Scientific Programming Assignment 2

MPhil in Computational Biology

November 1, 2022

If there are errors found, I will update the assignment on the web at
`http://github.com/sje30/sp2022/a2`

Due date: 2022-11-15 23:45

Please submit your report to Moodle as a single .Rnw file. Name your file `spa2_XXXXXX.Rnw`, where XXXXXX is your code that the graduate office will give you. You must ensure that the file can be compiled into a PDF by someone else on subliminal using:

```
Rscript -e "knitr::knit2pdf('yourfile.Rnw')
```

Your .Rnw file should dynamically compute and report answers, rather than you writing the code and then typing your answers manually into the latex part of the document.

Your report must be a maximum of ten pages, excluding the appendix. This course work will consist of 30% towards your overall mark for this module.

You must use only the core packages that are provided on subliminal. Your default setup on subliminal should include knitr, e.g. run this command to check where the package is being loaded from.

```
system.file(package='knitr')
```

On subliminal it should report the knitr package is in `/usr/lib/R/site-library/knitr`

In both questions, you should read in data in the format required using these two functions which you can copy. I may, or may not, then change the data files on github post submission, so your code should work with changes to the data files.

```
data1 <- function(d) {
  scan(quiet=TRUE,
       paste0("https://raw.githubusercontent.com/sje30/sp2022/",
              "main/assigns/a2/s/",d))
}
data2 <- function(d) {
  read.csv(paste0("https://raw.githubusercontent.com/sje30/sp2022/",
              "main/assigns/a2/p/",d))
}
```

```
par(mfrow=c(1,3),mar=c(2,2,1.5,1))
drawgrid(data1('s1.dat'), data1('g1.dat'), main='A')
drawgrid(data1('s2.dat'), data1('g2.dat'), main='B')
drawgrid(data1('s3.dat'), data1('g3.dat'), main='C')
```
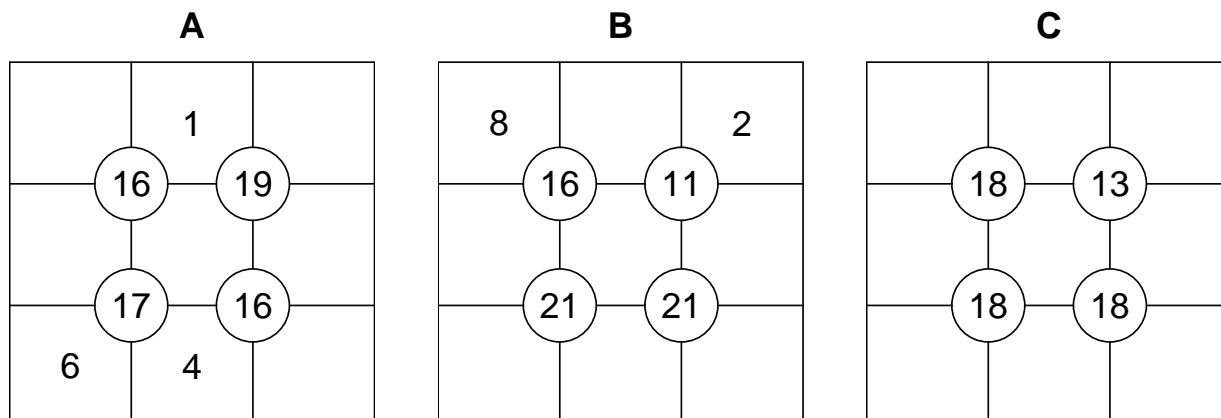

Figure 1: Three sujiko problems to be solved.

1 Sujiko problems [15 marks]

(Thanks to the Telegraph newspaper for these problems.) In this game, place the numbers 1 to 9 (without repetition) in the grid squares such that the total in each circle equals the sum of the numbers in the surrounding four squares.

1. Write the function "drawgrid" that given the four sums in the circles and any numbers in the squares, the problem is drawn. Use your function to draw the three grids shown in Figure 1. [5 marks]

2. Write a function with the following template:

```
sujiko <- function(sums, squares) {
  ## SUMS is a vector of length 4 containing the sums of the four circles.
  ## SQUARES is a vector of length 9 containing the solution so far.
  ##
  ## Return all completed solutions.
}
```

to solve a sujiko problem. [5 marks]

3. What are the solutions to the problems in Figure 1? (If you can't correctly write the function in part 2, write code that can be run to solve each problem.) [5 marks]

Hints: to solve this problem, the search space is small enough that you can exhaustively search all possible solutions. For drawing primitives, look at the functions: *segments()*, *symbols()*, *text()*. The following function taken may be copied if you wish:

```
## Taken from: http://stackoverflow.com/questions/11095992
permutations <- function(n){
    if(n==1){
        return(matrix(1))
```

2

```r
    } else {
        sp <- permutations(n-1)
        p <- nrow(sp)
        A <- matrix(nrow=n*p,ncol=n)
        for(i in 1:n){
            A[(i-1)*p+1:p,] <- cbind(i,sp+(sp>=i))
        }
        return(A)
    }
}
```

2 Permutation testing [15 marks]

For this you will need to first read about the Permutation test for measuring the signifance of an effect between two groups of measurements. Follow the definition in `https://en.wikipedia.org/wiki/Permutation_test`.

Your job is to use permutation test to examine the significance of two different experiments. The data are stored in `https://github.com/sje30/sp2022/tree/main/assigns/a2/p`

and should be read in using the data2 function, e.g.

```
expt1 = data2('expt1.dat')
summary(expt1)

##     group              val
##  Length:24        Min.   : 1.394
##  Class :character  1st Qu.: 6.121
##  Mode  :character  Median : 9.144
##                    Mean   : 8.476
##                    3rd Qu.:10.297
##                    Max.   :13.400
```

Column 1 is the group (control [ctl] or experimental [expt]) and column 2 is the measurement (arbitrary units). You can assume the data were collected independently on different indivdiuals. The two experiments are also independent, and should be analysed separately.

Your job is to write code that analyses the two experiments and summarises the results. You should write your code in such a way that the code should still work meaningfully if/when I change the datasets after the submission deadline.

You should generate some simulated data of your own to check that your test is working correctly. Explain how you checked your permutation test.

Your report should include graphical and textual summaries of the data and permutation tests. Compare the results from permutation tests with the t-test (available in R as `t.test`).