HW01p

Rebecca Strauss

February x, 2018

```
knitr::opts_chunk$set(error = TRUE) #this allows errors to be printed into the PDF
```

Welcome to HW01p where the "p" stands for "practice" meaning you will use R to solve practical problems. This homework is due 11:59 PM Satuday 2/24/18.

You should have RStudio installed to edit this file. You will write code in places marked "TO-DO" to complete the problems. Some of this will be a pure programming assignment. The tools for the solutions to these problems can be found in the class practice lectures. I want you to use the methods I taught you, not for you to google and come up with whatever works. You won't learn that way.

To "hand in" the homework, you should compile or publish this file into a PDF that includes output of your code. Once it's done, push by the deadline.

R Basics

First, install the package `testthat` (a widely accepted testing suite for R) from https://github.com/r-lib/testthat using `pacman`. If you are using Windows, this will be a long install, but you have to go through it for some of the stuff we are doing in class. LINUX (or MAC) is preferred for coding. If you can't get it to work, install this package from CRAN (still using `pacman`), but this is not recommended long term.

```
pacman::p_load(testthat)
```

1. Use the `seq` function to create vector `v` consisting of all numbers from -100 to 100.

```
v = seq(-100, 100)
```

Test using the following code:

```
expect_equal(v, -100 : 100)
```

If there are any errors, the `expect_equal` function will tell you about them. If there are no errors, then it will be silent.

2. Create a function `my_reverse` which takes as required input a vector and returns the vector in reverse where the first entry is the last entry, etc. No function calls are allowed inside your function (otherwise that would defeat the purpose of the exercise).

```
my_reverse <- function(vec) {
  j = length(vec)
  i = 1

  if(j %% 2 == 1) { #if length is odd
    while(TRUE) {
      temp = vec[i]
      vec[i] = vec[j]
      vec[j] = temp
      j = j - 1
      i = i + 1
      if(i == j) {
        break
```

```
      }
    }
  }else {
    while (i <= (j / 2) + 1) {
      temp = vec[i]
      vec[i] = vec[j]
      vec[j] = temp
      j = j - 1
      i = i + 1
    }
  }
  vec
}
```

Test using the following code:

```
expect_equal(my_reverse(c("A", "B", "C")), c("C", "B", "A"))
expect_equal(my_reverse(v), rev(v))
```

3. Let `n = 50`. Create a nxn matrix `R` of exactly 50% entries 0's, 25% 1's 25% 2's in random locations.

```
n = 50
stuff = c(rep(0, 1250), rep(1, 625), rep(2, 625)) #creates vector with 1250 0s, 625 1s, 625 2s. 1250+62
R = matrix(sample(stuff), nrow = n, ncol = n)#sample(stuff) shuffles the position of the entries. then
```

Test using the following and write two more tests as specified below:

```
expect_equal(dim(R), c(n, n))
#TO-DO test that the only unique values are 0, 1, 2
#TO-DO test that there are exactly 625 2's
```

4. Randomly punch holes (i.e. `NA`) values in this matrix so that approximately 30% of the entries are missing.

```
r = c(R) #flatten R
x = sample(length(r), .30*length(r))
for(i in x){
  R[i] = NA
}
```

Test using the following code. Note this test may fail 1/100 times.

```
num_missing_in_R = sum(is.na(c(R)))
expect_lt(num_missing_in_R, qbinom(0.995, n^2, 0.3))
expect_gt(num_missing_in_R, qbinom(0.005, n^2, 0.3))
```

