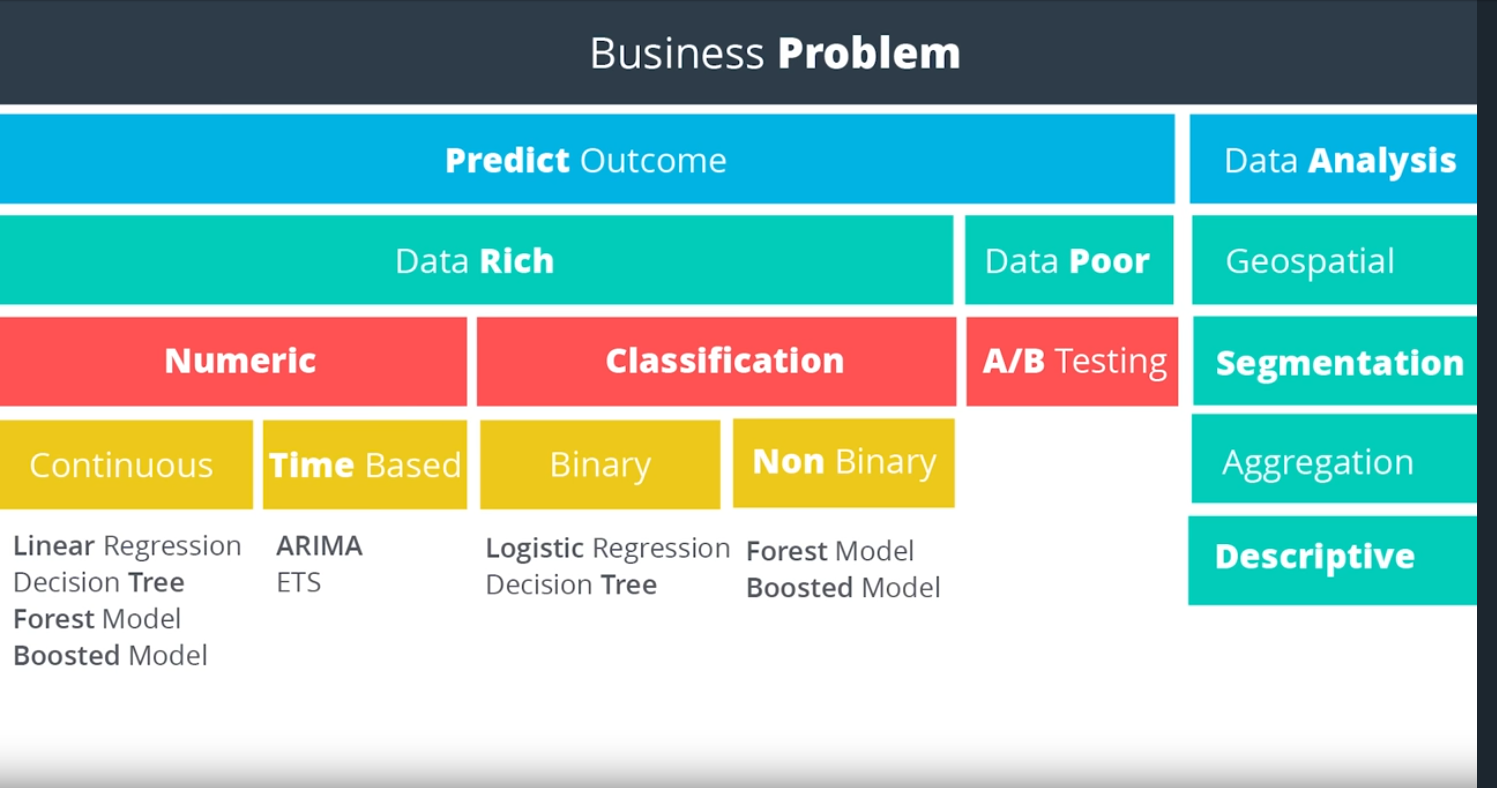
1. Time series forecasting

Time series forecasting is the use of a model to predict future values based on previously observed values.

1. Business Problem



ETS: Exponential Smoothing

ARIMA: Autoregressive Integrated Moving Average

1. Attributes of a Time Series

* It's over a continuous time interval
* There are sequential measurements across that interval
* There is equal spacing between every two consecutive measurements
* Each time unit within the time interval has at most one data point
* Order Matters: There is a dependency on time and changing the order could change the meaning of the data.

