the time series and remove parts obstructing regression modelling i.e., trend and seasonality.

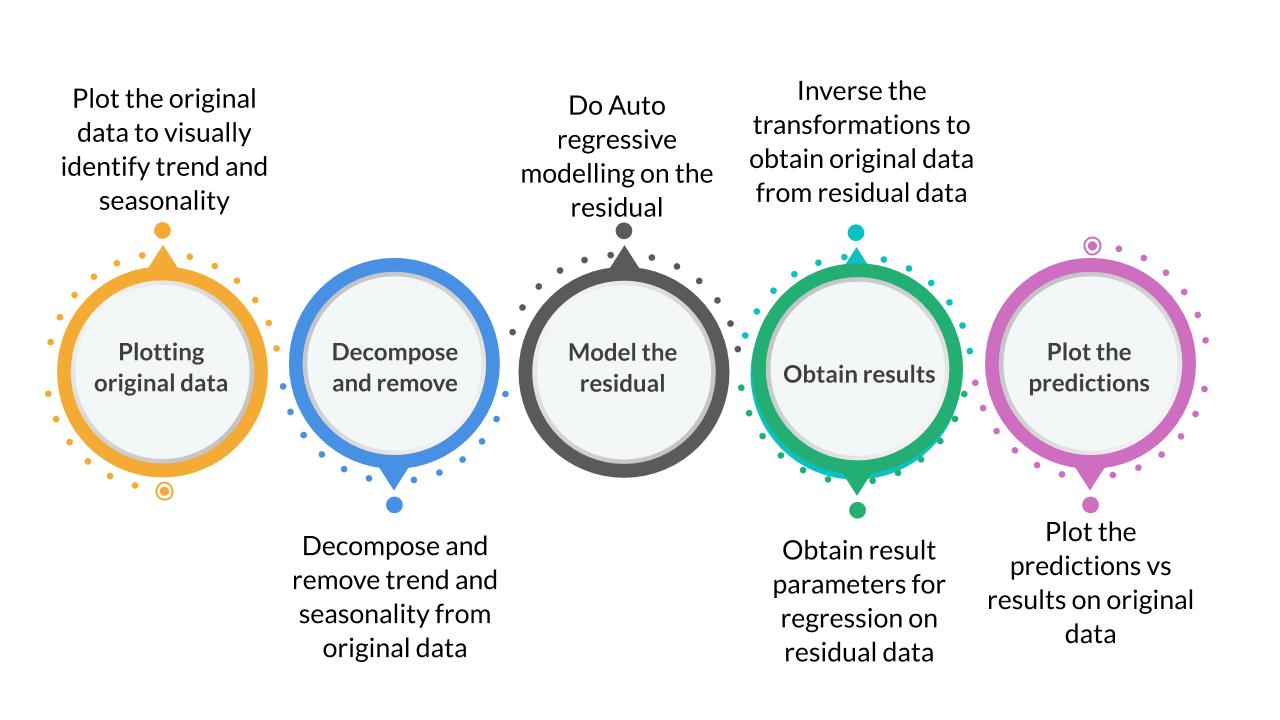
Trend Issue

Identifying and removing trend leads to simpler model design and better model performance.

Seasonality Issue

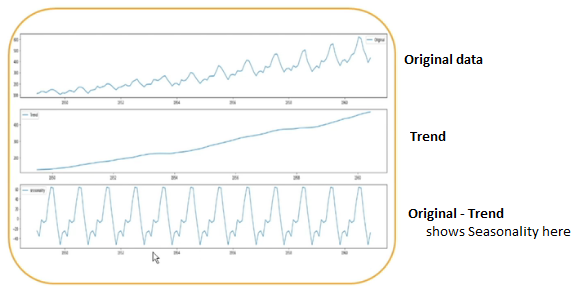
Identifying and removing seasonality can result in a clearer relationship between input and output variables.

**Auto-regressive modelling flow**



**Checking Trend and Seasonality**

Most of the times, the series has a clear trend or seasonal component which can be identified just visually.



However, there is a constraint that you cannot test whether its statistical properties like mean, variance etc.

**Augmented Dickey Fuller (ADF) Test**

ADF is a test for a statistical hypothesis.

The Null hypothesis of the test basically declares the presence of trend and seasonality in the time-series data.

* ***Null Hypothesis (H0)***: The series has trend and seasonality.
  + p-value >0.05
* ***Alternate Hypothesis (H1)***: The series do not have trend and seasonality.
  + p-value <=0.05

Interpretation of p-value

A p-value below a threshold (such as 5% or 1%) suggests we reject the Null Hypothesis, otherwise, a p-value above the threshold suggests we fail to reject the Null Hypothesis.

