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? Think carefully and explain your reasoning.
- **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? **Hint:** The necessary theory for this part of the exercise will be discussed in week 4 so you may consult the scriptum.

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`). **Hint:** The necessary theory for **c)** to **f)** will be discussed in week 4 so you may consult the scriptum.

- **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)** By how much does the fuel consumption increase, if the engine power increases by 10`hp`?
- **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? **Hint:** Find the confidence interval for the slope.
- **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?

3. Which of the following statements are false and why?

- **a)** In presence of outliers, in order to visualize the relationship between a response variable and a predictor variable it is better to use a Gaussian kernel smoother, rather than a LOESS smoother.
- **b)** The following is an example of a linear regression model:
  $y = \beta_0 + \beta_1 x_1 + \beta_2 (x_1)^2 + \beta_3 log(x_2) + \beta_4 x_1 x_2 + E$
- **c)** The following is an example of a simple linear regression model:
  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + E$
- **d)** Smoothing methods give a more accurate extrapolation prediction, as compared to linear modeling methods.
- **e)** If the coefficient of determination of a linear regression model with predictor variable x and response variable y is equal to or greater than 0.8, we can claim x is a cause of y.

Preliminary discussion: Monday, September 29.

Deadline: Monday, October 6.