1. Simple Exponential Smoothing in Excel

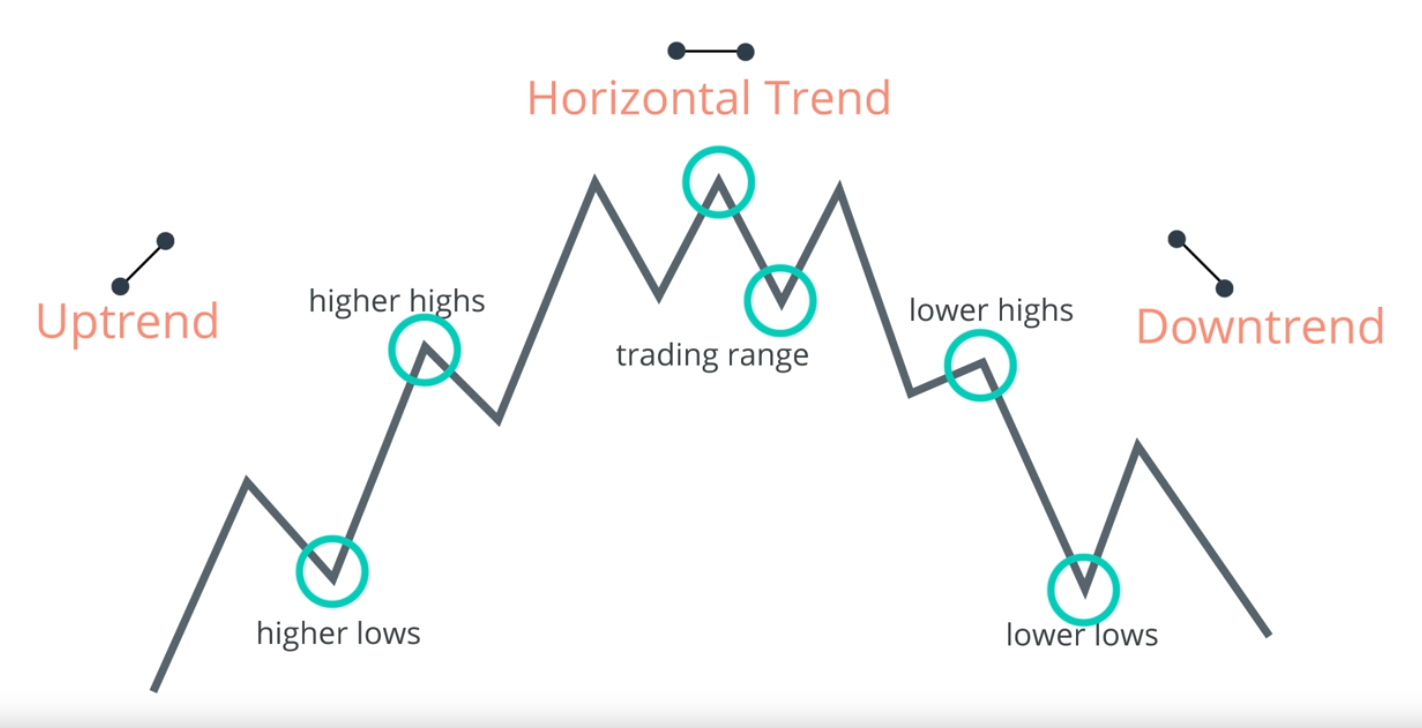
<http://www.excel-easy.com/examples/exponential-smoothing.html>

1. Time Series Components

* Trend (uptrend, downtrend, horizontal trend)
* Seasonal pattern
* Cyclical pattern

Seasonality is always of a fixed and known period. Hence, seasonal time series are sometimes called periodic time series. A **cyclic pattern** exists when data exhibit **rises and falls that are not of fixed period**

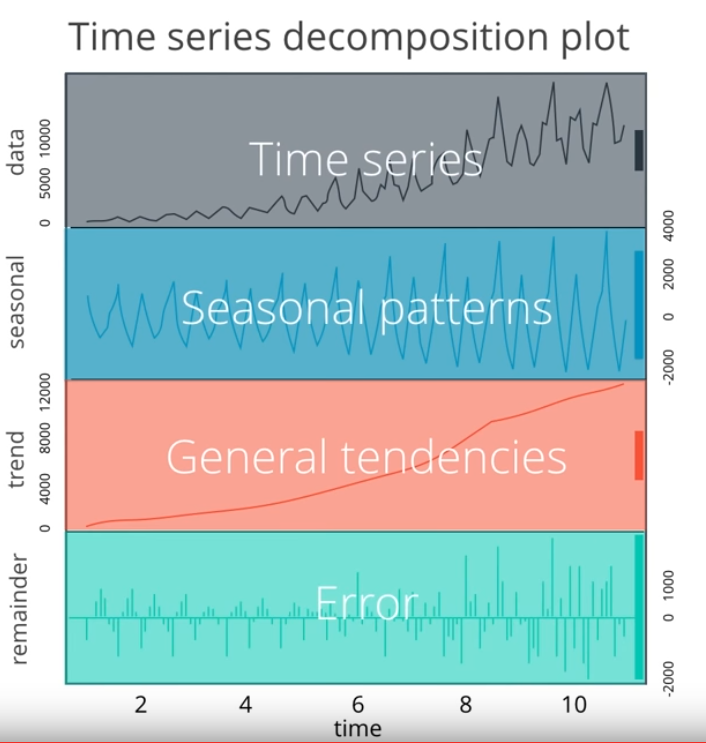
1. Trend



1. ETS (Error ,Trend, Seasonality)🡪**Additive** or **Multiplicative** terms

Exponential Smoothing -More weight on the most recent observations. The weights decreasing exponentially as the observations get older.







### Scenarios

Therefore the scenarios could be:

* No-Trend, No-Seasonal
* No-Trend, Seasonal-Constant
* No-Trend, Seasonal-Increasing
* Trend-Linear,No-Seasonal
* Trend-Linear,Seasonal-Constant
* Trend-Linear,Seasonal-Increasing
* Trend-Exponential,No-Seasonal
* Trend-Exponential,Seasonal-Constant
* Trend-Exponential,Seasonal-Increasing

As you can see there are nine possible scenarios.

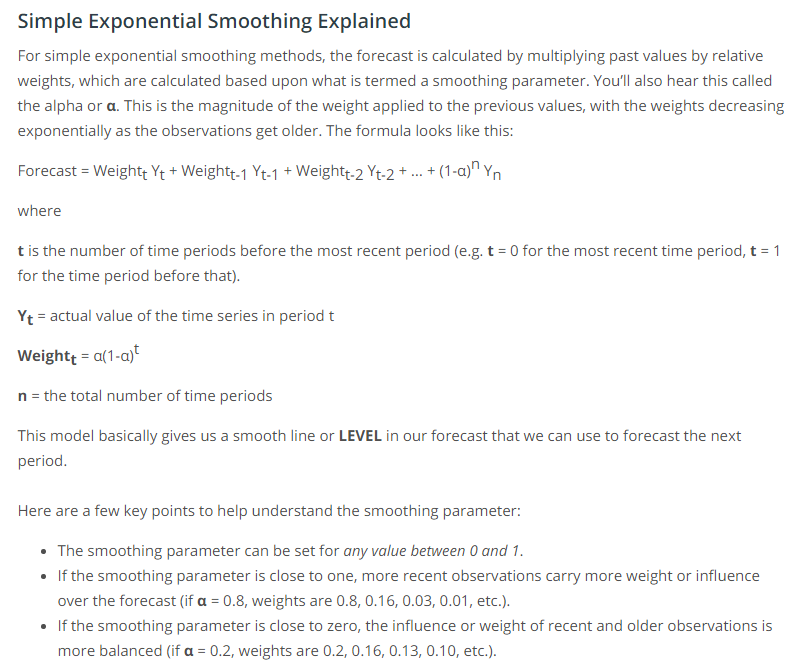
### ETS Models

We are going to explore four ETS models that can help forecast these possible time-series scenarios.

