Series 2

1. We would like to illustrate various methods for descriptive decomposition and elimination of trends using the data `hstart`.

This data contains monthly data on the start of residential construction in the USA within the time frame of January 1966 to January 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). In our opinion, this makes these data a good didactic choice for illustrating the theory.

(Source: U. S. Bureau of the Census, Construction Reports.)

Importing the data (without `header=T`!) and preparing them:

```
> hstart <- read.table("http://stat.ethz.ch/Teaching/Datasets/WBL/hstart.dat")
> hstart <- ts(hstart[,1], start=1966, frequency=12)
```

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

R-Hint:
The decomposition is made using `H.stl <- stl(hstart, s.window="periodic")`
Note: The smoothing parameter for the seasonal effect is chosen by means of `s.window`. If `s.window="periodic"`, the seasonal effect is estimated by averaging.
The trend estimation parameter can be set using `t.window`. Unlike `s.window`, this argument does have a default value (cf. the help file). Perhaps you could try to vary this parameter as well. The documentation for R and the help files give more details.
Trend, seasonal component and remainder of the STL are stored in

```
H.stl$time.series[,"trend"]
H.stl$time.series[,"seasonal"]
H.stl$time.series[,"remainder"]
```

Have a look at the output of `str(H.stl)` for more details.

2. We have seen several methods for descriptive decomposition and elimination of trends. In this exercise we will use the approach of parametric modeling for the Mauna Loa atmospheric CO_2 concentration data. This is a time series with monthly records from January 1959 to December 1997. The data is included in the default packages in R and you can load it with the command `data(co2)`.

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 sensibly be decomposed?

b) Decompose the time series into the components specified in a) using the R-Function `decompose()`. Run an additive decomposition. Why would you choose an additive decomposition? What can you say about the trend? Is it linear? What is the range and the frequency of the seasonal component?

c) Obviously there is a strong seasonal component. We investigate two ways of fitting the seasonal components. Fit a GAM model for the trend and encode each month with a factor. Plot the observed values and the fitted model. Also plot the residuals.
R-Hint: Use the following code:

```
> library(mgcv)
> tr <- as.numeric(..)
> months <- ...
> fit <- gam(co2 ~ s(tr) + months)
```

d) Fit the trend again using a GAM but this time we fit the seasonal component with an oscillation of the form $\beta_1 \sin(2\pi t) + \beta_2 \cos(2\pi t)$. Plot the observed and fitted time series together. Compare the residuals to the residuals of the factor seasonal effects.
R-Hint:

```
> fit <- gam(co2 ~ s(tr) + sin(2 *pi*tr)+cos(2*pi*tr))
```

3. The performance of a machine that analyses the creatine content of human muscular tissue is to be investigated using 157 known samples. These samples are given to the machine one at a time for it to determine their creatin content.

The data are from an investigation into the correct functioning of automated analysis machines. You can find them in the data

`http://stat.ethz.ch/Teaching/Datasets/WBL/kreatin.dat` .

In this exercise, we will focus on one of the variables in this data, namely `gehalt` (content).

a) Which stochastic model should this series of data follow if the machine is working correctly?

b) Use the time series plot, the autocorrelations and the partial autocorrelations to determine whether or not these data fit the ideal model found in Part a).
R-Hints:
Converting the data frame (`d.creatine`) to a time series:

```
> t.creatine <- ts(d.creatine[, 2], start = 1, frequency = 1)
```

Plotting ACF and PACF:

```
> acf(t.creatine, plot = TRUE)
> acf(t.creatine, type = "partial", plot = TRUE)
> ## or
> pacf(t.creatine)
```

Exercise hour: Monday, March 06.