STAT2005 Introduction to Programming Languages for Statistics
Sample Midterm Examination Paper

Answer ALL questions.

Question 1 (27 marks)

(a) (7 marks) Write the R codes to create the following object named **mylist**.

```
> mylist
$a
[1] 1 2 3 4 5

$b
[1] "a" "b"

$c
     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6
```

(b) (10 marks) A survey was conducted from a series of software workshops. The information collected from the survey include

- Workshop – software introduced at the workshop
- Gender – gender of participant
- Q1 – The instructor was well prepared.
- Q2 – The instructor communicated well.
- Q3 – The course materials were helpful.
- Q4 – Overall, I found this workshop useful.

The data are stored in a data frame named **survey** as shown below.

```
> survey
  workshop gender q1 q2 q3 q4
1        R Female  4  3  4  5
2     SPSS   Male  3  4  3  4
3     <NA>   <NA>  3  2 NA  3
4     SPSS Female  5  4  5  3
5    STATA Female  4  4  3  4
6     SPSS Female  5  4  3  5
```

Write the R codes to create this data frame.

(c) (3 marks) Create a data frame consisting of only the first two columns of **survey**.

(d) (3 marks) Create a data frame consisting of only the first and last row of **survey**.

(e) (4 marks) Replace all "**Female**" by "**F**" and "**Male**" by "**M**" in **survey**.

Question 2 (19 marks)

(a) (5 marks) With the use of **sample()** function, write down a command to generate a sample from the distribution $f_X(x) = \Pr(X = x)$ given below.

x	$f_X(x)$
1	0.2
2	0.4
3	0.3
4	0.1

(b) (8 marks) Generate 2,000 random sample from $f_X(x)$ and save them as **r**. Transform **r** into a 1,000-by-2 matrix and save them again as **r**, such that each row in **r** represents a bivariate sample (x_1, x_2).

(c) (6 marks) Produce a two-way table showing the frequency count for each combination of (x_1, x_2) using the sample obtained in part (b). A sample output is shown below.

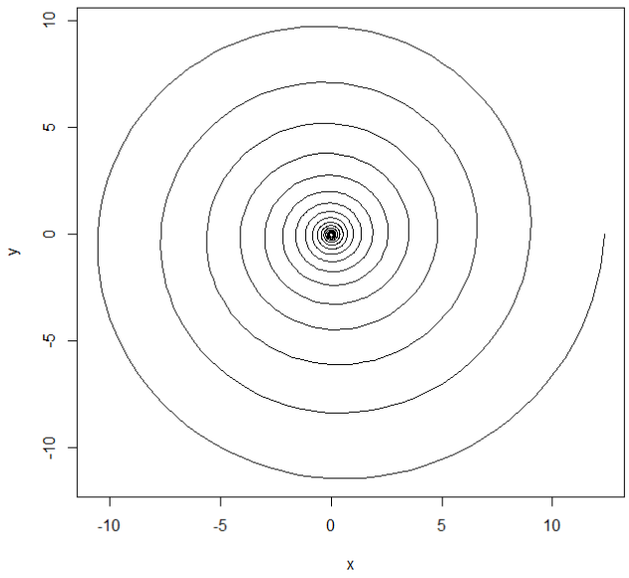
```
    1   2   3   4
1  45  84  58  26
2  76 155 114  40
3  53 124  86  32
4  22  45  29  11
```

Question 3 (18 marks)

A spiral can be described using the following equations.

$$\begin{cases} x = e^{0.05\theta} \cos\theta, \\ y = e^{0.05\theta} \sin\theta, \end{cases} \quad -16\pi \leq \theta \leq 16\pi.$$

Plot this spiral using R. A sample is shown below.



Hint: compute all the (x, y) coordinates along the given range of θ and then use **`plot()`**.

Question 4 (18 marks)

Each new term in the Fibonacci sequence is generated by adding the previous two terms. By starting with 1 and 1, the first 10 terms will be:

1, 1, 2, 3, 5, 8, 13, 21, 34, 55, ...