5. Sort the rows matrix `R` by the largest row sum to lowest. See 2/3 way through practice lecture 3 for a hint.

```
row_sums = rep(0, 50) #create vector of size 50 to hold each row sum

#loop to calculate each row sum value
for(i in 1:nrow(R)) {
  sum_temp = 0 #reset sum holder to 0 after each new row
  for(j in 1:ncol(R)) {
    if(is.na(R[i, j]) == FALSE) {
      sum_temp = sum_temp + R[i, j]
    }
```

```
  }
  row_sums[i] = sum_temp
}

#add the row_sums vector as the 51st column
cbind(R, row_sums)
```

```
##
##  [1,] NA  1  2  0 NA  0  2  0  1 NA NA  2  2 NA NA  0 NA  2 NA  0  0 NA  0
##  [2,]  1  2 NA NA  0  1 NA  0 NA  1  0  1 NA  1  0  0 NA NA  0  0  2  2  2
##  [3,]  2 NA  0  1  0  0  0  0  0 NA  2  0 NA  0  2 NA  0  2  1  2  0  0  1
##  [4,]  2 NA  0  1 NA  1  0  0  0  2  0  0  0  0  2  0  1  0 NA  2  2  2  2
##  [5,]  0  2  0 NA  2  0 NA  2 NA  2  2  0 NA  1 NA NA  1  0  0  0  0  0  2
##  [6,] NA  2 NA  0 NA  2  2 NA  0 NA NA  0  2  0  0 NA  0 NA  2 NA  0 NA NA
##  [7,]  0  0 NA  1 NA  0 NA  1 NA NA NA  0  1  0  0  0 NA NA  1  0  0  2 NA
##  [8,] NA  1 NA  1 NA  2  0 NA  0  0  1  1  0  0  2  0 NA  0  0  1  2  0  1
##  [9,]  2 NA  0  0 NA NA  1  1  1  1  0 NA  0  0  2  1  0  2 NA  1 NA NA  0
## [10,] NA  2  1  0 NA  1  0  0 NA NA  0 NA  0  2 NA  2 NA NA  1  2 NA  0  2
## [11,]  2  0  0  1  0  1  0  0  0  0  1 NA  1  0  1  1  0  2 NA  1 NA  2  0
## [12,]  2  2 NA NA  0  0  0  0  0 NA  0  0  2 NA NA NA  0  0  2  2  1  2  0
## [13,]  1  1 NA NA NA  2  0  0  1  1 NA  0  0  0 NA  0  2 NA  0  1  0  2  2
## [14,]  0  0  0  1  1  2  0  0 NA  0 NA  0  0 NA  1  2 NA  0  0  0  2  0 NA
## [15,] NA  0  0 NA  1  1 NA  2 NA  1 NA  2  0  0  0  0  1 NA  2  0  0  0  0
## [16,] NA  1 NA NA  0  0 NA  2 NA  0  0  1  1  0  1  0  1  2 NA NA  2  2 NA
## [17,] NA  1  1  0  1  2 NA  0  0  2  1 NA  0  1 NA NA NA NA  2  2  0  1  2
## [18,]  0  0  0  0  2  1 NA NA  2  0  1  2  2  1  1  2  0  1  1  1 NA  0  0
## [19,]  0  0  2  2  1 NA NA  0  1 NA  2  0  1  0  0 NA NA NA NA NA  0 NA  0
## [20,] NA  0  1  0  1 NA  0 NA  0 NA  1  1 NA  1  2  2  2 NA NA NA  2 NA NA
## [21,] NA  0 NA  0  0  0 NA  0  1  0 NA NA NA  1  0 NA NA NA  0  0 NA NA  0
## [22,]  0 NA  1  0 NA NA  0 NA  0  0  0  1  0  2  0  1  0 NA NA  1  2 NA NA
## [23,]  0 NA NA NA  0 NA NA  1  2  0 NA  0 NA  0  0  2  1 NA NA NA  2 NA  0
## [24,]  2  1 NA NA  1  1  1 NA NA NA  2  0  0  1  2 NA  1  0 NA  0 NA  2 NA
## [25,] NA NA  1  1 NA  1  0  0  1  0  0  0 NA  1  1  2  1 NA  0  1 NA  1  0
## [26,]  2 NA  1  2  2  0 NA  0  2 NA  0 NA  0  0 NA  0  2  2 NA NA  2  0  0
## [27,]  0  0 NA  2 NA  0  2  2  0  2  1 NA  1  0 NA  1  0  0  2  0  0  0 NA
## [28,]  1 NA  0  0  1 NA  0  0  0  0  0 NA  0 NA  0  0  0  0  1  2 NA  1  0
## [29,]  0  2  0  0  1  2  0 NA  0 NA  1  0  1  0  1 NA  0  2 NA  1  2  1  1
## [30,]  0 NA  2 NA  1  1  0  0  1  0  0  2  2  1  0 NA  0  1  1 NA NA  2  0
## [31,]  0 NA NA  0 NA  2 NA NA NA NA  0  0 NA  0  0  0  1  1 NA  0  0  0 NA
## [32,]  0  1  0  1  0  2  1  0  1  0 NA NA NA  0 NA NA  0  2 NA NA NA  0  1
## [33,]  2  1 NA  2  0  1  0  0  1 NA  0  0  1 NA  2 NA NA NA NA  2  1 NA  0
## [34,] NA  2  0  0 NA  1  0  2  2  0 NA NA NA  0  1  1  2  0  0  0  1  1  0
## [35,]  1 NA  2  1  0 NA  2  2  0  2  1 NA NA NA  1  2  1 NA  1 NA NA NA  0
## [36,]  2  0  0  0  2 NA  1 NA NA NA NA  2  0  0  0 NA NA  0 NA  0  1  0  2
## [37,]  0 NA NA  0 NA  0  1 NA  2  2  0 NA NA  1  0  1 NA  0  2  0 NA NA  0
## [38,] NA  0  0  2  1  0 NA  0  1  2  2  0  1  0 NA  2  0  0 NA  2  1  1  0
## [39,]  2  2  1  2  2  1  1 NA  1 NA  2  2 NA  0  0  1 NA  0 NA NA  2  0 NA
## [40,]  2  1  0  2  0  1  1 NA  0  1  2 NA  1  0  0  0  0  0  2  0  0  2  1
## [41,]  0 NA NA  2 NA  0 NA NA NA NA  0 NA NA  1  2  0 NA NA  0  1  1  1  0
## [42,]  2  0  0  0  2 NA  0  1 NA  2  0 NA  0 NA  2 NA  0  2 NA  1  0  0 NA
## [43,]  1 NA NA  1  1 NA  1 NA NA  1 NA  1 NA  1 NA  1  0  0  0  1  0  0  1
## [44,]  1  0  1  0  0  0  0  0  1  1 NA NA  1  2 NA  1  1  0  0  0  0  1  2
## [45,]  2 NA  1  0  0  1  1  0  2  0  2 NA  1 NA NA  0  0 NA NA  0  1  2 NA
## [46,]  0  1  2  2  0  2  1 NA  0 NA  0 NA NA  0  0  0 NA  0  0  1 NA  1  2
```