1. Simple Exponential Smoothing Method (suitable only for time series without trend and seasonality)
2. Holt's Linear Trend Method
3. Exponential Trend Method
4. Holt-Winters Seasonal Method
5. Simple Exponential Smoothing Method

The series does not have a trend line and does not have seasonality component. We should use a Simple Exponential Smoothing model.





#### Choosing the Smoothing Parameter α

Choosing the correct smoothing parameter is often an iterative process. Luckily, advanced statistical tools, like Alteryx, will select the best smoothing parameter based upon minimizing forecasting error. Otherwise, you will need to test many smoothing parameters against each other to see which model best fits the data.

The advantage of exponential smoothing methods over simple moving averages is that new data is depreciated at a constant rate, gradually declining in its impact, whereas the impact of a large or small value in a moving average, will have a constant impact. However, this also means that exponential smoothing methods are more sensitive to sudden large or small values.

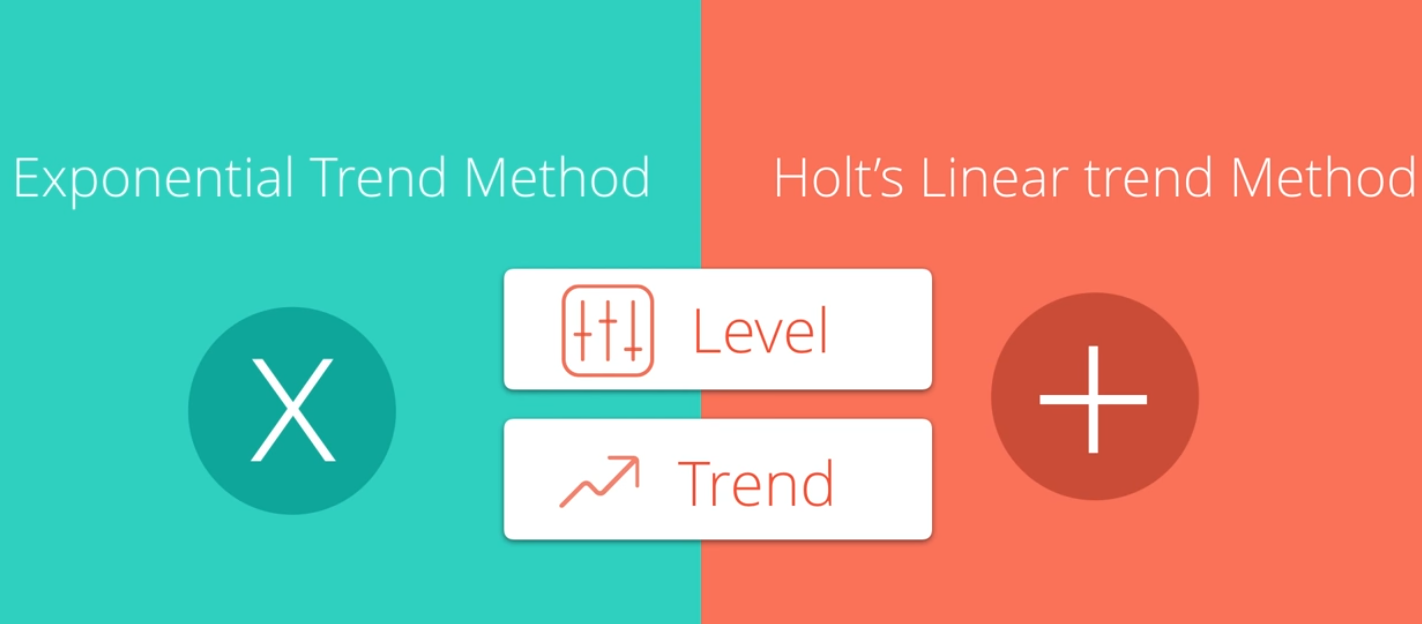
The simple exponential smoothing method does not account for any trend or seasonal components, rather, it only uses the decreasing weights to forecast future results. This makes the method suitable only for time series without trend and seasonality.

Reference: <https://www.otexts.org/fpp/7/1>

1. Holt's Linear Trend Method



1. Exponential Trend Method



1. Holt-Winters Seasonal Method



## What You've Learned So Far

Let's take a step back and understand what we've learned so far.

### Methods

There are several methods we need to pick in order to model any given time series appropriately:

1. Simple Exponential Smoothing
   * Finds the level of the time series
2. Holt's Linear Trend
   * Finds the level of the time series
   * Additive model for linear trend
3. Exponential Trend
   * Finds the level of the time series
   * Multiplicative model for exponential trend
4. Holt-Winters Seasonal
   * Finds the level of the time series
   * Additive for trend
   * Multiplicative and Additive for seasonal components

These methods help deal with different scenarios in our time series involving:

1. Linear or exponential trend
2. Constant or increasing seasonality components

For trends that are exponential, we would need to use a **multiplicative** model.

For increasing seasonality components, we would need to use a **multiplicative model** model as well.

### ETS

Therefore we can generalize all of these models using a naming system for ETS:

#### ETS (Error, Trend, Seasonality)

Error is the error line we saw in the time series decomposition part earlier in the course. If the error is increasing similar to an increasing seasonal components, we would need to consider a multiplicative design for the exponential model.

Therefore, for each component in the ETS system, we can assign None, Multiplicative, or Additive (or N, M, A) for each of the three components in our time series.

