# ARIMA Modelling

To predict the development of the dengue cases with an Auto-Regressive Integrated Moving Average model (ARIMA), a time series with the total dengue cases of Thailand was created. A time series can be decomposed in the components trend, seasonality and random.

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 (Prabhakaran, 2017). 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. 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. In contrast to ADF-test, for the KPSS-test the null hypothesis assumes stationarity. 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 (Hyndman, Athanasopoulos, 2018).

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 (Venkat, 2018).

To find the optimal values for p and q, the Autocorrelation function (ACF) and partial Autocorrelation function (pACF) were evaluated. The ACF plot shows the correlations of a time-series with lags of itself. The pACF calculates the relationship between a time series and its lag, excluding the influence of linear dependencies among other lags (Prabhakaran, 2017).

A second evaluation tool is the auto.arima function. The function automatically fits the best ARIMA model by minimizing the Akaike’s Information Criterion (AIC), which is a criterion for the quality of the model. The auto.arima function can also consider seasonal models.

Optimal models will yield uncorrelated residuals with zero mean and constant variance. This can be evaluated by plotting the ACF of the residuals and by performing a portmontreau test, for example the Ljung-Box test (Hyndman, Athanasopoulos, 2018).

Quellen:

Selva Prabhakaran (2017), Time Series Analysis, r-statistics.co

Aishwarya Venkat (2018), Time Series Analysis for Epidemiological Data, <https://rstudio-pubs-static.s3.amazonaws.com/354672_b7cb732a2b61469390d6fc72621bc9c4.html>

Hyndman, R.J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 12.07.2023

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

2017 Selva Prabhakaran, Time Series Analysis, r-statistics.co

* Stationary
* acf pacf

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

* ADF und KPPS

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

Time Series Analysis for Epidemiological Data

Aishwarya Venkat

2018-02-01

AR und MA

Hyndman, R.J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 12.07.2023

für später:

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.