```
## [47,]  0 NA  0  2  2  1  0  1  0  0  1  2 NA  0  0  2  1  0  2  1  2  2  0
## [48,] NA NA NA  0  0  0  2  1  2 NA  2  2  0 NA  0  0 NA  0 NA NA  1  0 NA
## [49,]  0  0  0  0  0 NA  2  1  0  1  0 NA  2  2  1 NA NA NA NA NA  0 NA  1
## [50,]  2  2  1  0  0  0 NA NA  1  0  2  0  0  1  0  2  0  1 NA  2 NA NA NA
##
##  [1,]  2 NA  1  0  2  0  0 NA  2  1 NA NA  1  0  0 NA  2  2  0  1  1 NA NA
##  [2,]  0  0  0  2 NA NA NA NA  0  1 NA  0  1 NA NA  0 NA  2 NA  2 NA  2  0
##  [3,]  0 NA  2  0  1  0 NA  0  1  0  0 NA  0  1 NA NA  2  0  2 NA  0 NA  1
##  [4,]  2 NA  1  1 NA NA  0  0 NA NA  0 NA NA  2  1  0  2 NA  2 NA  1  0  0
##  [5,]  0 NA  0 NA  0 NA  2 NA  1  1 NA  2 NA  1  2  2  1  0  0 NA NA  0 NA
##  [6,]  0 NA NA  0  1  0  0  0  1 NA  0 NA  1  1 NA  2  1  2  0 NA  1 NA NA
##  [7,]  0  1  2 NA NA  2  0 NA NA  2  2  0 NA  0  0  0 NA  2 NA  2 NA NA  1
##  [8,] NA  1 NA  0 NA  0  0 NA  2 NA NA  0 NA NA  2  0  1  2  0 NA  0  0  2
##  [9,]  2 NA  0  2  1  2 NA  1  1 NA NA  1 NA  0  0  1  0  0  1 NA  0  1 NA
## [10,]  0  2  1  0 NA  2  1  1  2  0  0  2  0 NA  1  0 NA  0  0 NA  2  1  2
## [11,] NA NA  0  2  2  0  1  1  1  1  0  0 NA  0  0  2 NA  1 NA NA  2  2  2
## [12,] NA  2  1  0  1 NA  0  1  1  0 NA  0  1  2  1  0  1  0 NA  1 NA  2 NA
## [13,]  1 NA  0  0 NA NA  0  0  0  0  2  0  2  2  2  2  2 NA  2  0  0  0  0
## [14,] NA  0  1  0 NA  2  0 NA  1 NA NA  0 NA  0  1 NA  0  2  1  1  0 NA NA
## [15,]  0  1 NA NA  0 NA  0  0  0  0  0 NA NA NA  1  0  2 NA  2 NA NA  0  0
## [16,]  0  1 NA NA  0  0  0 NA  0 NA  0  1  0 NA NA  0  0  2  2  2  0 NA  1
## [17,]  0  0  1 NA  2  0 NA  0  0  1  1 NA  0  2 NA  0 NA  0 NA NA  0  0  0
## [18,]  2 NA  2  0  1  0  2  0  2  2 NA  0 NA  1  2  0  0  2  2  2  1  0 NA
## [19,]  2 NA  0 NA  2  1 NA NA  2  0  0  1 NA  0 NA NA  2  2 NA  1  2 NA  0
## [20,] NA  1 NA  0  0  1 NA  1  0  0  0  1  0 NA  1 NA NA  1  1  2 NA  0  0
## [21,]  2  2  1 NA  0  0 NA  1 NA NA  0  0  2  2 NA  0 NA  0 NA  0 NA  1  2
## [22,] NA  0  0  0  2  0 NA  0 NA  0 NA  0 NA NA  1  1 NA NA NA NA NA  1  2
## [23,] NA  1  0 NA  2  0 NA NA  0 NA  1 NA NA  1 NA  2  1  0  0  1 NA  2  0
## [24,] NA  0  0  2  1 NA  0 NA  1  0  1  1  0  2  2  0  0 NA  0 NA  2  0  0
## [25,]  0  2  1  0  0 NA  0 NA  0  2  0  0  2  2  1 NA  2  2  0 NA NA  2  0
## [26,] NA  0  0  2  0  2  0  0  2  2  0  1  2  1  0  0  0 NA NA NA NA NA  2
## [27,] NA  0 NA  2  0 NA  1  0  0  0  2 NA  1  0  1  0  2  0 NA  1 NA NA  1
## [28,]  0 NA NA  1  0  2  0  0 NA  2 NA  2  1  0 NA  0 NA NA  1  0 NA NA  0
## [29,] NA  0  0  1  2  1  0  0  1  0  0 NA  2  0  0  0 NA  0  2  1  0  2  0
## [30,]  1  1  0 NA NA NA  0  0  2  1  0  2 NA  0 NA NA  0 NA NA  2  2 NA  1
## [31,] NA NA  0  0  0  0 NA NA  1  0  1  0  2  1 NA  1  0  0 NA  1 NA NA  2
## [32,]  0 NA  1  0  0  0 NA  2 NA  0 NA NA NA NA  0  0  2 NA  2  2  0 NA  2
## [33,]  0  0  1  0  0  0  1 NA NA  0 NA  0  1  0 NA  1 NA  2 NA NA  0  0  0
## [34,] NA NA  2  0 NA  1  2  0  1 NA  2  0 NA  0  0  1  1  2  2 NA  0 NA NA
## [35,]  2  2  2  0  1  0  0  0 NA  0  0  2  1 NA  0  2 NA  0 NA NA  0  1 NA
## [36,] NA  0  1 NA  2  1  0  1 NA NA  0  0  1  0  0  0  0 NA  0 NA  2  1  1
## [37,]  1  2 NA  1  0  2  0 NA  1  0  0 NA  0  0  0 NA NA NA  0  2  2  0 NA
## [38,]  0  0  0 NA  0 NA  1  1  0 NA  0  0 NA  0  1  2 NA  0  2 NA  2  0 NA
## [39,]  1 NA  0  1  1  1 NA  0  0 NA  2  0 NA  1 NA  0  1 NA  1  1  0  0  0
## [40,]  0  1  0 NA  0  0 NA  0  1  2  1 NA  0  1 NA  2 NA  0 NA  1  0 NA  0
## [41,]  2 NA  0  1 NA  0  2  0  0  0 NA NA  0 NA  0  2 NA  2 NA  0 NA NA  0
## [42,]  0  0  0  0  2 NA  1  1  0  0  1  2 NA  0  2  0 NA  0 NA  0  2  1  1
## [43,] NA  1  2 NA  0  2  0  0  0  0 NA NA NA  0  0  1  0  0 NA  1 NA  0  1
## [44,]  1  0  1 NA  0 NA NA NA  0  2  1  0  1 NA NA  0  2  2  1  0  0  1  1
## [45,]  0  0  2  1  2  0  2  0  0  1  0  0 NA  2 NA NA NA  0  2 NA  2  1  2
## [46,]  2 NA  2 NA  0  0 NA NA  1  2  0  0  2  0 NA  1  0 NA  0 NA  1  1 NA
## [47,]  0 NA  2  1  0 NA  0  2 NA  0 NA  0 NA  2  2 NA  0 NA  0  0  0  0 NA
## [48,] NA  2 NA  0  2  0  2 NA NA NA  0 NA  2  2  2  0  0 NA  1  0  0  0  0
## [49,]  0  2  2  1  2  0  0  2  2  0 NA NA NA  1  2  1  2 NA NA  2  1  0 NA
```