#### Examples

A time series model that has a constant error, linear trend, and increasing seasonal components means we would need to use an ETS model of:

##### ETS(N,A,M)

A time series model that has increasing error, exponential trend, and no seasonality means we would need to use an ETS model of:

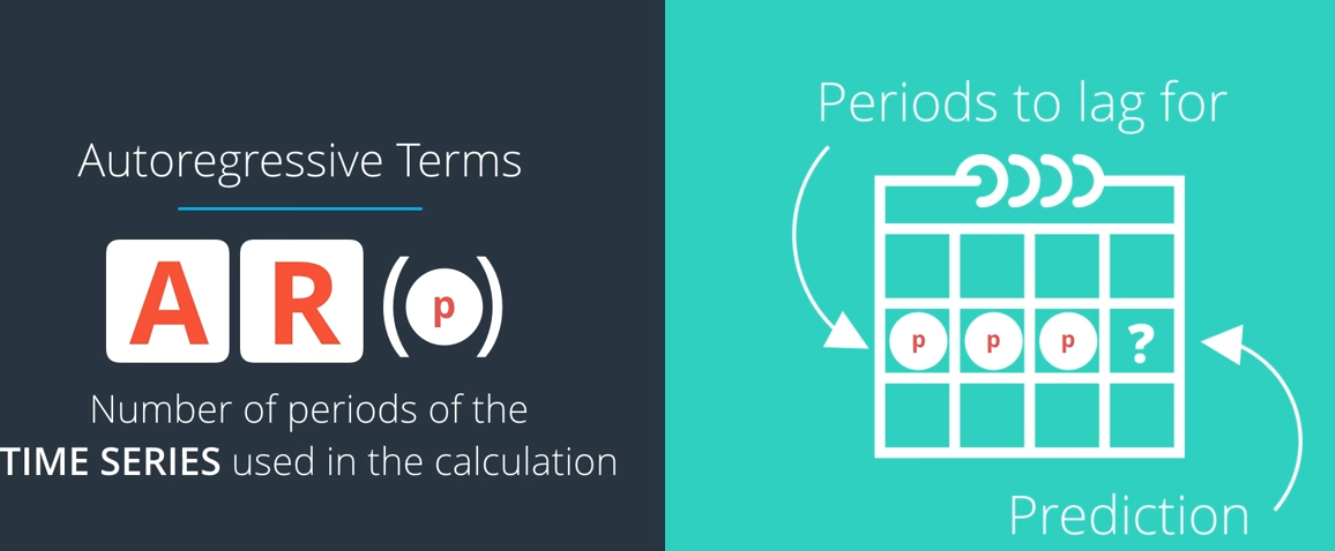
##### ETS(M,M,N)

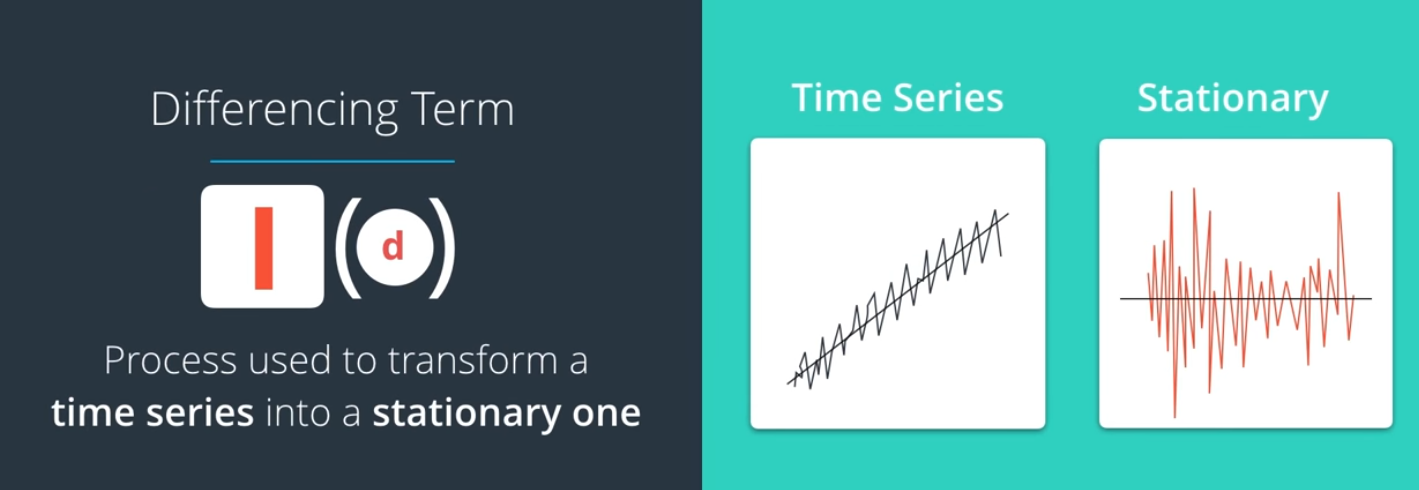
1. ARIMA (Autoregressive Integrated Moving Average)

