# Initial Posts

|  |
| --- |
| **Time Series Analysis**  Time series analysis is the process of finding the relationship that certain intervals of time have on an outcome. For example, if the price of cell phones spiked up 3 months ago, what sort of effect will that have, if any, on the price of cell phones today? Time series analysis has many applications, one of the major ones being the stock market and predicting future stock prices based on previous stock prices. |
| **Time vs. Non-Time as Predictors**  When developing a predictive model using time as a predictor, you might be tempted to fit a model similarly to the way you would with a non-time predictor. It is important however to understand why time is different than a non-time variable when used as a predictor.  When predicting future data using time as the predictor you are basically predicting something that has never happened before (aka tomorrow or some other date in the future). When using other non-time variables however, you are basing your prediction on something that has happened. For example, predicting sunscreen prices using temperature as the predictor. The temperatures you're interested in are probably within a range of temperatures that have existed on Earth before...but technically you don't have to...(*I predict the sale of sunscreen will be 0.1x10^-100 if the temperature is 1,000,000F* 🥵). |
| **Why Weighting in Time Series Analysis is Important**  Weighting rolling aggregations (aka EWMA) is important in Time Series Analysis because it more accurately represents the predictive power that further away intervals have on future intervals. This makes sense if you think about it in terms of stock prices. The stock price yesterday is probably better at predicting the stock price tomorrow vs. using the stock price 15 days ago as the predictor. Weighted rolling aggregations account for this by exponentially dropping predictive power the further away you are from the current interval being analyzed. |
| **Handling Missing Values**  In Think Stats, Downey describes 2 different methods to fill missing values where an interval does not have any data. The first method is to calculate some sort of moving average and then fill the missing data with that value. "A drawback of this method is that it understates the noise in the series. We can solve that problem by adding in resampled residuals" (Downey, 2014). To do this, we go a step further than just filling missing values with a moving average and instead:   1. Calculate the residual from the actual value and moving average. 2. Create a new column that adds a randomly selected residual from the residuals calculated in step #1 to the moving average also calculated in step #1. 3. If the original data set as a null value, fill it with the value calculated in step #2   Reference: Downey, Allen B.. Think Stats: Exploratory Data Analysis (p. 212). O'Reilly Media. |
|  |

# Replies

|  |
| --- |
| Yeah, that function looks like it has been replaced by ewm(). You'll have to do something like below: |
| Well this post sent me down a Google rabbit hole but now I know what ARIMA is...at least I think I do. From what I've found there are multiple different models you can build from a time series.   * AR (Auto-regression aka autocorrelation). The Autocorrelation Function or ACF can help determine if AR is a good model for the data and also can help determine the order of AR vs. MA in an ARMA, ARIMA, or SARIMA model. * MA (Moving Average). The Partial Autocorrelation Function or PACF can help determine if MA is a good model for the data and also can help determine the order of AR vs. MA in an ARMA, ARIMA, or SARIMA model. * ARMA (Auto Regressive Moving Average). Incorporating both AR & MA. Without the 'I' like in the ARIMA model, this indicates that the initial data was stationary and no differencing was needed to transform the data into a stationary form. * ARIMA (Auto Regressive Integrated Moving Average). Same as ARMA but the data had to be transformed into a stationary form before analysis. * SARIMA (Seasonal ARIMA). Same as ARIMA but you are decomposing seasonal trends to make the data stationary. |
| Very nice find, thank you Amelia. I love that the article shows examples of code for each of the methods described. The article also does a good job explaining what the differences are between each of the methods. |
| Another great article, thank you Anjani. This article also has a lot of great python examples. Will have to bookmark both of these. |
| Great summary of moving average Chandrasekhar. As you mentioned, moving average is great at smoothing out unexpected variances in the data which would cause the function to fit the line to be more complex. |