```
## [50,] NA NA  0  2  1  2 NA  1 NA  1  1  0  1 NA NA  1 NA  0  2  1 NA  0  1
##                  row_sums
##  [1,]  0  1 NA  1       29
##  [2,]  1  1  2  0       27
##  [3,]  2  0 NA NA       25
##  [4,] NA  0  2 NA       31
##  [5,]  0  2 NA  1       29
##  [6,]  2  0  0  2       24
##  [7,]  1  0 NA NA       21
##  [8,]  2  0  1  0       25
##  [9,] NA NA  2  0       27
## [10,]  1  0  1 NA       32
## [11,]  0  0  0 NA       30
## [12,]  1  1  0  2       31
## [13,]  2  0  2  0       32
## [14,]  0  1  1  2       22
## [15,]  2  0  1  0       19
## [16,]  2  0  2  2       28
## [17,]  0  2  0 NA       25
## [18,]  1  0  2 NA       41
## [19,] NA NA  0 NA       24
## [20,]  2  2  2  2       30
## [21,]  2 NA  0 NA       17
## [22,] NA  0 NA  2       17
## [23,] NA  2 NA  0       21
## [24,]  0 NA NA  0       26
## [25,]  0  0 NA NA       27
## [26,] NA  0  0  2       31
## [27,]  0 NA  2  0       26
## [28,]  0  1  0  0       16
## [29,]  0  0 NA NA       27
## [30,]  0  0  0  2       28
## [31,] NA  0  0  0       13
## [32,] NA NA  2  0       22
## [33,]  0 NA NA  2       21
## [34,]  0  0  0  0       27
## [35,]  0  0 NA  0       29
## [36,]  1  1  2  0       24
## [37,]  1  0  1 NA       22
## [38,]  0  0  2  0       26
## [39,]  0  0  0  2       31
## [40,]  0  0 NA NA       25
## [41,] NA  1  0  1       19
## [42,]  2 NA  2  0       29
## [43,]  1  2  2  2       25
## [44,]  1 NA  2  0       28
## [45,]  1  2 NA NA       33
## [46,]  0  0  0 NA       24
## [47,]  1 NA NA NA       29
## [48,]  0  0  1  0       24
## [49,]  0  1  0 NA       31
## [50,]  0  0  0 NA       28
```

```r
#sort by the 51st column
R[order(R[,ncol(R)]),]
```

```
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
##  [1,]    1    2   NA   NA    0    1   NA    0   NA     1     0     1    NA
##  [2,]   NA    1   NA    1   NA    2    0   NA    0     0     1     1     0
##  [3,]    2   NA    0    0   NA   NA    1    1    1     1     0    NA     0
##  [4,]    1    1   NA   NA   NA    2    0    0    1     1    NA     0     0
##  [5,]   NA    0    0   NA    1    1   NA    2   NA     1    NA     2     0
##  [6,]    0   NA   NA   NA    0   NA   NA    1    2     0    NA     0    NA
##  [7,]    2    1   NA   NA    1    1    1   NA   NA    NA     2     0     0
##  [8,]    0    0   NA    2   NA    0    2    2    0     2     1    NA     1
##  [9,]    1   NA    0    0    1   NA    0    0    0     0     0    NA     0
## [10,]    0   NA   NA    0   NA    2   NA   NA   NA    NA     0     0    NA
## [11,]    0    1    0    1    0    2    1    0    1     0    NA    NA    NA
## [12,]   NA    2    0    0   NA    1    0    2    2     0    NA    NA    NA
## [13,]    1   NA    2    1    0   NA    2    2    0     2     1    NA    NA
## [14,]    2    0    0    0    2   NA    1   NA   NA    NA    NA     2     0
## [15,]   NA    0    0    2    1    0   NA    0    1     2     2     0     1
## [16,]    2    0    0    0    2   NA    0    1   NA     2     0    NA     0
## [17,]    1    0    1    0    0    0    0    0    1     1    NA    NA     1
## [18,]   NA   NA   NA    0    0    0    2    1    2    NA     2     2     0
## [19,]   NA    1    2    0   NA    0    2    0    1    NA    NA     2     2
## [20,]    0    2    0   NA    2    0   NA    2   NA     2     2     0    NA
## [21,]    0   NA   NA    2   NA    0   NA   NA   NA    NA     0    NA    NA
## [22,]   NA    2   NA    0   NA    2    