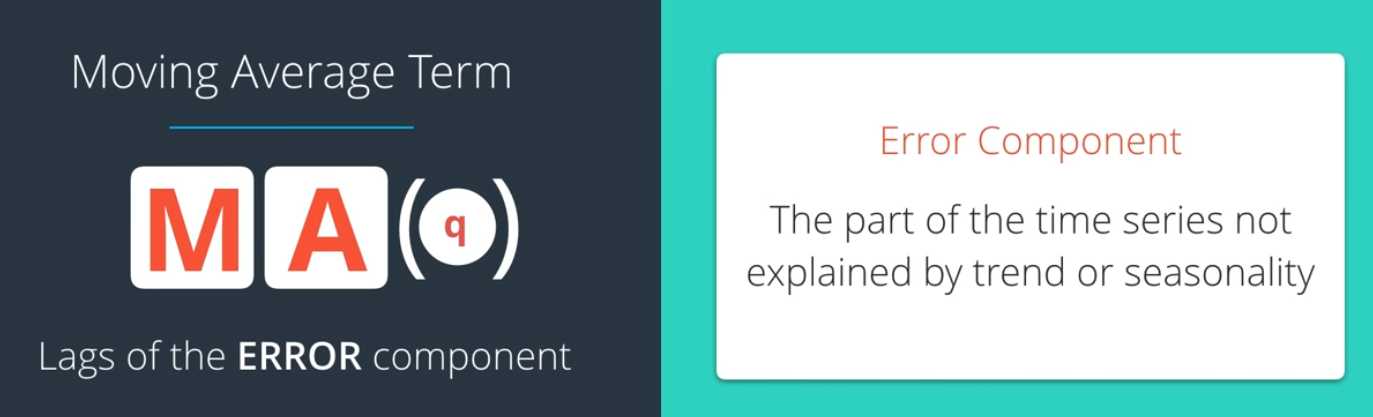
Seasonal ARIMA + Non seasonal ARIMA



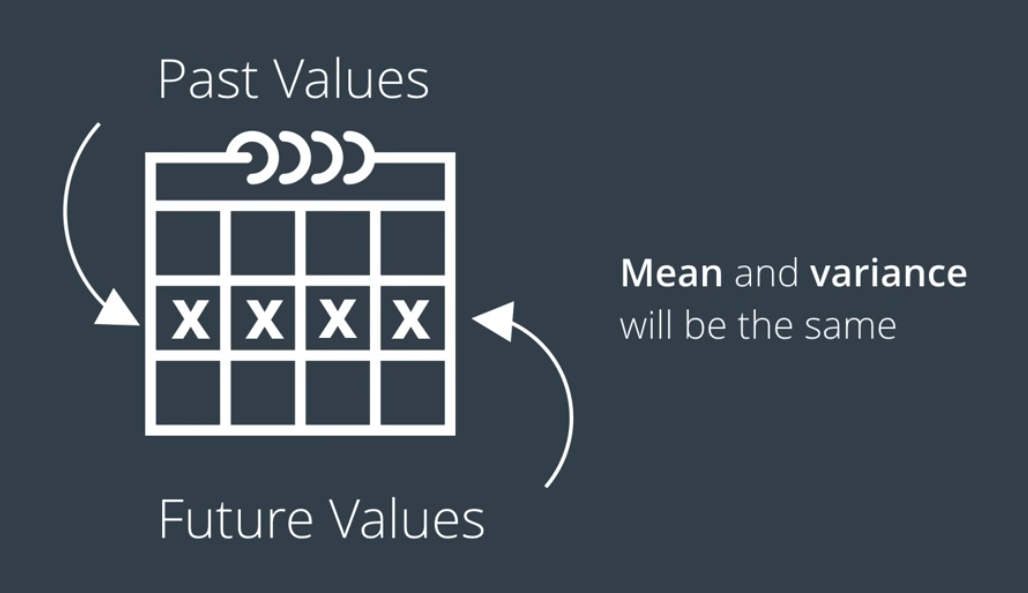
(p,d,q)







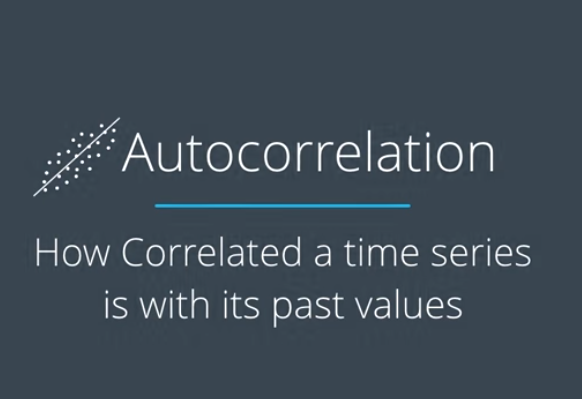
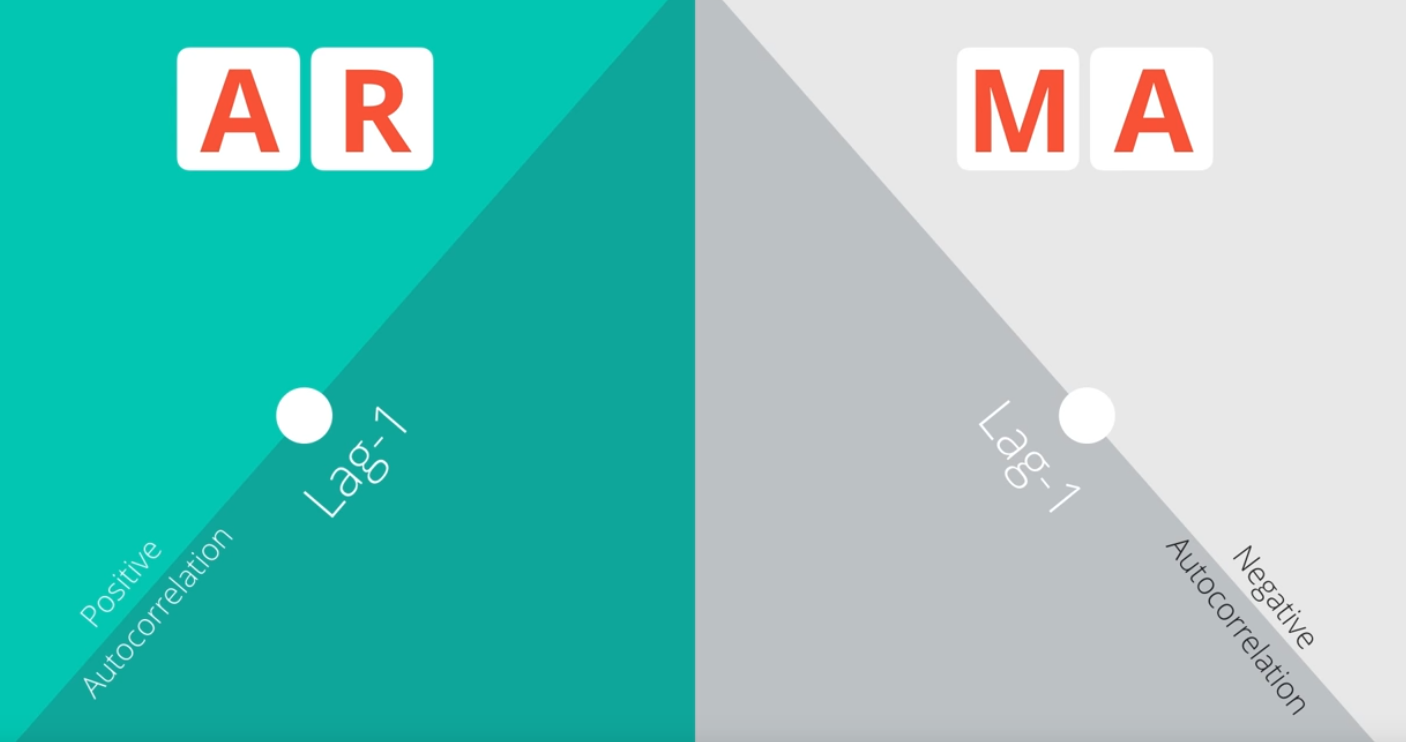
Stationary time series: Mean and the variance are constant over time.



