# ARIMA Modelling

To predict the development of the dengue cases with an Autoregressive Integrated Moving Average model (ARIMA), a time series with the total dengue cases of Thailand was created.

A requirement for fitting an ARIMA model is a stationary time-series. This is obtained, when 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 (ADF) test examines whether the time series has a unit root, indicating non-stationarity. The null hypothesis assumes the presence of a unit root, implying non-stationarity, while the alternative hypothesis suggests stationarity. Therefore a p-value below the significance level supports the conclusion that the time series is stationary. On the other hand, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is also a unit root test but focuses on the presence of a deterministic trend in the series. The null hypothesis assumes stationarity, and a high p-value indicates that the time series is indeed stationary. In contrast to AGF-test, for the KPSS-test the null hypothesis assumes stationarity, and a high p-value indicates that the time series is indeed stationary. If the time series is initially found to be non-stationary, the differences between consecutive observations can be calculated, and the stationarity tests can be applied again.

ARIMA models combine 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.

To find the best values for p and q, the Autocorrelation function (ACF) and partial Autocorrelation function (pACF) can be evaluated. The ACF plot shows the correlations of a time-series with lags of itself, while the pACF additionally removes the effects of lags.

A second evaluation tool is the auto.arima function. It automatically fits the best ARIMA model by minimizing the Akaike’s Information Criterion (AIC). The aout.arima function can also consider seasonal models.

A good forecasting method will yield residuals with the following properties:

1. The residuals are uncorrelated. If there are correlations between residuals, then there is information left in the residuals which should be used in computing forecasts.
2. The residuals have zero mean. If the residuals have a mean other than zero, then the forecasts are biased.
3. The residuals have constant variance.
4. The residuals are normally distributed.

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>

When data have a trend, the autocorrelations for small lags tend to be large and positive because observations nearby in time are also nearby in size. So the ACF of trended time series tend to have positive values that slowly decrease as the lags increase.

When data are seasonal, the autocorrelations will be larger for the seasonal lags (at multiples of the seasonal frequency) than for other lags.