2   NA    0    NA    NA     0     2
## [23,]    2    2   NA   NA    0    0    0    0    0    NA     0     0     2
## [24,]    0    0    0    1    1    2    0    0   NA     0    NA     0     0
## [25,]   NA    1   NA   NA    0    0   NA    2   NA     0     0     1     1
## [26,]   NA    0    1    0    1   NA    0   NA    0    NA     1     1    NA
## [27,]    0   NA    1    0   NA   NA    0   NA    0     0     0     1     0
## [28,]    2   NA    1    2    2    0   NA    0    2    NA     0    NA     0
## [29,]    0   NA    2   NA    1    1    0    0    1     0     0     2     2
## [30,]    2    1   NA    2    0    1    0    0    1    NA     0     0     1
## [31,]    2    2    1    2    2    1    1   NA    1    NA     2     2    NA
## [32,]    1   NA   NA    1    1   NA    1   NA   NA     1    NA     1    NA
## [33,]    2   NA    0    1    0    0    0    0    0    NA     2     0    NA
## [34,]    2   NA    0    1   NA    1    0    0    0     2     0     0     0
## [35,]    0    0   NA    1   NA    0   NA    1   NA    NA    NA     0     1
## [36,]   NA    2    1    0   NA    1    0    0   NA    NA     0    NA     0
## [37,]    2    0    0    1    0    1    0    0    0     0     1    NA     1
## [38,]   NA    1    1    0    1    2   NA    0    0     2     1    NA     0
## [39,]    0    0    0    0    2    1   NA   NA    2     0     1     2     2
## [40,]    0    0    2    2    1   NA   NA    0    1    NA     2     0     1
## [41,]   NA    0   NA    0    0    0   NA    0    1     0    NA    NA    NA
## [42,]   NA   NA    1    1   NA    1    0    0    1     0     0     0    NA
## [43,]    0    2    0    0    1    2    0   NA    0    NA     1     0     1
## [44,]    0   NA   NA    0   NA    0    1   NA    2     2     0    NA    NA
## [45,]    2    1    0    2    0    1    1   NA    0     1     2    NA     1
## [46,]    2   NA    1    0    0    1    1    0    2     0     2    NA     1
## [47,]    0    1    2    2    0    2    1   NA    0    NA     0    NA    NA
## [48,]    0   NA    0    2    2    1    0    1    0     0     1     2    NA
## [49,]    0    0    0    0    0   NA    2    1    0     1     0    NA     2
## [50,]    2    2    1    0    0    0   NA   NA    1     0     2     0     0
```

```
##       [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
##  [1,]     1     0     0    NA    NA     0     0     2     2     2     0
##  [2,]     0     2     0    NA     0     0     1     2     0     1    NA
##  [3,]     0     2     1     0     2    NA     1    NA    NA     0     2
##  [4,]     0    NA     0     2    NA     0     1     0     2     2     1
##  [5,]     0     0     0     1    NA     2     0     0     0     0     0
##  [6,]     0     0     2     1    NA    NA    NA     2    NA     0    NA
##  [7,]     1     2    NA     1     0    NA     0    NA     2    NA    NA
##  [8,]     0    NA     1     0     0     2     0     0     0    NA    NA
##  [9,]    NA     0     0     0     0     1     2    NA     1     0     0
## [10,]     0     0     0     1     1    NA     0     0     0    NA    NA
## [11,]     0    NA    NA     0     2    NA    NA    NA     0     1     0
## [12,]     0     1     1     2     0     0     0     1     1     0    NA
## [13,]    NA     1     2     1    NA     1    NA    NA    NA     0     2
## [14,]     0     0    NA    NA     0    NA     0     1     0     2    NA
## [15,]     0    NA     2     0     0    NA     2     1     1     0     0
## [16,]    NA     2    NA     0     2    NA     1     0     0    NA     0
## [17,]     2    NA     1     1     0     0     0     0     1     2     1
## [18,]    NA     0     0    NA     0    NA    NA     1     0    NA    NA
## [19,]    NA    NA     0    NA     2    NA     0     0    NA     0     2
## [20,]     1    NA    NA     1     0     0     0     0     0     2     0
## [21,]     1     2     0    NA    NA     0     1     1     1     0     2
## [22,]     0     0    