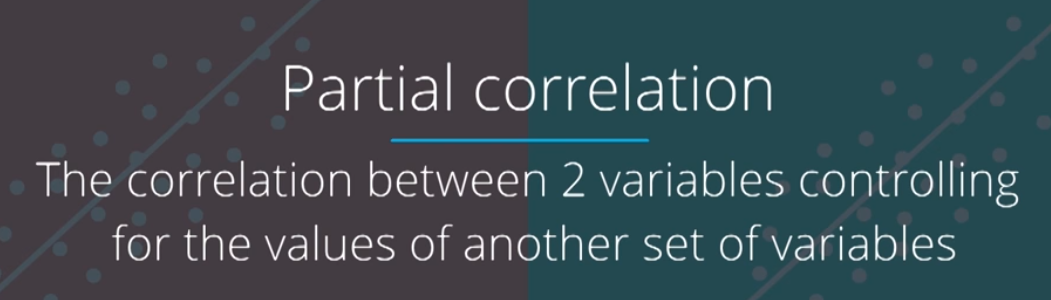
1. Differencing

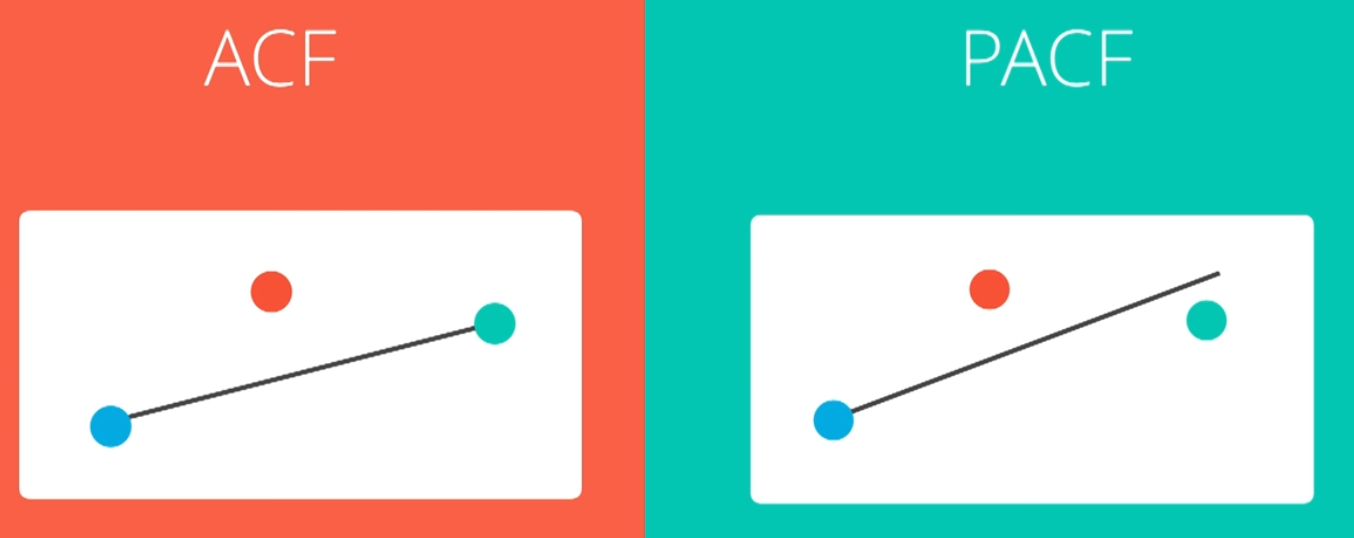
Differencing is a method of transforming a non-stationary time series to a stationary one. This is an important step in preparing data to be used in an ARIMA model. Let’s go through an example to understand differencing.

