# ARIMA Modelling

To predict the development of the dengue cases with an ARIMA model, a time series with the total dengue cases of Thailand was created.

A requirement for fitting an ARIMA model is that the time-series is stationary. Stationary time series are time-series where the mean value doesn’t change over time, the variance doesn’t increase and the seasonality effect is minimal. Two tests were used to test for stationarity of the data. The Augmented Dickey-Fuller test (AGF) checks whether the time series has a unit root, which would be evidence that the time series is non-stationary. The null-hypothesis is that the time-series has a unit-root and therefore is non-stationary. The alternative hypothesis is that the data is stationary. Therefore a p-value lower than the significance level indicates that the time series is stationary. The Kwiatkowski-Phillips-Schmidt-Shin test (KPSS) is also a unit root test. It tests for the stationarity of a given series around a deterministic trend. In contrast to AGF the null-hypothesis is stationarity. Therefore a high p-value indicates that the time-series is stationary. If the time series is found to be non-stationary the differences between consecutive observations can be calculated and tested for stationarity again.

ARIMA models combine different components. An autoregressive model AR(p) and a moving average model MA(q).

The autoregressive model computes the current value from previous values and the error term:

yt=c+ϕ1yt−1+ϕ2yt−2+⋯+ϕpyt−p+εt

εt = white noise  
 ϕ1,…,ϕp = parameters  
yt-1,…, yt-p = lagged values

For the moving average the current value consists of the mean value of the time series and weighted current and past error terms:

yt=c+εt+θ1εt−1+θ2εt−2+⋯+θqεt−q

θ1,…, θq = parameters

I(t) is the number of times differencing was performed to make the time series stationary

The auto.arima function also considers a

ACF & PACF !!!

A time series can be decomposed in the components trend, seasonality and random. For additive decomposition the original time series is the sum of the different components.

<http://r-statistics.co/Time-Series-Analysis-With-R.html?utm_content=cmp-true>

<https://www.analyticsvidhya.com/blog/2021/06/statistical-tests-to-check-stationarity-in-time-series-part-1/#How_to_Check_Stationarity>?

https://rstudio-pubs-static.s3.amazonaws.com/354672\_b7cb732a2b61469390d6fc72621bc9c4.html