SS2864B, 2020

Assignment #3 due to March 6, 11:55pm, 2020

Instructions Submit an electronic version (pdf, words, etc) of your solutions (appropriately annotated with comments, plots, and explanations) to owl. Save all your R codes in one script file with proper comments and submit it as well to owl.

1. Use the plot function **persp** to plot the function (2+sin(x))(cos(2y)) with x and y both ranging from $-\pi$ to π. Experiment with using three different values of the **theta** and **phi** argument to persp, which give angles defining the viewing direction. **theta** gives the azimuthal direction and **phi** the colatitude. Please add proper sub or main titles.

2. Write an R function to calculate both

   (a) $\sum_{i=1}^{n} \min(2^i, i^3)$,

   (b) $\sum_{i=1}^{n} \max(2^i, i^3)$,

   i.e., both sums need to be calculated in one function and should return two values. Make sure to have two error checkings on input n and test your function for a few bad/wrong n's. Then execute your function for $n = seq(200, 5000, by = 600)$.
   Note: **for**, **while**, or **repeat** looping is not allowed. Please check the usages of R functions **pmin** and **pmax**. You can use R function **sapply** to carry out the last step calculation.

3. Implement an R function, say **IQR.outliers**, to compute the inter-quartile-range (IQR), find outliers if they exist, and produce a boxplot. The function should have one argument x. In the function body,

   (a) Do at least two error checkings on x before any computation.

   (b) Compute the IQR as Q_3 (3rd quartile) - Q_1 (1st quartile).

   (c) Use 1.5*IQR rule to detect suspected outlier(s) if they exist. First check if there is(are) value(s) of x that is(are) less than Q_1-1.5*IQR. If so, this (those) is(are) suspected outlier(s) on left tail. Similarly, checks any value(s) of x over Q_3+1.5*IQR on right tail.

   (d) Please produce a boxplot of x (a side-effect).

   (e) Please choose a proper output object to represent required information. The return values must contain IQR and outlier(s) from left and/or right tail.

   Test your function with the variables **dist** and **speed** in the data.frame **cars**. Also test your function with wrong inputs (checking if testing procedures are working or not).

4. NASA's GISS Surface Temperature Analysis (GISTEMP) is an estimate of global surface temperature change recorded monthly. The detail information can be found in https://data.giss.nasa.gov/gistemp. A file, GLB.Ts_dSST.csv, has uploaded to owl in Data sets folder. It contains monthly temperatures from 1880 to 2019. Please do the following steps to extract some basic information. Any looping such as for, while, repeat is not allowed.

(a) Import the dataset into R as a data frame. Create another data frame to keep only **Years**, **Jan**, ..., **Dec** 13 columns.

(b) Write an R function with input x to calculate the mean of x without the first element $x[1]$. Do one error checking on x. Then with the help of R function **apply** to generate a vector of yearly average temperatures from 1880 to 2019 and plot it as a time series. Comment your findings.

(c) Plot (time series plot) monthly temperatures from Jan to Dec for the year 1880 with proper labels and sub/main title. Then add monthly temperatures for every two decades (1990,1920,.., 2000) and 2019 with different colors and line types. Try to find an automatic way to add lines rather than adding one line a time. Finally adding a legend to the plot and comment your findings.

5. For iid sample $X_1, \ldots, X_n$ from population distribution $F(x)$, the empirical cumulative distribution function is defined as

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq y),$$

where $I(X_i \leq y)$ is an indicator function, i.e., $I(X_i \leq y) = 1$ if $X_i \leq y$ is true and $= 0$ if false. Write an R function called **my.ecdf** with arguments y and x (x is the vector $x_1, \ldots, x_n$) and return the value $F_n(y)$. Your function should check if y is a single number (a vector with only one element) or not. Otherwise, an error message should be returned. The **for** loop is not allowed. Test your function with one sample x=rnorm(20) and y=-2, median(x), and 2. Report your findings.