Series 2

1. In this problem we look at a data set containing measurements of solar radiation hitting the earth. The data is stored in the file `solar.radiation.rda`, which you can download from `http://stat.ethz.ch/education/semesters/as2012/asr`. It contains 29 measurements between 1963 and 2003. Note that for some years there are no measurements and some values might be corrupted!

a) Load the data in R and create a scatter plot. Add various smoothers to the scatter plot (Running Mean & Gaussian Kernel Smoother with bandwidth 10; LOESS).

b) Did the radiation intensity decrease over the years?

c) Fit an ordinary least squares regression to the data and plot the regression line in the scatter plot. Look at the summary of the fit – what can you say now about the radiation intensity over the years?

2. Various data on cars and their fuel consumption is stored in the data set `my.mtcars.rda`. We will look at the connection between the engine power (variable `hp`) and fuel consumption (variable `l.100km`).

a) Load the data set in R and create a scatter plot of `hp` and `l.100km`. Perform a linear regression, plot the regression line in the scatter plot, and print the summary output.

b) What is the estimated error standard deviation?

c) Use your model to predict the fuel consumption if the engine power is 100. Conversely, using the same model what would be the predicted engine power if the fuel consumption is 15?

d) Some people use the rule of thumb: "10 additional horse powers yield half a litre more fuel consumption per 100km". Does the data support this rule?

e) Draw the 95% confidence interval for the regression line as well as the 95% prediction interval into the scatter plot.

f) Analyze the residuals of the fit to assess whether the model assumptions are satisfied. Is the model appropriate?

g) Perform a log transformation of both variables and repeat the regression analysis (fit and plot the model, perform residual diagnostics). How does it compare with the original model?

h) The linear regression model for the transformed variables is

$$\log(\texttt{l.100km}) = \beta_0 + \beta_1 \log(\texttt{hp}) + \epsilon$$

Express this as a relation between the original variables `l.100km` and `hp`.

i) Plot the model curve of the log-transformed model in the scatter plot of the original model.

3. In an experimental setup oceanic bacteria were exposed to x-ray in 15 six-minutes intervalls. These are the results:

No. of bac.	355	211	197	166	142	106	104	60	56	38	36	32	21	19	15
Intervall	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

a) What does the relation between number of surviving bacteria and time of exposure look like? Is a linear regression reasonable?
Hint: `plot`.

b) Try to transform the data such that a straight line fits better.
Hint: One theory assumes that within each intervall the relative amount of bacteria killed is equal.

c) Estimate the size of the starting population of bacteria based on a regression analysis. Estimate also the relative decrease within each intervall.
Hint: `lm, summary`, see the R-Tutorial.

4. The file `gas.dat` contains the gas consumption (in kWh) and the differences of temperature (in ℃) inside and outside of 15 houses which are heated with gas. The measures were collected over a long time span and then averaged.

a) Read in the data from the internet using
`read.table("http://stat.ethz.ch/Teaching/Datasets/gas.dat", header = TRUE)`.
Hint: Alternatively the data can be downloaded from the web using a browser and read in from the local drive using also `read.table()`. This could be necessary if you get an error reading it directly (e.g. caused by a stringent firewall).
Illustrate the data graphically. What does the relation look like?

b) Compute an ordinary linear regression `mod1` of the consumption versus the temperature difference. Compare the output when calling `mod1` and `summary(mod1)`.

c) Perform a diagnosis of the model. Does the residual analysis look satisfying?
Hints:
`plot(fitted(mod1),resid(mod1)), abline(h=0)`,
`plot(gas$temp,resid(mod1)), abline(h=0)` and
`qqnorm(resid(mod1))`, `qqline(resid(mod1))`.
Or `plot(mod1)`, which generates directly the above plots and an additional one.
If necessary, try to find an alternative model.

d) What kind of consumption do you expect when the difference in temperature is 14℃ ?
Give also the confidence intervall for the expected consumption. Hint: `predict()`.

Preliminary discussion: Monday, October 01.

Deadline: Wednesday, October 10.