NA     0    NA     2    NA     0    NA    NA     0
## [23,]    NA    NA    NA     0     0     2     2     1     2     0    NA
## [24,]    NA     1     2    NA     0     0     0     2     0    NA    NA
## [25,]     0     1     0     1     2    NA    NA     2     2    NA     0
## [26,]     1     2     2     2    NA    NA    NA     2    NA    NA    NA
## [27,]     2     0     1     0    NA    NA     1     2    NA    NA    NA
## [28,]     0    NA     0     2     2    NA    NA     2     0     0    NA
## [29,]     1     0    NA     0     1     1    NA    NA     2     0     1
## [30,]    NA     2    NA    NA    NA    NA     2     1    NA     0     0
## [31,]     0     0     1    NA     0    NA    NA     2     0    NA     1
## [32,]     1    NA     1     0     0     0     1     0     0     1    NA
## [33,]     0     2    NA     0     2     1     2     0     0     1     0
## [34,]     0     2     0     1     0    NA     2     2     2     2     2
## [35,]     0     0     0    NA    NA     1     0     0     2    NA     0
## [36,]     2    NA     2    NA    NA     1     2    NA     0     2     0
## [37,]     0     1     1     0     2    NA     1    NA     2     0    NA
## [38,]     1    NA    NA    NA    NA     2     2     0     1     2     0
## [39,]     1     1     2     0     1     1     1    NA     0     0     2
## [40,]     0     0    NA    NA    NA    NA    NA     0    NA     0     2
## [41,]     1     0    NA    NA    NA     0     0    NA    NA     0     2
## [42,]     1     1     2     1    NA     0     1    NA     1     0     0
## [43,]     0     1    NA     0     2    NA     1     2     1     1    NA
## [44,]     1     0     1    NA     0     2     0    NA    NA     0     1
## [45,]     0     0     0     0     0     2     0     0     2     1     0
## [46,]    NA    NA     0     0    NA    NA     0     1     2    NA     0
## [47,]     0     0     0    NA     0     0     1    NA     1     2     2
## [48,]     0     0     2     1     0     2     1     2     2     0     0
## [49,]     2     1    NA    NA    NA    NA    NA     0    NA     1     0
## [50,]     1     0     2     0     1    NA     2    NA    NA    NA    NA
##       [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35]
##  [1,]     0     0     2    NA    NA    NA    NA     0     1    NA     0
##  [2,]     1    NA     0    NA     0     0    NA     2    NA    NA     0
```

```
##  [3,]    NA     0     2     1     2    NA     1     1    NA    NA     1
##  [4,]    NA     0     0    NA    NA     0     0     0     0     2     0
##  [5,]     1    NA    NA     0    NA     0     0     0     0     0    NA
##  [6,]     1     0    NA     2     0    NA    NA     0    NA     1    NA
##  [7,]     0     0     2     1    NA     0    NA     1     0     1     1
##  [8,]     0    NA     2     0    NA     1     0     0     0     2    NA
##  [9,]    NA    NA     1     0     2     0     0    NA     2    NA     2
## [10,]    NA     0     0     0     0    NA    NA     1     0     1     0
## [11,]    NA     1     0     0     0    NA     2    NA     0    NA    NA
## [12,]    NA     2     0    NA     1     2     0     1    NA     2     0
## [13,]     2     2     0     1     0     0     0    NA     0     0     2
## [14,]     0     1    NA     2     1     0     1    NA    NA     0     0
## [15,]     0     0    NA     0    NA     1     1     0    NA     0     0
## [16,]     0     0     0     2    NA     1     1     0     0     1     2
## [17,]     0     1    NA     0    NA    NA    NA     0     2     1     0
## [18,]     2    NA     0     2     0     2    NA    NA    NA     0    NA
## [19,]    NA     1     0     2     0     0    NA     2     1    NA    NA
## [20,]    NA     0    NA     0    NA     2    NA     1     1    NA     2
## [21,]    NA     0     1    NA     0     2     0     0     0    NA    NA
## [22,]    NA    NA     0     1     0     0     0     1    NA     0    NA
## [23,]     