Data Analysis with R, fall 2010

Demonstrations 2

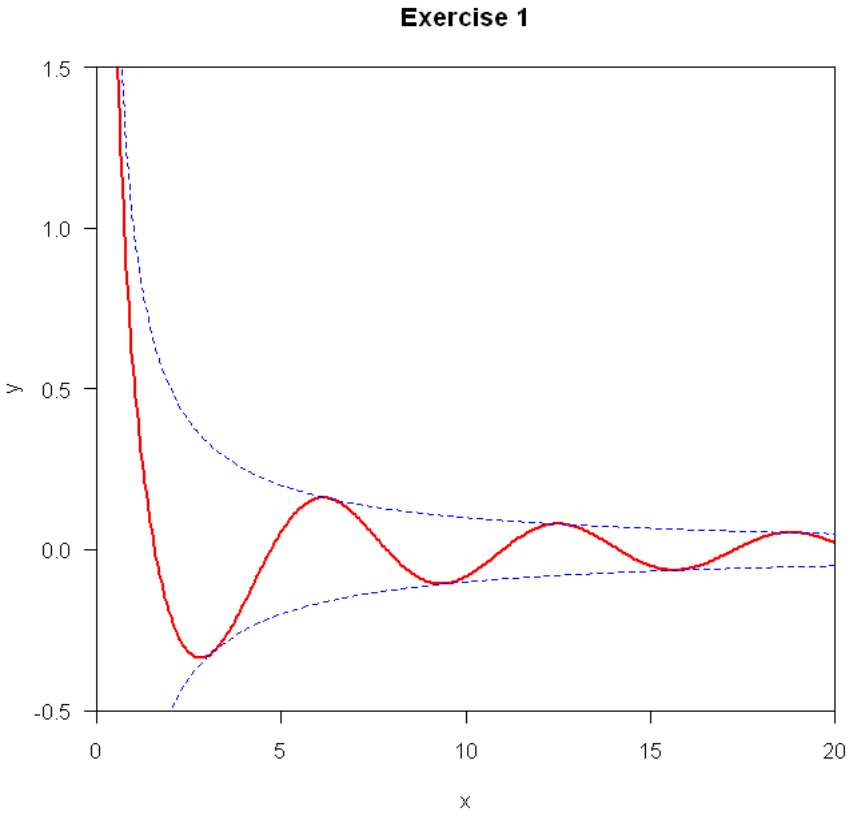
1. First represent the date 1 Jan 2010 as an object of class Date. Then create a vector of dates representing the first day of each month in the year 2010 using a suitable function call. Finally find out on which weekdays these dates fall.

2. The following code supposedly simulates n = 1000 values from the chi-squared distribution with nu = 4 degrees of freedom.

   ```
   n <- 1000; nu <- 4
   mat <- matrix(rnorm(n * nu), nrow = n)
   x <- apply(mat^2, 1, sum)
   ```

   To check the claim, plot the probability density histogram of the simulated values and the graph of the chi-squared density in the same figure. Draw the histogram using more bins than what hist() gives you by default. Specify the axis limits so that both the histogram and the density function fit nicely in the plotting area, and give the plot a meaningful title and axis labels.

3. Write R code for producing the figure below. The function cos(x)/x has been plotted on (0, 20) using a wide solid red line. The figure also contains the graphs of the functions 1/x and -1/x drawn with thin blue dashed. Use necessary arguments to produce **EXACTLY** the same kind of plot shown below.



Data Analysis with R, fall 2010

4. Let's consider a data frame called Pima.tr2 (*MASS* package). The eventual interest is in using variables in the first seven columns to classify diabetes according to *type*. Here, we explore the individual columns of the data frame.
   (a) Several columns have missing values. Determine the number of missing values in each column, broken down by *type.*
   (b) Create a version of the data frame Pima.tr2 that has *anymiss* as an additional column. *anymiss* gets the value of "nomiss", if all the values are present in a certain row and the value "miss" if one or more values are missing in a certain row (function called *complete.cases* might help…)
   For the first seven columns, compare the means for the two levels of *anymiss*, separately for levels of type. (So 4 different means per column)

5. Draw four random samples. One from Normal distribution, one from Chi-square distribution, one from Gamma-distribution and one from Cauchy-distribution. Plot the samples so that all of them are in one graph sheet. Use different point characters (symbols) and colours in each of the plot. (You can use your own selection of parameters when drawing the samples)

6. Input the data from *bostonc.txt*. The data is available at http://www.maths.anu.edu.au/~johnm/datasets/text/bostonc.txt Examine the contents of the initial lines of the file before trying to read it in. It will be necessary to change some options of *read.table* to get this to work. Going through the help-page of *read.table* will be helpful.