Time Series

* **Time series**: when all other factors are constant prediction of future values.
* **Different Methods of doing Time series Analysis and Forecasting**
  + **ARIMA model.**
  + **Seasonally ARIMA.**  Most used.
  + **Holt Winter Exponential Smoothing.** Easiest and effective model. [Link](https://youtu.be/O6cUkdQeLUQ)

1. **Important Concepts and terminology** in Time series Analysis.
   * Stationarity. To know everything follow the [link.](https://youtu.be/R69TZFNEao4)
     + A stationary time series is one whose properties (i.e mean, variance, autocorrelation) does not depend on the time.
   * Autoregression AR.
   * Moving Average MA
   * Integration & Difference
   * ACF and PACF Plots
   * **Time series components:**
     + Trend: long term smooth movement, upward or downward
     + Seasonal: periodic fluctuation, less than 1 year, most commonly found in industry.
     + Cyclical: periodic fluctuation, more than 1 year.
     + irregularity: random movement.

Arima

* ARIMA is the Most common model used for time series forecasting. It has 3 components.
  1. Autoregression AR.
  2. Moving Average MA
  3. Integrated

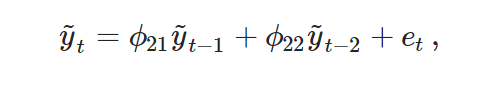
1. **Autoregression AR.**
   * Future values of Y is dependent of previous lagged values of Y.

regression of yt on yt-1, yt-2 .

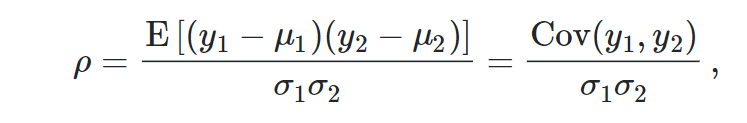
* + **P = ORDER OF AR**; current value of y is dependent on how many previous lagged value of current Y. if p=2 that means yt is dependent on yt-1 and yt-2.
  + **P** from **PACF**
  + Interpretation of PACF:

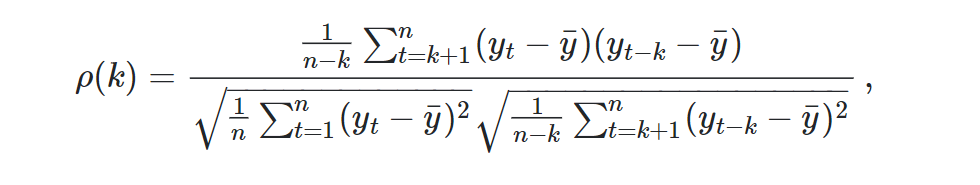
1. **Moving Average MA.**
   * Future values of Y is dependent of previous lagged values of **white noise** i.e. the irregular component. white noise is just the error. error is the difference between the actual value and predicted value. So, we take into consideration the error also to predict the future value.
   * autocorrelation between the errors.
   * Trend, s, c components of TS is captured in AR whereas the irregular comp is captured in MA.
   * q is order of MA.
   * **ACF** gives q.
2. **Integrated**
   * Intergrated means no of times we difference the data then we have to integrated it back to get the original series back.
   * We difference to remove trend and seasonality to it stationary series as only after making a series stationary we can implement AR and MA.

ACF and PCF



**ACF** is an (complete) auto-correlation function which gives us values of auto-correlation of any series with its lagged values. We plot these values along with the confidence band and tada! We have an ACF plot. In simple terms, it describes how well the present value of the series is related with its past values. A time series can have components like trend, seasonality, cyclic and residual. ACF considers all these components while finding correlations hence it’s a ‘complete auto-correlation plot’.





**PACF** is a partial auto-correlation function. Basically, instead of finding correlations of present with lags like ACF, it finds correlation of the residuals (which remains after removing the effects which are already explained by the earlier lag(s)) with the next lag value hence ‘partial’ and not ‘complete’ as we remove already found variations before we find the next correlation. So, if there is any hidden information in the residual which can be modeled by the next lag, we might get a good correlation and we will keep that next lag as a feature while modeling. Remember while modeling we don’t want to keep too many features which are correlated as that can create multicollinearity issues. Hence, we need to retain only the relevant features