作業繳交說明

- 作業均以附檔上傳方式繳交.
- 作業檔名要包含學號.
- 作答檔案必須是類型爲txt的純文字檔案, 可以在Windows系統下用記事本開啓, 開啓後內容全選複製, 再貼上到 R 指令列執行後, 要能跑出所有答案. 例如, 如果要作答下面的二個問題:
  - 問題100. 寫出R指令計算以下結果
    (a) $\sqrt{3}$.
    (b) $2^{1.2/\pi}$.
  - 問題101. 寫出R指令計算$1+2+\cdots+10$.

  那麼解答要寫成

```
#100
#(a)
sqrt(3)

#(b)
2^(1.2/pi)

#101
sum(1:10)
```

  這樣以上九行(含二行空白行)全選複製後, 貼上到 R 指令列執行, 就可以跑出全部答案(三個數值). 注意題號部份都放在標注符號#之後, 因此並不會被作爲指令執行.

作業問題從下一頁開始

Problems

1. (20 pts) Evaluate the following expressions.
   (a) $2^{2.1}$.
   (b) $\sqrt{2}$.
   (c) $4\tan^{-1}(1)$. Here $\tan^{-1}$ is the inverse function of tan.
   (d) ae^{-ax} for $a = 2.1$ and $x = 3.2$. Here $e = \lim_{n\to\infty}(1 + 1/n)^n$.

2. (30 pts) Suppose that $x = (1, 3, 5, \ldots, 19)$. Let x_i denote the i-th element in x for $i = 1, \ldots, 10$. Evaluate the following expressions.
   (a) $\sum_{i=1}^{10} x_i$.
   (b) $\prod_{i=1}^{10} x_i = x_1 x_2 \cdots x_{10}$.
   (c) $\sum_{i=1}^{5} \frac{1}{x_{2i}}$.
   (d) $\sum_{x_i<5} x_i$.
   (e) $\sum_{x_i<5 \text{ or } x_i>15} x_i$.
   (f) $\sum_{\tan(x_i)<3 \text{ and } x_i>15} x_i$.

3. (15 pts) Suppose that $x = (1, 3, 5, \ldots, 19)$ and $y = (2, 4, \ldots, 20)$. Let x_i and y_i denote the i-th element in x and y respectively for $i = 1, \ldots, 10$. Evaluate the following expressions.
   (a) $\prod_{i=1}^{10}(x_i + y_i)$.
   (b) $\sum_{i=1}^{10} \frac{x_i}{y_i}$.
   (c) $\sum_{i=1}^{10} x_i^3 y_i^2$.

4. (10 pts) Suppose that X is a discrete random variable such that $P(X = k) = 1/20$ for $k = 1, \ldots, 20$. Compute $E(X)$ and $E(X^2)$. Recall that for a discrete random variable X, $E(X)$ is the weighted average of all possible values of X, where the weight for a possible value x is $P(X = x)$.

5. (8 pts)
   (a) Write a function `even.x`. The function input is a vector `x` and the output is the vector `x[c(2,4,6,...)]`, the components of `x` with even indexes. You may assume that the length of `x` is an even number.
   (b) Use the function `even.x` to compute $x_2 + x_4 + \cdots + x_{200}$ for $x = (1, 3, 5, \ldots, 399)$. Write down the R commands.

6. (8 pts)

(a) Write a function `mean.fun`. The function has two input variables `n` and `x`, where `n` is a positive integer and `x` is a vector of numbers. The function output is the average of values in `x` excluding the largest `n` values and the smallest `n` values.

Note.

- You may assume that the length of `x` is larger than `2n`.
- The input vector `x` may not be sorted. The R command to sort `x` is `sort(x)`.

(b) Use the function `mean.fun` to compute the average of numbers in $x = (1, 3, 5, \ldots, 399)$ excluding the smallest two numbers and the largest two numbers in x. Write down the R commands.

7. (8 pts)

(a) Write a function `prop.fun`. The function has two input variables k and `x`, where $k > 0$ and `x` is a vector of numbers. Let m denote the sample mean of `x` and let s denote the sample standard deviation of `x`. The function returns a list of two numbers p and q, where p is the proportion of numbers in `x` that are outside the interval $[m - ks, m + ks]$, and $q = 1/k^2$.

(b) Use the function `prop.fun` to compute the proportion of the numbers in x that are outside the interval $[m - 2s, m + 2s]$ for $x = (1, 3, 5, \ldots, 399)$, where m and s are the sample mean and sample standard deviation of the numbers in x respectively. Write down the R commands.

