Introduction to R programming: Homework 4

Due on: March 10, 2017 11:59PM

Problem 1 (35 + 3 points)

Submit the R package you built during the lab. I expect to receive a single file called `coolpkg_1.0.tar.gz` or `coolpkg_1.0.zip` (version number might vary on different systems). The package must:

1. Be installable and loadable in R. (15 points)
2. Contain a function called `myRowSums` that works as intended. (10 points)
3. Include a function help file for the `myRowSums` function, with title, description, usage, value and an example. (10 points)
4. **Bonus 3 points**: Includes a package vignette written in Sweave/knitr.

Problem 2 (30 + 3 points)

Write a brief report addressing the following research question: **What is the cross-sectional relationship between age and physical activity in American children?** Use the `NHANES` dataset from the statistical analysis lecture. Use the variable **cpm** as a measure of physical activity, and include children age 6 to 17 years in your analysis. Here are some more detailed instructions:

1. Write an Introduction section providing a short description of the data. No more than 5 sentences. (10 points)
2. Create a boxplot showing physical activity vs. age (10 points).
3. Fit and interpret a linear regression model for physical activity vs. age. Limit your interpretation to no more than 5 sentences. (10 points)
4. **Bonus 3 points**: Write the report in Sweave/knitr.

Problem 3 (20 points)

Provide simulation codes to assess the coverage rate of a 95% confidence interval, under the same linear regression setting as in the simulation lecture.

Problem 4 (15 points)

Buffon's needle problem is a question posed in the 18th century by **Georges-Louis Leclerc, Comte de Buffon**: "*Given a needle of length a and an infinite grid of parallel lines with common distance d between them, what is the probability that a needle, tossed at the grid randomly, will cross one of the parallel lines?*".

The problem is solved. When $a < d$, the answer is $p = 2a/d\pi$. This result can be used for estimating π: $\pi = 2a/dp$. The probability p can be obtained from experiment: you keep throwing the needle and count the percentage of times it crosses the line. The experiment, however, can be realized computationally by a simulation.

Implement Buffon's needle idea and present the estimate of π (mean and standard deviation over many trials) using sample sizes 1000, 10000 and 100000. Submit the simulation code, and a short description of the method.