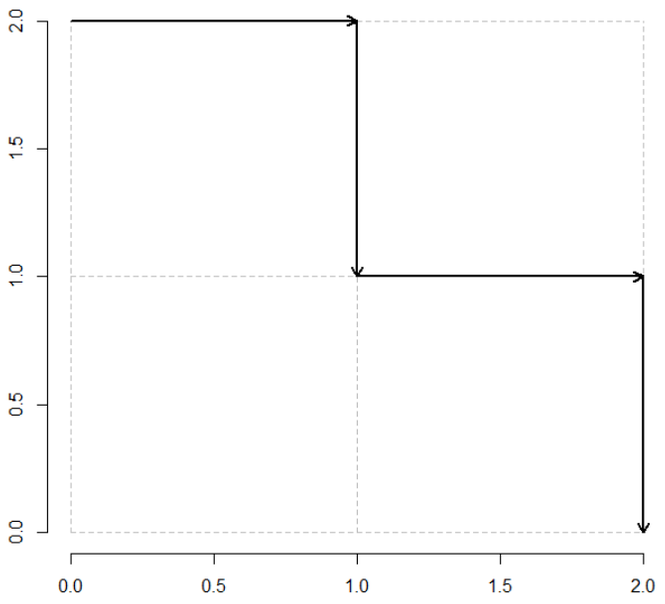
By considering the terms in the Fibonacci sequence whose values do not exceed 1,000,000,000, find the sum of the even-valued terms.

Question 5 (18 marks)

Use the following command to generate an empty plot.

```
plot(0, 0, type="n", xlim=c(0,2), ylim=c(0,2), bty="n",
xlab="", ylab="")
```

Use low level graphic functions to generate the following plot.



Note: the dashed lines are of line type 2, the arrows are of double line width and the arrow heads have length 0.1.

End of Questions

Q1(a)

```
mylist ← list ("a" = c(1,2,3,4,5), "b" = c("a", "b"),
               "c" = matrix(1:6, nrow=2))
```

(b)

```
survey ← data.frame(
    workshop = c("R", "SPSS", NA, "SPSS", "STATA", "SPSS")   ← can be factor or not
    gender = c("Female", "Male", NA, "Female", "Female", "Female")
    q1 = cc 4, 3, 3, 5, 4, 5)
    q2 = cc 3, 4, 2, 4, 4, 4)
    q3 = c(4, 3, NA, 5, 3, 3)
    q4 = c(5, 4, 3, 3, 4, 5)
)
```

(c)

```
first2 col ← survey[, c(1,2)]
```

(d)

```
first last ← survey[c(1, 6), ]
```

nrow(survey)

(e)

```
survey$gender ← factor(survey$gender)
level(survey$gender) ← c("F", "M")
```

Q2(a)

```
sample(c(1,2,3,4), size = 1, prob = c(0.2, 0.4, 0.3, 0.1))
```

replace=T..

(b)

```
r← sample(c(1,2,3,4), size = 2000, prob = c(0.2, 0.4, 0.3, 0.1), replace = T)
r← matrix(r, nrow = 1000)
```

(c)

```
table( r[,1], r[,2])
```

Q3

pi/50

```
thetas <- seq(-16*pi, 16*pi, 0.0001)
xs <- exp(0.05 * thetas) * cos(thetas)
ys <- exp(0.05 * thetas) * sin(thetas)
plot(xs, ys, type="l")
```

Q4

```
max_num <- 1 000 000 000

sum <- 0

fibs <- 1
fibl <- 1


while (fibl <= max_num ){
    if(fibl %% 2 == 0 ){
        sum <- sum+ fibl
    }

    tmp <- fibl+ fibs
    fibs <- fibl
    fibl <- tmp
}
cat (Sum)
```

Q5

```
grid( lwd=2, lty=2)

arrows( 0, 2, 1, 2, lwd = 2, head.length = 0.1)
arrows(1, 2, 1, 1, lwd = 2, head.length = 0.1)
arrows(1, 1, 2, 1, lwd = 2, head.length = 0.1)
arrows(2, 1, 2, 0, lwd = 2, head.length = 0.1)
```

```
# Q5
plot(0, 0, type="n", xlim=c(0,2), ylim=c(0,2), bty="n", xlab="", ylab="")
segments(0, 0, 2, 0, col="grey", lty=2) # "grey" is optional
segments(0, 0, 0, 2, col="grey", lty=2)
segments(0, 1, 2, 1, col="grey", lty=2)
segments(1, 0, 1, 2, col="grey", lty=2)
segments(0, 2, 2, 2, col="grey", lty=2)
segments(2, 0, 2, 2, col="grey", lty=2)
arrows(0, 2, 1, 2, lwd=2, length=0.1)
arrows(1, 2, 1, 1, lwd=2, length=0.1)
arrows(1, 1, 2, 1, lwd=2, length=0.1)
arrows(2, 1, 2, 0, lwd=2, length=0.1)
```