2     1     0     1    NA     0     1     1     0    NA     0
## [24,]     0     1     0    NA     2     0    NA     1    NA    NA     0
## [25,]     1    NA    NA     0     0     0    NA     0    NA     0     1
## [26,]     1    NA     0     0     1    NA     1     0     0     0     1
## [27,]     0     0     0     2     0    NA     0    NA     0    NA     0
## [28,]     0     0     2     0     2     0     0     2     2     0     1
## [29,]     1     0    NA    NA    NA     0     0     2     1     0     2
## [30,]     0     1     0     0     0     1    NA    NA     0    NA     0
## [31,]    NA     0     1     1     1    NA     0     0    NA     2     0
## [32,]     1     2    NA     0     2     0     0     0     0    NA    NA
## [33,]    NA     2     0     1     0    NA     0     1     0     0    NA
## [34,]    NA     1     1    NA    NA     0     0    NA    NA     0    NA
## [35,]     1     2    NA    NA     2     0    NA    NA     2     2     0
## [36,]     2     1     0    NA     2     1     1     2     0     0     2
## [37,]    NA     0     2     2     0     1     1     1     1     0     0
## [38,]     0     1    NA     2     0    NA     0     0     1     1    NA
## [39,]    NA     2     0     1     0     2     0     2     2    NA     0
## [40,]    NA     0    NA     2     1    NA    NA     2     0     0     1
## [41,]     2     1    NA     0     0    NA     1    NA    NA     0     0
## [42,]     2     1     0     0    NA     0    NA     0     2     0     0
## [43,]     0     0     1     2     1     0     0     1     0     0    NA
## [44,]     2    NA     1     0     2     0    NA     1     0     0    NA
## [45,]     1     0    NA     0     0    NA     0     1     2     1    NA
## [46,]     0     2     1     2     0     2     0     0     1     0     0
## [47,]    NA     2    NA     0     0    NA    NA     1     2     0     0
## [48,]    NA     2     1     0    NA     0     2    NA     0    NA     0
## [49,]     2     2     1     2     0     0     2     2     0    NA    NA
## [50,]    NA     0     2     1     2    NA     1    NA     1     1     0
##       [,36] [,37] [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46]
##  [1,]     1    NA    NA     0    NA     2    NA     2    NA     2     0
##  [2,]    NA    NA     2     0     1     2     0    NA     0     0     2
##  [3,]    NA     0     0     1     0     0     1    NA     0     1    NA
##  [4,]     2     2     2     2     2    NA     2     0     0     0     0
##  [5,]    NA    NA     1     0     2    NA     2    NA    NA     0     0
```

```
##  [6,]    NA     1    NA     2     1     0     0     1    NA     2     0
##  [7,]     0     2     2     0     0    NA     0    NA     2     0     0
##  [8,]     1     0     1     0     2     0    NA     1    NA    NA     1
##  [9,]     1     0    NA     0    NA    NA     1     0    NA    NA     0
## [10,]     2     1    NA     1     0     0    NA     1    NA    NA     2
## [11,]    NA    NA     0     0     2    NA     2     2     0    NA     2
## [12,]    NA     0     0     1     1     2     2    NA     0    NA    NA
## [13,]     1    NA     0     2    NA     0    NA    NA     0     1    NA
## [14,]     1     0     0     0     0    NA     0    NA     2     1     1
## [15,]    NA     0     1     2    NA     0     2    NA     2     0    NA
## [16,]    NA     0     2     0    NA     0    NA     0     2     1     1
## [17,]     1    NA    NA     0     2     2     1     0     0     1     1
## [18,]     2     2     2     0     0    NA     1     0     0     0     0
## [19,]     1     0     0    NA     2     2     0     1     1    NA    NA
## [20,]    NA     1     2     2     1     0     0    NA    NA     0    NA
## [21,]     0    NA     0     2    NA     2    NA     0    NA    NA     0
## [22,]     1     1    NA     2     1     2     0    NA     1    NA    NA
## [23,]     1     2     1     0     1     0    NA     1    NA     