Autocorrelation

Autocorrelation analysis is an important step in the Exploratory Data Analysis of time series forecasting.
The autocorrelation analysis helps detect patterns and check for randomness.

Autocorrelation is the correlation between a time series with a lagged version of itself.

Any autocorrelation that may be present in time series data is determined using a correlogram, also known as an ACF plot. This is used to help you determine whether your series of numbers is exhibiting autocorrelation at all, at which point you can then begin to better understand the pattern that the values in the series may be predicting.

An **autoregressive model** is when a value from a time series is regressed on previous values from that same time series.

for example, y_t on y_{t-1}:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t.$$

Autocorrelation

The coefficient of correlation between two values in a time series is called the **autocorrelation function** (**ACF**) For example the ACF for a time series yt is given by:

$$\mathrm{Corr}(y_t, y_{t-k}), k = 1, 2, \ldots.$$

The ACF is a way to measure the linear relationship between an observation at time *t* and the observations at previous times.

The key statistics in time series analysis is the autocorrelation coefficient (or the correlation of the time series with itself, lagged by 1, 2, or more periods), which is given by the following formula,

$$r_k = \frac{\sum_{t=k+1}^{n} (Y_t - \bar{Y})(Y_{t-K} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}$$

The value should be between -1 to +1

ACF Example:

Y	Y_{t-1}	$Y_t-\bar{Y}$	$Y_{t-1}-\bar{Y}$	$(Y_t-\bar{Y})(Y_{t-1}-\bar{Y})$	$(Y_t-\bar{Y})(Y_t-\bar{Y})$
2		-4			16
3	2	-3	-4	12	9
5	3	-1	-3	3	1
7	5	1	-1	-1	1
9	7	3	1	3	9
10	9	4	3	12	16
Total				29	52

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t-\overline{Y})(Y_{t-K}-\overline{Y})}{\sum_{t=1}^{n}(Y_t-\overline{Y})^2} = \frac{29}{52} = 0.557692$$