Series 6

1. There is a study on the development of beluga whales that focusses on the nursing behaviour of mother and calf. During a total of 160 time periods (each lasting 6 hours) subsequent to birth, the following variables were observed for "Hudson", a beluga calf. Zoologists use this data to ascertain the health of this young whale.

PERIOD	Index of the time period
BOUTS	Square root of the number of nursing bouts
LOCKONS	Square root of the number of lock-ons ("docking attempts")
DAYNIGHT	Day ($= 1$, 8am – 8pm) or night ($= 0$, 8pm – 8am) indicator
NURSING	Square root of the number of seconds spent successfully nursing during the period.

   A nursing bout is defined as a successful nursing episode where milk was obtained. We would like to model the nursing time by means of the other variables. Count variables have already undergone a square root transformation to stabilize their variance ("first-aid-transformation"). You will find the data in the file beluga.dat.

   Load the data in the usual way and create a time series matrix:
   > d.beluga <- read.table("http://stat.ethz.ch/Teaching/Datasets/WBL/beluga.dat",
   header=T)
   > d.beluga <- ts(d.beluga)

   a) Fit the model

   $$\text{NURSING} = \beta_0 + \beta_1 \text{ PERIOD} + \beta_2 \text{ BOUTS} + \beta_3 \text{ LOCKONS} + \beta_4 \text{ DAYNIGHT}$$

   using ordinary linear regression. Check the independence of the residuals. What conclusions can zoologists draw from this analysis?

   b) Due to the correlations involved, an AR model should be assumed for the residuals. Determine the order p of this model, and estimate the parameters $\alpha_1, \ldots, \alpha_p$.

   c) What transformation should you apply to obtain a linear model with independent errors? State it as a formula.

   d) (*) How would you perform this transformation (or these transformations) in R? Use the transformed time series to carry out another regression, and look at the correlation structure for the errors!

   R-hint: lag()

   e) Estimate the regression coefficients and the AR parameters using Maximum Likelihood.

   R-hints:
   > library(nlme)   *# Load the package containing the procedure gls()*
   > r.bel.gls <- gls(NURSING ∼ BOUTS + LOCKONS + DAYNIGHT + PERIOD,
   + data = d.beluga, correlation = corARMA(form= ∼ PERIOD,
   + p = r.burg$order, q = 0, fixed = FALSE), method = "ML")
   > summary(r.bel.gls)
   > d.resid <- ts(resid(r.bel.gls))
   > plot(d.resid)
   > acf(d.resid)
   > pacf(d.resid)

   To ensure convergence of the algorithm, known estimates of the AR parameters can be passed to corARMA() as starting values using the optional argument values. In this particular case, this does not change the outcome.(correlation = corARMA(..., value = r.burg$ar, ...))

f) Simplify the model where possible.

2. In this exercise we investigate the connection between advertising expenditure and sales. Our example here is that of the annual advertising expenditure (`ADVERT`, in $1000) for, and sales revenue (`SALES`, in $1000) from, a particular brand of vegetable stock (Lydia Pinkham's Vegetable Compound, 1907 - 1960). These data are contained in the file `advert.dat`. Load this file and create a time series matrix:

```
> d.advert <- read.table("http://stat.ethz.ch/Teaching/Datasets/WBL/advert.dat",
header=T)
> d.advert.ts <- ts(d.advert)
```

a) Make plots of both these time series, either simultaneously or below each other.

R-hints:
`plot(..., ylim=...)` and `lines()` or more elegantly:
`plot.ts(d.advert[...], plot.type="multiple"/"single", lty=c(1,2))`

b) Investigate the dependence of sales revenue (`SALES`) on advertising expenditure in the current year (`ADVERT`) as well as the last (`ADVERT1`) by means of a linear model. Investigate whether or not the residuals are independent. What are the consequences of these results?

R-hints:
Those observations containing missing values are omitted in regression using `lm()`, and the following warning is given:
`(2 observations deleted due to missingness)`.
Objects other than time series do not yield warnings for missingness.
```
> r.lm1 <- lm(SALES ~ ADVERT + ADVERT1, ...)
> summary()
> plot()
> acf()
> pacf()
```

c) Extend this model to include the previous year's sales revenue (Variable `SALES1`), and analyse the residuals of this extended model (using Tukey-Anscombe and normal plots as well).

d) Another way to build a model is to look at the changes in sales revenue compared to the changes in advertising expenditure.
What fitted model do you obtain in such a case? Look at the residuals, too.

R-hints:
```
> d.dsales <- d.advert$SALES - d.advert$SALES1
> d.dadvert <- d.advert$ADVERT - d.advert$ADVERT1
```

e) Compare the models obtained in c) and d).

Preliminary discussion: Monday, March 31.