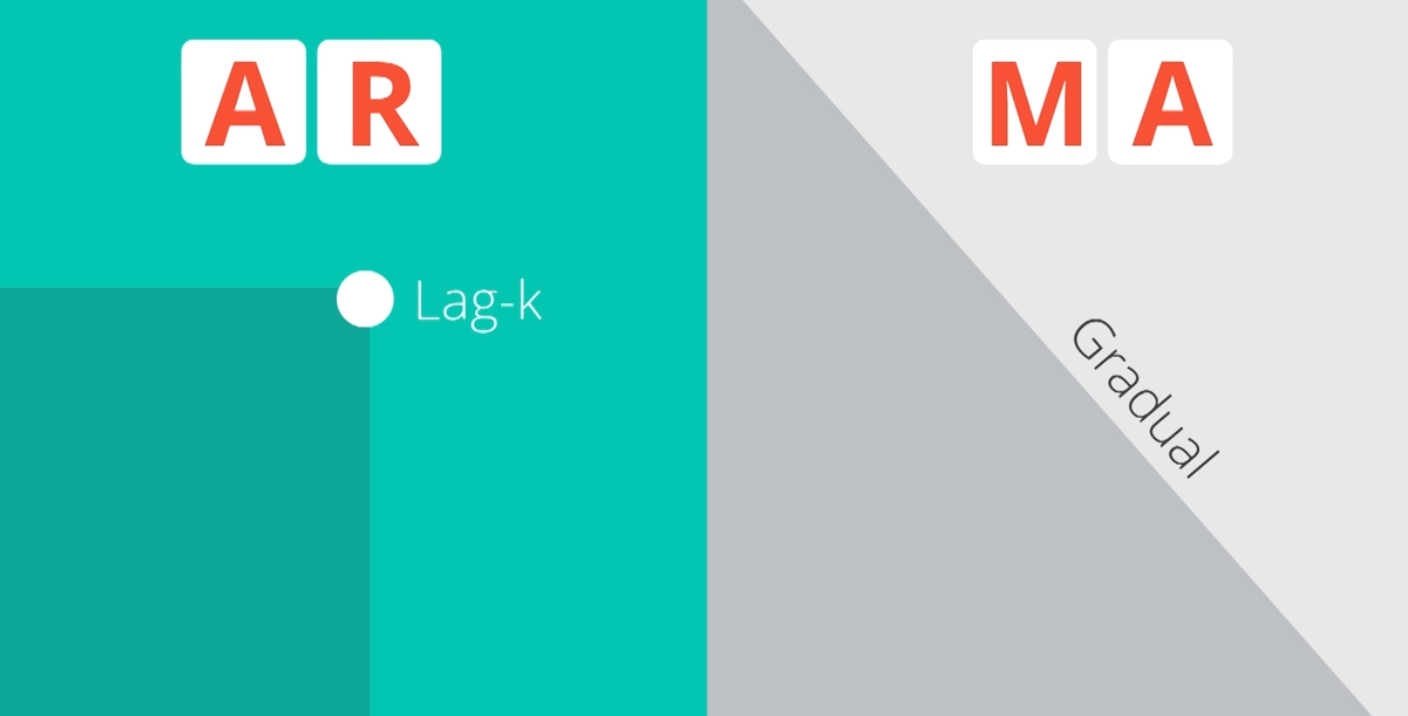
1. Autocorrelation

1. Partial-Autocorrelation







Partial-correlation: <https://onlinecourses.science.psu.edu/stat510/node/62>

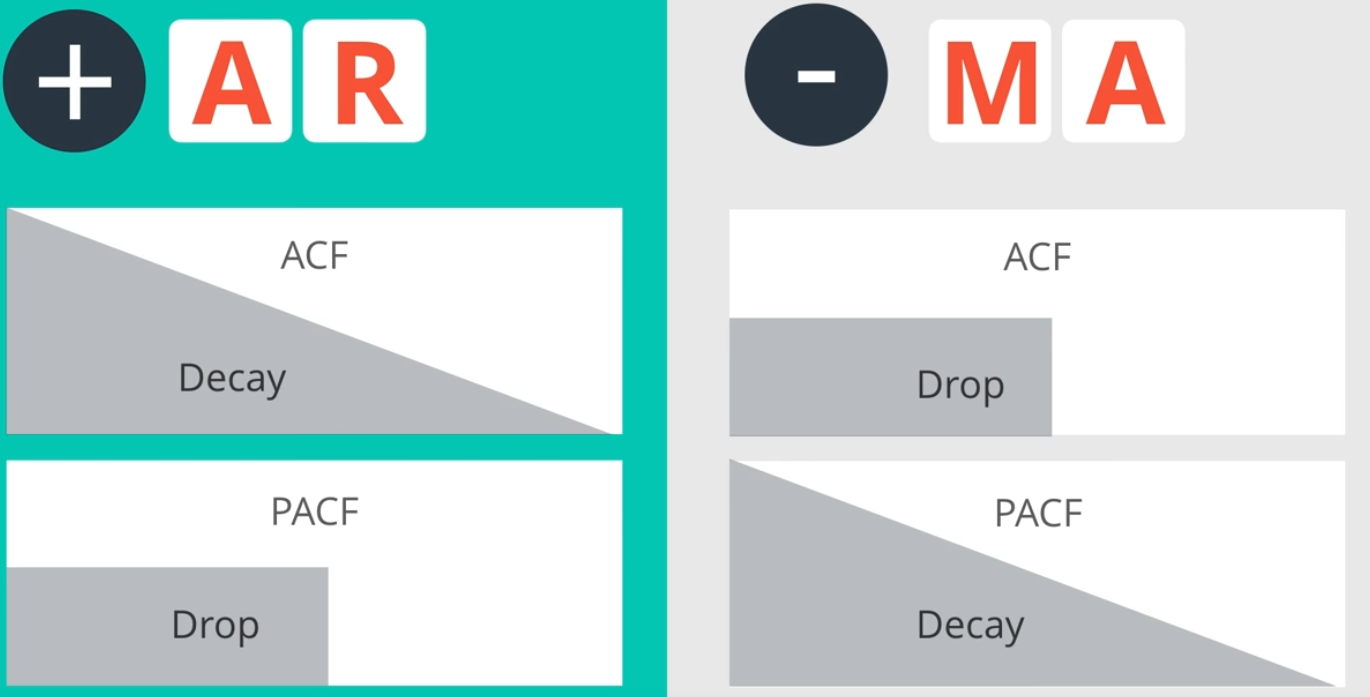
Duke: <http://people.duke.edu/~rnau/411arim3.htm>