8. (5 pts) Let

$$f(x) = \frac{2\sqrt{1+x} - x - 2}{x^2}$$

for $x > 0$. Write down the R commands for plotting f on $[0.001, 1]$.

9. (5 pts) Write a function `sq.fun`. The function does not have any input variable. When running the function, the user will be asked to enter a number and then the following output will appear on the screen:

```
The square of x is xx
```

where `x` is the number entered by the user and `xx` is the square of `x`.

10. (6 pts) Write a function `median.fun`. The function input is a vector `x` and the output is the median of `x`. You may use the command `sort(x)` to sort `x` and the command `floor(x/2)==(x/2)` to check wether `x` is an even number. The command `floor(x)` gives the largest integer that is less than or equal to `x`. Do not use the function `median` in your function.

11. (6 pts) Write a function `cor.fun`. The function has two input variables `x` and `y`, and the function output is the sample correlation between `x` and `y` unless any of the following happens.

(a) `x` is not numeric (`is.numeric(x)` is `FALSE`).

(b) `y` is not numeric (`is.numeric(y)` is `FALSE`).

(c) `x` and `y` do not have the same length.

(d) The sample variance of `x` is 0.

(e) The sample variance of `y` is 0.

If (a)–(e) happens, the function should return the following string:

```
"The sample correlation cannot be computed!"
```

Write down the R commands for computing the correlation between `x` and `y` for the following `x` and `y` using your function `cor.fun`:

- `x <- c("a","b"); y <- c(1,3)`
- `x <- 1:4; y <- c(1,3)`
- `x <- rep(0,4); y <- 1:4`
- `x <- seq(1.1, 3.3, by=1.1); y <- c(1,2,4)`

You may use the R functions `mean` and `var` in your function. You may use the R function `cor` to check your answer, but do not use the function `cor` in your function. Recall that the sample correlation between $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ is given by

$$\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}},$$

where $\bar{x} = \sum_{i=1}^n x_i/n$ and $\bar{y} = \sum_{i=1}^n y_i/n$.

12. (6 pts) Let

$$f(x) = \begin{cases} \sin(x)/x & \text{if } |x| > 10^{-6}; \\ 1 & \text{otherwise.} \end{cases}$$

Write a function `f.fun` that returns $f(x)$ for an input value x. You may use the `abs` function in your function.

13. (30 pts) Download the data file

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

and save it to a file called `test.txt` in the working directory of R.

(a) Write down the R commands that read the data into an R object `X`.

(b) Write down the R command that gives the number of columns of `X`.

(c) Write down the R command(s) for computing the sample correlation between the first two columns of `X`. You may use your `cor.fun` in Problem 11 or the R function `cor`.

(d) Write down the R command that creates a matrix `Z` that is composed of the first 10 rows of `X`.

(e) Write down the R command that creates a matrix `W` that is composed of the last 10 rows of `X`.

(f) For `Z` and `W` in Parts (d) and (e), let `Y` be the matrix (`Z`, `W`). That is, `Y` is a $10 \times 2n$ matrix, where n is the number of columns of `X`, `Z` is composed of the first n columns of `Y`, and `W` is composed of the last n columns of `Y`. Write down the R command(s) that write the data in `Y` to a file `y.txt` located in the R working directory. In the file `y.txt`, the columns should be separated by space and column names and row names should not be included.

14. (20 pts) Run the following commands in R to create objects `x`, `y` and `z` in the current R workspace.

```
x <- matrix(1:8, 4, 2)
y <- t(x) %*% x
z <- 1:3
```

Write down the R commands for completing the following tasks.

(a) Save `x` and `y` to a R workspace file `xy.RData` in the R working directory.

(b) Remove all the objects in the current R workspace.

(c) Load `xy.RData` (located in the R working directory) into the current R workspace.

(d) List all the objects in the current R workspace.

15. (8 pts) For $x \in (-\infty, \infty)$, let

$$f(x) = \begin{cases} -1 & \text{if } x < 0; \\ 1 & \text{if } x > 0; \\ 0 & \text{if } x = 0. \end{cases}$$

(a) Define an R function `f.fun` that computes $f(x)$ for a given x. When the function input is a vector $(x_1, \ldots, x_n)$, the function output is the vector $(f(x_1), \ldots, f(x_n))$.

(b) Plot f on $[-2, 2]$ using the R function `curve` and your R function `f.fun` from Part (a).

16. (8 pts) Plot $z = x^2 - y^2$ for $x \in \{-0.9, -0.8, \ldots 0.9\}$, $y \in \{-2.2, -2.1, \ldots, 2.2\}$ using the R function `persp`.

