MCA 103 - Assignment 1 **MCA First Year 2017**

Date of Issue 1st September 2017 **Due Date 12th September 2017**

General Instructions

- Programming language to be used is R for every aspect of calculation. No manual calculation should be done
- Do it in groups of one/two but not in more than two
- **Please give all numerical answers to 10 digits of precision. Partial credit will be given to answers that agree to less than 10 digits**
- CR will collect all the folders of all the groups in a pen-drive and submit it to me for evaluation maximum by 4pm.
- You may refer to internet for solutions and executing various commands. You may also refer to e-books for syntax of specific commands.

Programming Instructions

- Note that all your programs should have proper alignment, indentation and proper comments.
- All constants / variables / functions etc. should have meaningful names.
- Overall, programs should be readable. If programs fails to execute in R, you will get zero for everything.
- Submission files – Q1.R, Q2.R, Q1_readme.txt, Q2_readme.txt
- Read me files should give information about code, functions and data structures used, diagrammatic representation of the concepts, etc. You may refer to preparation of readme file from here.

Q1: A Tower of Hanoi setup consists of N disks, weighing $1,\ldots, N$ kilos. Each disk sits in one of M stacks, at positions $0,\ldots, M-1$. All disks start stacked at position zero, ordered by weight with the lightest on top. At each point in time, a valid move can be made. A valid move consists of taking the top disk off of one of the stacks and moving it to the top of the stack to the immediate left or right (that is, at position one less or one greater than its current position), provided that the

moved disk weighs less than all other disks at its destination stack. For example, in a N=M=3 problem with disks 1,2,3 at positions 0,1,2, respectively, the only valid moves are for disk 1 to move to position 1 and disk 2 to move to position 2. We choose moves uniformly at random amongst all valid moves at a given time. I am interested in the position of the center of mass after T moves, that is

$$\sum d \times p_d \; / \; \sum d$$

Where, d is the disk weight and p_d is the position of a disk weighing d.

(a) For M=3, N=3, what is the mean of the center of mass after T=16?

(b) For M=3, N=3, what is the standard deviation of the center of mass after T=16?

(c) For M=6, N=6, what is the mean of the center of mass after T=256?

(d) For M=6, N=6, what is the standard deviation of the center of mass after T=256?

Q2. Lending Club is a platform for non-traditional loans. Applicants submit an application for a loan, and individual investors can choose to fund part of the loan. To help investors make intelligent decisions, Lending Club has released data on past loan performance. For these questions, I will be looking at the lending club loans issued in 2014 and 2015. Download the relevant CSV files, as well as a spreadsheet describing the columns (i.e. Data Dictionary).

a) What is the median loan amount?

b) Each loan is categorized into a single purpose. What fraction of all loans are for the most common purpose?

c) Calculate the average interest rate across loans for each purpose. What is the ratio of minimum average rate to the maximum average rate? (The ratio should be less than 1.)

d) What is the difference in the fraction of the loans with a 36-month term between 2014 and 2015?

e) I will consider all loans that are not in the 'Fully Paid', 'Current', 'In Grace Period' statuses to be in default. Calculate the ratio of the time spent paying the loan, defined as the

difference between the last payment date and the issue date, divided by the term of loan. What is the standard deviation of this ratio for all the loans in default?

f) What is the Mean, Median, Mean Absolute Deviation, Variance, IQR, Skewness and Kurtosis for the total rate of return, as figured from the total payments and the loan amount, and the interest rate? Consider only loans that have reached the end of their term. [*Summary function NOT to be used here*]

g) Let's find a loan purpose that shows up abnormally often in one state. Call *A* the probability of a loan going to a specific purpose nationwide. Call *B* the probability of a loan going to a specific purpose for each state. Out of all (state, purpose) pairs with at least 10 loans, what is the highest ratio of B / A (*i.e.* the most surprising)?

h) Group all loans by their sub-grade and calculate their average interest rate, average default rate, and percentage of loan status categories.

All the Best