* ***p-value >0.05***: Fail to reject the Null Hypothesis (H0).
* ***p-value <=0.05***: Reject the Null Hypothesis (H1).

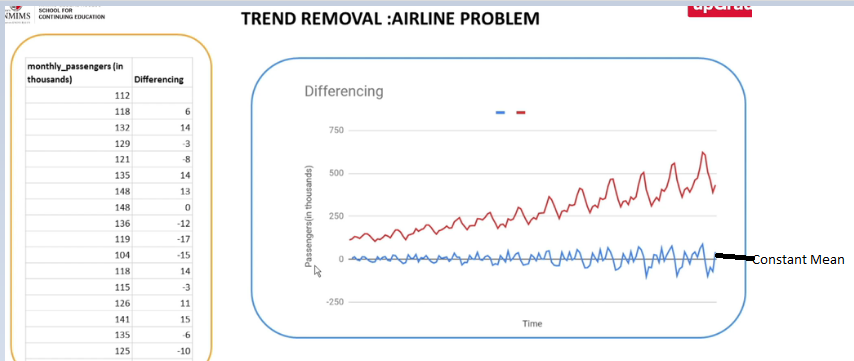
**Trend and Seasonality Removal**

**TREND Removal**

Simplest way of removing trend is ***Differencing Method***.

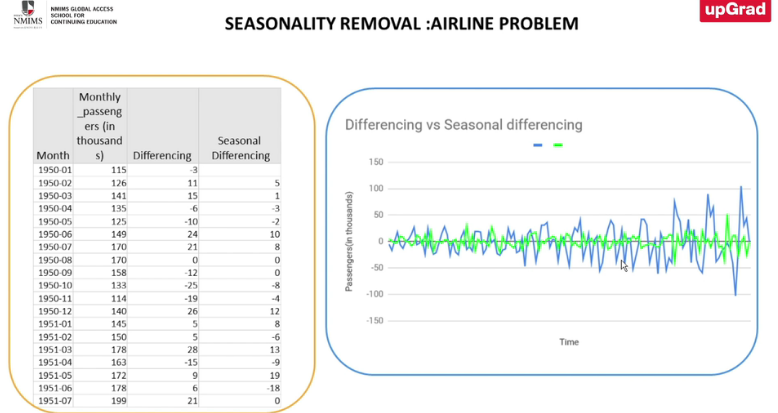
Differencing Method

* Key points for performing differencing on time-series data:
* Take the gap between the data samples in an attempt to make the sequence stationary.
* Let’s say, the original time series was Z1, Z2, Z3, . . . . , Zn.
* Series with a difference of Degree 1 becomes: Z2 – Z1, Z3 – Z2, …. Zn – Zn-1.
* Take the difference, plot the sequence to see if there is any change in the curve. If the change shows unsatisfactory results, you can implement a 2nd or 3rd order differencing.



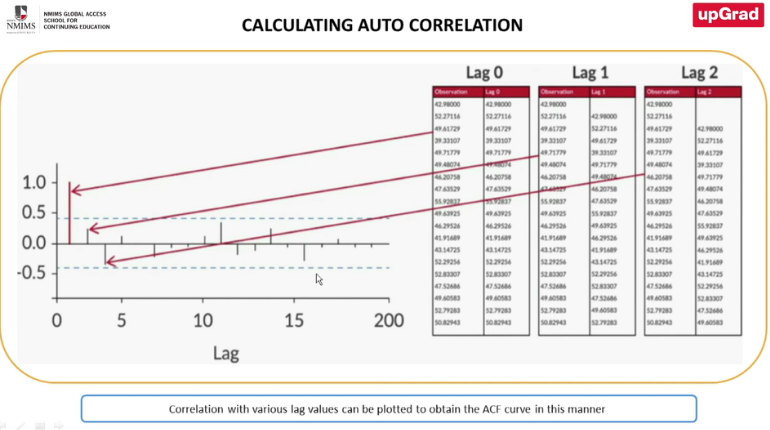
**Seasonality Removal**

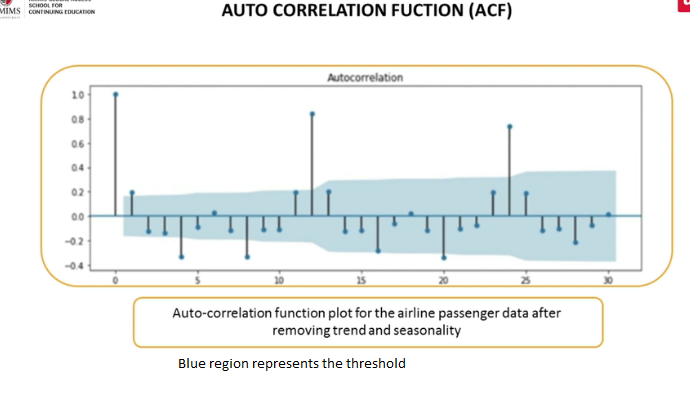
Here, we obtain datapoints by differencing with the values seasonally lagging behind to current values.