17. (6 pts) Write down the R commands for defining a matrix `M` that is a 100×100 matrix of the form

$$\begin{pmatrix} 1 & 2 & 3 & 4 & \cdots & 100 \\ 2 & 1 & 2 & 3 & \cdots & 99 \\ 3 & 2 & 1 & 2 & \cdots & 98 \\ & & & \vdots & & \\ 100 & 99 & 98 & \cdots & 2 & 1 \end{pmatrix}$$

18. (6 pts) Let

$$s_n = 1^{1.1} + 2^{1.1} + \cdots + n^{1.1}$$

for $n \geq 1$. Let k be the largest integer n such that $s_n \leq 1000$. Write down the R commands for computing k and s_k and then printing these two numbers on the screen.

19. (6 pts) Write an R function `test.fun` that takes a square matrix `M` as input and returns a vector `x` whose j-th element is the sum of the first j elements in the j-th row of `M`. In addition, write down the R command(s) for finding `test.fun(M1)`, where `M1` is the matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 6 & 9 \end{pmatrix}.$$

20. (6 pts) Let $\{a_n\}_{n=1}^{\infty}$ be the sequence such that $a_1 = 1$, $a_2 = 1$ and

$$a_{n+2} = a_n + a_{n+1}$$

for $n \geq 1$. Let $b_n = a_{n+1}/a_n$ for $n \geq 1$. Let m be the smallest integer n such that $|b_n - b_{n-1}| < 0.001$. Write down R commands for finding b_m.

21. (6 pts)

    (a) Write down an R function `count.fun` with two input variables `x` and `brks`. Both `x` and `brks` are numeric vectors. Suppose that `brks` $= (b_1, \ldots, b_m)$ and $b_1 < b_2 < \cdots < b_m$. The output of `count.fun` is the vector $(n_1, \ldots, n_{m-1})$, where n_i is the number of values in `x` that are in the interval $[b_i, b_{i+1})$ for $i = 1, \ldots, m-1$.

    (b) Run the following R commands to obtain a vector `x`:

    ```
    set.seed(1)
    x <- runif(100, 0,2)
    ```

    Let I_1, I_2 and I_3 be the three intervals $[0, 1)$, $[1, 1.5)$, $[1.5, 2)$ respectively, and let n_j be the number of elements in `x` that are in the interval I_j for $j = 1, 2, 3$. Write down the R commands to compute the vector (n_1, n_2, n_3) using your function `count.fun` in Part (a).

22. (6 pts) Write down R commands for computing the following quantities:

    (a) $P(N(3, 4) > 2)$

    (b) the number z such that $P(N(0, 1.21) \leq z) = 0.75$

23. (6 pts) Suppose that X is a random variable whose distribution is Poisson with mean 2.5. Write down R commands for computing the following quantities:

    (a) $P(X \leq 3)$

    (b) the maximum of $a_0, \ldots, a_{100}$, where

    $$a_k = |P(X \leq k) - P(Y \leq k)|$$

    for $0 \leq k \leq 100$ and Y is a random variable whose distribution is $Bin(100, 2.5/100)$.

    Note.

    - If you do not know the name for Poisson distribution in R, you can use the `help.search` command to look it up.
    - $a_0, \ldots, a_{100}$ should be small since the Poisson distribution with mean μ can be approximated well by $Bin(n, \mu/n)$ for large n.

24. (6 pts) Consider the iris data set in R. You may enter

    ```
    iris
    ```

    in R to view the data set, and enter

    ```
    help("iris")
    ```

    in R to get a description for the data set. Write down R commands for completing the following tasks:

(a) drawing a matrix of scatter plots for the first 4 columns in `iris`.

(b) drawing a historgram for the first column in `iris` using break points 4, 4.5, 5, ..., 8.

25. (6 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)

26. (10 pts) Run the following R command

```
x <- c("ab.tt", "abtt", "att", "bt1")
```

    For each of the following parts, a string property is described. For each property, write down an R command for listing the component(s) of `x` satisfying the property.

    (a) beginning with "`ab`"

    (b) ending with "`.tt`"

    (c) beginning with "`a`" and ending with "`.tt`"

    (d) containing "`bt`"

    (e) beginning with "`a`" or "`b`"

27. (10 pts) Run the following R commands to create two files `f1.txt` and `ff.txt`:

```
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).

28. (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`.

29. (10 pts) A student stored his R programming homework in a file `hk.txt` and emailed the file to the TA. The first 11 lines in `hk.txt` are given below (the file 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 in your R working directory. Write down the R commands for removing the extra "`+`" appearing in the beginning of each line in `hk.txt`.