作業加分題(選擇性繳交,每週持續更新至期末)

- 注意：
  - 作業加分題必須以附檔繳交，若未以附檔繳交不會被批改。
  - 所有作業加分題必須合併在一個文字附檔(可接受檔案類型：以記事本可開啓的.txt檔、以R可開啓的.R檔或以Word可開啓的.docx)繳交，若繳交多個附檔，只有第一個附檔會被批改。
  - 繳交的檔案需以註解方式標註每一題題號(在題號行開頭加上#)，所有非程式碼的部份也需以註解方式處理(在每一行開頭加上#)。
  - 繳交的檔案以記事本（或R、Word)開啓後，選取全部內容貼到R執行必須能跑出所有答案。

1. (15 pts) The practice problem in the handout "Empirical CDF and the Kolmogorov—Smirnov statistic". The file is located at

   `https://stat.walkup.tw/teaching/programming_R/handouts/ks.pdf`

2. (10 pts) Run the following R commands to create two files `f1.txt` and `ff.txt` in the R working directory:

   ```
   cat("f1\n", file="f1.txt")
   cat("ff\n", file="ff.txt")
   ```

   Write down R commands for completing the following tasks.

   (a) List all files with file names beginning with `f` and ending with `.txt` in the R working directory. List the file names only.

   (b) Remove all files listed in Part (a) from the R working directory.

3. (10 pts) Run the following R commands to create a file `test.txt` and read its content into an R string vector `x`:

   ```
   cat("Using lasso or elastic net for variable selection\n", file="test.txt")
   cat("lasso variable selection requires\n",  file="test.txt", append=TRUE)
   cat("setting penalty parameters\n", file="test.txt", append=TRUE)
   x <- scan(file="test.txt", what="char")
   ```

   Write down R commands for completing the following tasks.

   (a) Compute the number of components of `x` that contain the string "`las`".

   (b) Create a vector `y` obtained by replacing each "`las`" with "`LAS`" in the components of `x`.

4. (15 pts) A student stored his R programming homework in a file `hk.txt` and emailed the file to the TA. The content in `hk.txt` looks like the following but is much longer:

   ```
    test <- function(A, B){
   +   m <- dim(A)[1]
   +   n <- dim(B)[2]
   +   if (dim(B)[1]!=dim(A)[2]) return("Error!")
   +   D <- matrix(0, m, n) ##
   ```

```
+   for (i in 1:m){
+     for (j in 1:n){ D[i,j] <- sum(A[i,]*B[,j]) }
+   }
+   D <- D+D
+   return(D)
+ }
```

The TA asked the student to remove the extra "+" appearing in the beginning of each line in `hk.txt` and then turn in the modified file by email. Suppose that the student's file `hk.txt` is copied to your R working directory. Write down the R commands for removing the extra "+" appearing in the beginning of each line in `hk.txt`. 注意：在R正規表示式中，符號"+"爲特殊字元，若要作爲一般字元使用時，在表示式中要使用"[+]"來表示。

5. (10 pts) Write down R commands for completing the following tasks:
   (a) generating a sample of 1000 IID data from the distribution of a random variable X, where the possible values for X are 0, 1, 2 and $P(X = x) = (x + 1)/6$ for $x = 0, 1, 2$
   (b) finding the proportions of 0, 1, 2 respectively in the sample generated in Part (a)

6. (15 pts) Write down the R commands for computing the integral

$$\int_{\{(x,y,z):x^2+y^2+z^2\le 1, x\ge 0, y\ge 0, z\ge 0\}} e^{x^2+y^2+z^2} d(x, y, z)$$

   using Monte Carlo method. Take the Monte Carlo sample size to be 10^6.

7. (30 pts; 15 pts for #1 and 15 pts for #2) 做決策樹報告檔案`tree.pdf`最後一頁的題目。決策樹報告檔案連結如下：

   `https://stat.walkup.tw/teaching/programming_R/hk_files/tree.pdf`

   資料檔`babies.txt` 連結如下：

   `https://stat.walkup.tw/teaching/programming_R/data/babies.txt`

8. (30 pts; 7 pts for each part of #1 and 8 pts for each part of #2) 做二項分布報告檔案 `binom_1205.pdf`最後的作業題目。二項分布報告檔案連結如下：

   `https://stat.walkup.tw/teaching/programming_R/hk_files/binom_1205.pdf`

9. (40 pts; 10 pts for #1, 20 pts for #2(a) and 10 pts for #2(b)) Do the practice problems in the file `gradient.pdf`. The file is located at

   `https://stat.walkup.tw/teaching/programming_R/handouts/gradient.pdf`

10. (40 pts; 10 pts for #1, 15 pts for #2(a) and 15 pts for #2(b)) Do the practice problems in the file `mle.pdf`. The file is located at

    `https://stat.walkup.tw/teaching/programming_R/handouts/mle.pdf`

11. (40 pts; 20 pts for #1, 20 pts for #2) Do the practice problems in the file `km.pdf`. The file is located at

    `https://stat.walkup.tw/teaching/programming_R/handouts/km.pdf`