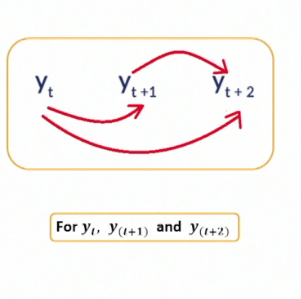
**Auto-correlation Function**

* To evaluate the relation among time series data, we use a function called as autocorrelation function.
* Autocorrelation captures the relationship between observations Yt at time t and Yt-k at time k time period before t.
* Autocorrelation measures the linear relationship between the different entries’ data similar to how correlation evaluates.
* The autocorrelation value can be found out by evaluating correlation among values data points.
* The ACF values that cross a ‘threshold’ mark are relevant for forecasting modelling.
* To calculate the autocorrelation among different timestamps, the time series is lagged accordingly. E.g.,
* Lag 1 evaluates auto correlation between T1 and T2.
* Lag 2 evaluates auto correlation between T1 and T3.
* ACF values that exceed a *threshold* are relevant for forecast modelling.





**ACF Constraints**



* *Autocorrelation function captures both direct and indirect relationships with its lagged values*.
* When you are doing a higher period autocorrelation calculation, you would see that it is **NOT a pure effect**. i.e., when you look at correlation between Yt and Yt+2, you cannot say that entire correlation is because these two values are correlated because Yt+1 may also be playing a role.
* So, auto-correlation captures both the direct effect and also the indirect relationship between Yt and Yt+2 through something in between ( e.g., Yt+1)
* So, we cannot differentiate out only the direct relationship . This problem is solved by PACF (Partial Auto-correlation function).

**PACF (Partial Auto-correlation function)**

PACF is the correlation of the **residuals**. Purpose is that it **captures ONLY direct** relationships.

**What is a white noise series?**

* White noise is a series which has zero autocorrelation (meaning no values cross the threshold). For such series, no prediction models can be built.
* A white noise has zero mean, and its variance remains constant over time.

**Auto – Regressive Time Series Models**

**Simple Auto-Regressive (AR) Model**

A simple auto-regression model is a simple linear regression model between target variable and **previous results** of the variable.

An auto-regressive model of order ‘p’ is written as AR(p) model, where order defines the highest number of lag observations considered for creating this regression model.

Determining parameter ‘p’

1. **Plot PACF Curve** – Autoregression relates to PACF curve to measure direct impacts of previous time series.
2. **Identify significance**- consider only those variables which cross the threshold and hold significant importance.
3. **Determine p** – Select p as the highest lag where partial autocorrelation is significantly high. Hence, we select p =12(example)
4. **Modelling** – Build a regression model from the independent variable.
   * Yt = aYt-1 + bYt-2 + cYt-4

(only those terms are considered whose coefficients are greater than threshold)

**Moving Average Model (MA)**

The moving average model uses previous **forecast errors** in a regression model instead of incorporating previous values in a regression.

A moving average model of order ‘r’ is written as MA(r) model, where order defines the highest number of lag observations considered for creating this model.

Determining parameter ‘r’

1. **Plot ACF Curve**.
2. **Identify significance** – consider only those values which are above threshold.
3. **Determine ‘r’** – Select ‘r’ as the highest lag where autocorrelation is significantly high.
4. **Modelling** – Build a moving average from the independent variable’s past observations Yt-1, Yt-2 (which are greater than threshold and hence significant) to predict the dependent variables.

**Auto Regressive Moving Average Model (ARMA)**

The ARMA model takes into account the combined effect of the previous time series values as well as the errors in their prediction to come up with a suitable prediction.

ARMA model is written as ARMA (p,r) model, a model with autoregression of order ‘p’ and moving average of order ‘r’.

Determining parameters ‘r’ and ‘p’

* Plot the PACF curve.
* Determine autoregression order ‘p’ for the model.
* Plot the ACF curve.
* Determine the moving average order ‘r’ for the model.
* Generate an ARMA (p,r) model.