CKME 132 Summer 2016 - Assignment #3

Tamer Abdou

July 26, 2016

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

Use RStudio for this assignment. Edit the file `assignment-2.Rmd` and insert your R code where wherever you see the string "INSERT YOUR ANSWER HERE"

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.

This assignment questions 1 - 4 make use of data that is provided by the `ISwR` package.

```
library(ISwR)
```

```
## Warning: package 'ISwR' was built under R version 3.2.5
```

Sample Question and Solution

Use `seq()` to create the vector $(1, 2, 3, \ldots, 10)$.

```
seq(1,10)
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

Question 1 - 35%

In this question you will explore the airquality dataset available in the datasets library in the ISwR package.

a) Display the first 6 rows of the airquality dataset.

```
#Insert your code here.
```

b) Display the type of each column of the airquality dataset, use only one function in R to do so.

```
#Insert your code here.
```

c) Use a histogram to assess the normality of the `Ozone` variable.(In order to get the output diagram inserted in your answer use attach(dataframe name))

```
#Insert your code here.
```

d) Does it appear normally distributed?

```
#Insert your code here. (Do not remove the #)
#
```

e) Create a boxplot which shows the distribution of `Ozone` in each month.Use different colors for each month.

```
#Insert your code here.
```

f) Create one scatter plot matrix of the numeric variable(Ozone, Solar.R,Wind,Temp) within the airquality dataset. (Hint investigate pairs())

```
#Insert your code here
```

Question 2 - 10%

a)Use simulation to estimate the mean and variance of a binomial random variable with n = 18 and p = 0.76.

```
#Insert your code here
```

b) Calculate the values using the theroy (state the value and the equation in your answer),compare the values you get with the values you got in (a), wirte one sentence sumurizing the comparision.

```
#Insert your answer here (Do not remove the #)
```

Question 3 - 10%

a)Estimate the mean and variance of a Poisson random variable whose mean is 7.2 y simulating 10,000 Poisson random numbers.

```
#Insert your code here
```

b)Compare the mean value you got in (a),with the one stated in the question. wirte one sentence summarizing the comparision.

```
#Insert your answer here (Do not remove the #)
#
```

Question 4 - 10%

Simulate 100 realizations of a normal random varialbe having a mean of 51 and a standard deviation of 5.2.

```
#Insert your code here
```

Question 5 - 35%

This question makes use of pakcage “RCurl”, accordingly carry out the following:

```
#install.packages("RCurl")
library(RCurl)
```

```
## Loading required package: bitops
```

First we read the computers.csv file and load the price using the following:

```
u <- getURL("http://vincentarelbundock.github.io/Rdatasets/csv/Ecdat/Computers.csv")
cprices <- read.csv(text =u)
```

a) Display the first 6 rows of cprices and make note of all the variables.

```
#Insert your code here
```

b) Calculate the mean,variance and standard diviation of price by omitting the missing values, if any.

```
#Insert your code here
```

c) Use `ram` to predict price and build a univariate linear regression model, display a summary of your model indicating Residuals, Coefficients..etc.

```
#Insert your code here
```

d)Based on the output of your model, predicted the expected price when ram is set to 8 MB

```
#Insert your answer here
```

e)Find Pearson correlation between hard disk and speed.

```
#Insert your code here
```

f)Write the correlation matrix of the variables:price,speed,hd and ram.

```
#Insert your code here
```

END of Assignment #3.

Bonus Question

π appears in the formula for the standard normal distribution, the most important probability distribution in statistics. Why not give it a try to calculate π using statistics! In fact, you'll use a simulation technique called the *Monte Carlo Method*.

Recall that the area of a circle of radius r is $A = \pi r^2$. Therefore the area of a circle of radius 1, aka a *unit circle*, is π. You'll compute an approximation to the area of this circle using the Monte Carlo Method.

a) The Monte Carlo Method uses random numbers to simulate some process. Here the process is throwing darts at a square. Assume the darts are uniformly distributed over the square. Imagine a unit circle enclosed by a square whose sides are of length 2. Set an R variable `area.square` to be the area of a square whose sides are of length 2.

b) The points of the square can be given x-y coordinates. Let both x and y range from -1 to +1 so that the square is centred on the origin of the coordinate system. Throw some darts at the square by generating random numeric vectors x and y, each of length `N = 10,000`. Set R variables `x` and `y` each to be uniformly distributed random numbers in the range -1 to +1. (hint: runif() generates random number for the uniform distribution)

c) Now count how many darts landed inside the unit circle. Recall that a point is inside the unit circle when $x^2 + y^2 < 1$. Save the result of sucessfull hits in a variable named hit. (hint: a for loop over the length of x and y is one option to reach hit)

d) The probability that a dart hits inside the circle is proportional to the ratio of the area of the circle to the area of the square. Use this fact to calculate an approximation to π and print the result

Wow you got the first estimate for `pi` π, congratulations you have completed the first run of the Monte Carlo simulation.

Bonus question end.

If there is further intrest put all the above logic in a function, and call it 50 times store the results in a vector called pi then take the mean of pi. vector.