1. Seasonal ARIMA Model

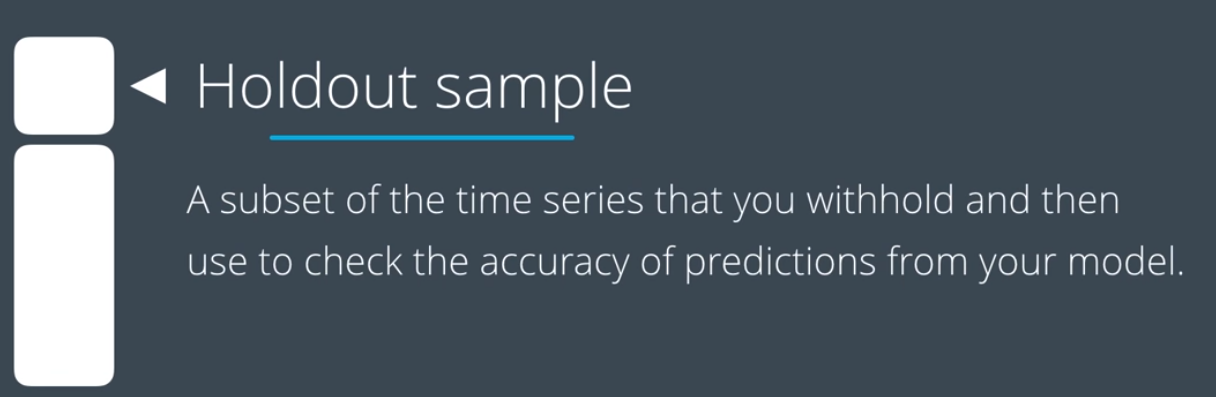




**Partial Autocorrelation Function (PACF)**



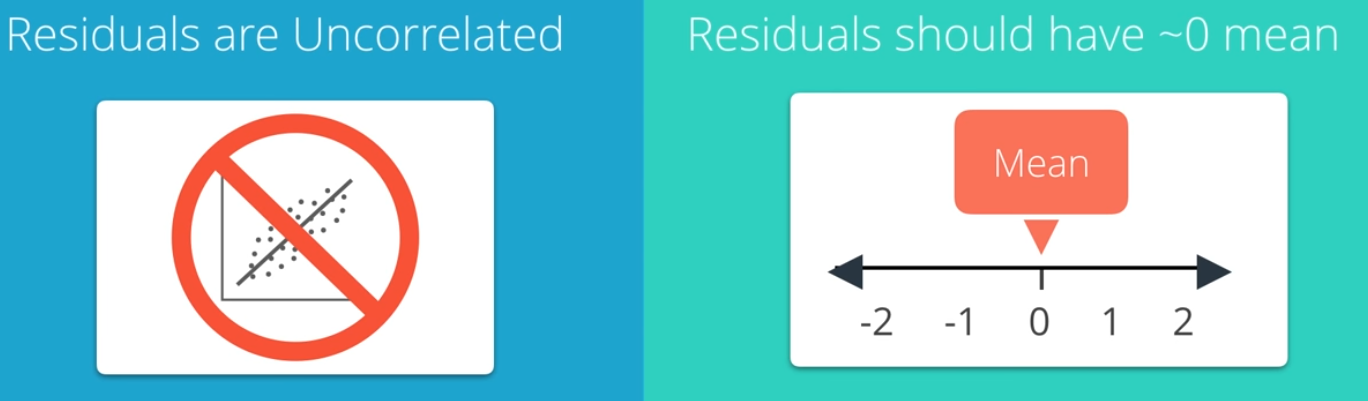
1. Holdout Sample



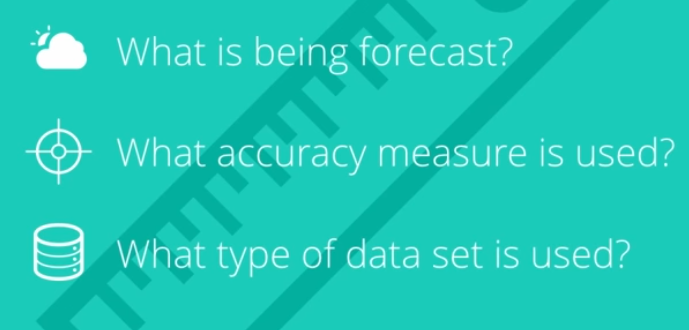
1. Residual



* Residuals are uncorrelated
* Residuals should have 0 mean



1. Calculating Error



## Interpreting Measures of Error

### Scale Dependent Errors

Scale dependent errors, such as mean error (ME) mean percentage error (MPE), mean absolute error (MAE) and root mean squared error (RMSE), are based on a set scale, which for us is our time series, and cannot be used to make comparisons that are on a different scale. For example, we wouldn’t take these error values from a time series model of the sheep population in Scotland and compare it to corn production forecast in the United States.

* **Mean Error (ME)** shows the average of the difference between actual and forecasted values.
* **Mean Percentage Error (MPE)** shows the average of the percent difference between actual and forecasted values. Both the ME and MPE will help indicate whether the forecasts are biased to be disproportionately positive or negative.
* **Root Mean Squared Error (RMSE)** represents the sample standard deviation of the differences between predicted values and observed values. These individual differences are called residuals when the calculations are performed over the data sample that was used for estimation, and are called prediction errors when computed out-of-sample. This is a great measurement to use when comparing models as it shows how many deviations from the mean the forecasted values fall.
* **Mean Absolute Error (MAE)** takes the sum of the absolute difference from actual to forecast and averages them. It is less sensitive to the occasional very large error because it does not square the errors in the calculation.

### Percentage Errors

Percentage errors, like MAPE, are useful because they are scale independent, so they can be used to compare forecasts between different data series, unlike scale dependent errors. The disadvantage is that it cannot be used if the series has zero values.

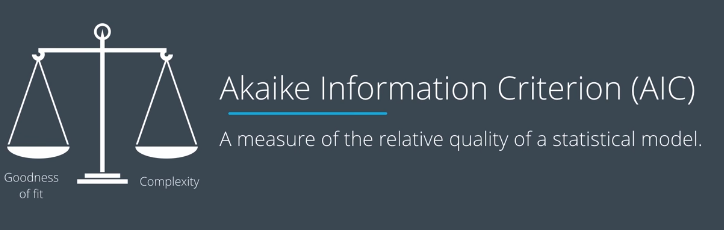
* **Mean Absolute Percentage Error (MAPE)** is also often useful for purposes of reporting, because it is expressed in generic percentage terms it will make sense even to someone who has no idea what constitutes a "big" error in terms of dollars spent or widgets sold.

### Scale-Free Errors

Scale-free errors were introduced more recently to offer a scale-independent measure that doesn't have many of the problems of other errors like percentage errors.

* **Mean Absolute Scaled Error (MASE)** is another relative measure of error that is applicable only to time series data. It is defined as the mean absolute error of the model divided by the the mean absolute value of the first difference of the series. Thus, it measures the relative reduction in error compared to a naive model. Ideally its value will be significantly less than 1 but is relative to comparison across other models for the same series. Since this error measurement is relative and can be applied across models, it is accepted as one of the best metrics for error measurement.

1. Akaike Information Criterion (AIC)



AIC: <https://coolstatsblog.com/2013/08/14/using-aic-to-test-arima-models-2/>