# RTP Exercise Sheet

## Series 1

### Exercise 1.1

Apply backward differencing on a timeseries created by the following model:

$$X_t \sim 2t + 0.5 + U_t \quad , \text{where} \quad U_t \sim U(-5, 5)$$

Plot the result.

R-hints:

```
# create timeseries object
t <- seq(1, 100, length = 100)
data <- 2 * t + 0.5 + runif(100, -1, 1)
ts <- ts(data)

# have a look at
`?`(diff)

# compare lengths of arrays with
length()
```

### Exercise 1.2

Remove the linear trend by applying backward differencing on timeseries created from the following models:

a) $X_t \sim 0.5t + 1 + U_t \quad \text{where } U_t \sim U(-1, 1)$

b) $X_t \sim 2t^2 + 3t - 1 + U_t \quad \text{where } U_t \sim U(-200, 200)$

Plot the results.

### Exercise 1.3

To test ideas and algorithms, R comes with built-in data sets. The data used in this exercise is called co2 and contains atmospheric concentrations of $CO_2$ in parts per million. In the R-hints below is shown how to load the data into R.

Use backward differencing on the co2 data to abolish the seasonality effect. Figure out what value for the lag is to choose for an optimum reduction of the seasonality? What happens if you choose other values for the lag?

R-hints:

```r
# load co2 data
data(co2)
plot(co2, main = "co2 data")
```

## Exercise 1.4

Once again have a look at the co2 data set. In this exercise you should try to decompose the series into trend, seasonality and random parts using a linear additive filter. For the seasonal part, the hints below should help you calculate the means over the same months in different years.

R-hints:

```r
# R-help:
`?`(filter)

# seasonal part:
trend_removed <- series - series.trend
month <- factor(rep(1:12, 39))
seasn.est <- tapply(trend_removed, month, mean, na.rm = TRUE)
```

## Exercise 1.5

In this exercise, you are dealing with daily rainfall data, which is available as **rainDay.txt**[1]. The data consists of the date (01.01.2000 until 31.12.2007) and the rainfall on that day in mm.

R-hints:

```r
# R-help:
read.table()
as.Date()
`?`(ts())
```

a) Read in the data rainDay.txt and tell R that the column DATE is a date.

b) Define your data (without the DATE-column) correctly as a time series of class **ts**.

c) Use the R-Functions weekdays(), months() and quarters() to create these factors. Combine them together with the rainfall data and the date into one dataframe.

[1] http://stat.ethz.ch/Teaching/Datasets/WBL/rainDay.txt

d) Should a log transformation be applied to the rainfall data? If yes, why is the log transformation sensible in this case? Plot the boxplots of the rainfall data with weekday, month and quarter as grouping variables.

e) Plot the entire time series, as well as the part from 2006 to 2007.

## Exercise 1.6

What is the expected period (time period of repetition) and the time step for the following timeseries:

a) Sunshine duration per month in Basel from 1990 to 2000.

b) Number of newborn babies in the city of Zurich per year from 2000 to 2011.

c) Number of reservations in a restaurant for every night during 4 weeks.

d) Water runoff of a river. The data has been collected every day for 4 years.

## Exercise 1.7

Using the data **hstart.dat**[2], we illustrate various methods for descriptive decomposition and elimination of trends. The data contains monthly data on the start of residential construction in the USA within the time frame of January 1966 to December 1974. The data have undergone some transformation unknown to us (perhaps an index over some baseline value has been calculated, or perhaps the data are to be read as $x \cdot 10^?$ construction permits).

R-hint:

```
hstart <- read.table("hstart.dat")
```

a) Make a time series plot. Is this a stationary time series? If not, what kind of non-stationarity is evident? Into which components might this time series be decomposed sensibly?

b) Decompose the time series in trend, seasonal component and remainder using the non-parametric STL method, and plot the results.

c) The special filter $Y_t = \frac{1}{24}(X_{t-6} + 2X_{t_5} + ... + 2X_t + ... + X_{t+6})$ can be used for computing a trend estimate. Plot this, the STL trend and the data in a single plot. What are the differences between the two methods?

d) Try to remove the trend and seasonal effects by computing differences. After removing seasonal effects, choose some linear trend elimination method and plot the outcome.

[2] http://stat.ethz.ch/Teaching/Datasets/WBL/hstart.dat

## Exercise 1.8

Simulate timeseries according to the following models:

a) Y1: $Y_t = E_t - 0.5 \cdot E_{t-1}$ , where $E_t \sim N(0,1)$ i.i.d. $E_0 = 0$

b) Y2: $Y_t = Y_{t-1} + E_t$ , where $E_t \sim N(0,1)$ i.i.d. $Y_0 = 0$

c) Y3: $Y_t = 0.5 \cdot Y_{t-1} + E_t$ , where $E_t \sim N(0,1)$ i.i.d. $Y_0 = 0$

d) Y4: $Y_t = Y_{t-1} \cdot E_t$ , where $E_t \sim U(0.95, 1.05)$ i.i.d. $Y_0 = 1$

Use a time series plot to decide whether or not these processes are stationary. If a process is not stationary, suggest a simple method to make it stationary, and try this method in R.

*Disclaimer: Parts of the exercises are adopted from 'Applied Time Series Analysis' course at ETHZ by Marcel Dettling.*