2    NA
## [24,]    NA     0     1    NA     0     2     1     1     0    NA    NA
## [25,]     0    NA    NA     0     0     2     2     2     0    NA     1
## [26,]     0    NA     1    NA    NA     1     1     2    NA     0     0
## [27,]    NA    NA     1     1    NA    NA    NA    NA    NA     1     2
## [28,]     2     1     0     0     0    NA    NA    NA    NA    NA     2
## [29,]    NA     0    NA    NA     0    NA    NA     2     2    NA     1
## [30,]     1     0    NA     1    NA     2    NA    NA     0     0     0
## [31,]    NA     1    NA     0     1    NA     1     1     0     0     0
## [32,]    NA     0     0     1     0     0    NA     1    NA     0     1
## [33,]     0     1    NA    NA     2     0     2    NA     0    NA     1
## [34,]    NA     2     1     0     2    NA     2    NA     1     0     0
## [35,]    NA     0     0     0    NA     2    NA     2    NA    NA     1
## [36,]     0    NA     1     0    NA     0     0    NA     2     1     2
## [37,]    NA     0     0     2    NA     1    NA    NA     2     2     2
## [38,]     0     2    NA     0    NA     0    NA    NA     0     0     0
## [39,]    NA     1     2     0     0     2     2     2     1     0    NA
## [40,]    NA     0    NA    NA     2     2    NA     1     2    NA     0
## [41,]     2     2    NA     0    NA     0    NA     0    NA     1     2
## [42,]     2     2     1    NA     2     2     0    NA    NA     2     0
## [43,]     2     0     0     0    NA     0     2     1     0     2     0
## [44,]     0     0     0    NA    NA    NA     0     2     2     0    NA
## [45,]     0     1    NA     2    NA     0    NA     1     0    NA     0
## [46,]    NA     2    NA    NA    NA     0     2    NA     2     1     2
## [47,]     2     0    NA     1     0    NA     0    NA     1     1    NA
## [48,]    NA     2     2    NA     0    NA     0     0     0     0    NA
## [49,]    NA     1     2     1     2    NA    NA     2     1     0    NA
## [50,]     1    NA    NA     1    NA     0     2     1    NA     0     1
##       [,47] [,48] [,49] [,50]
##  [1,]     1     1     2     0
##  [2,]     2     0     1     0
##  [3,]    NA    NA     2     0
##  [4,]     2     0     2     0
##  [5,]     2     0     1     0
##  [6,]    NA     2    NA     0
##  [7,]     0    NA    NA     0
##  [8,]     0    NA     2     0
```

```
##  [9,]     0     1     0     0
## [10,]    NA     0     0     0
## [11,]    NA    NA     2     0
## [12,]     0     0     0     0
## [13,]     0     0    NA     0
## [14,]     1     1     2     0
## [15,]     0     0     2     0
## [16,]     2    NA     2     0
## [17,]     1    NA     2     0
## [18,]     0     0     1     0
## [19,]     0     1    NA     1
## [20,]     0     2    NA     1
## [21,]    NA     1     0     1
## [22,]     2     0     0     2
## [23,]     1     1     0     2
## [24,]     0     1     1     2
## [25,]     2     0     2     2
## [26,]     2     2     2     2
## [27,]    NA     0    NA     2
## [28,]    NA     0     0     2
## [29,]     0     0     0     2
## [30,]     0    NA    NA     2
## [31,]     0     0     0     2
## [32,]     1     2     2     2
## [33,]     2     0    NA    NA
## [34,]    NA     0     2    NA
## [35,]     1     0    NA    NA
## [36,]     1     0     1    NA
## [37,]     0     0     0    NA
## [38,]     0     2     0    NA
## [39,]     1     0     2    NA
## [40,]    NA    NA     0    NA
## [41,]     2    NA     0    NA
## [42,]     0     0    NA    NA
## [43,]     0     0    NA    NA
## [44,]     1     0     1    NA
## [45,]     0     0    NA    NA
## [46,]     1     2    NA    NA
## [47,]     0     0     0    NA
## [48,]     1    NA    NA    NA
## [49,]     0     1     0    NA
## [50,]     0     0     0    NA
```

Test using the following code.

```
for (i in 2 : n){
  expect_gte(sum(R[i - 1, ], na.rm = TRUE), sum(R[i, ], na.rm = TRUE))
}
```

```
## Error: sum(R[i - 1, ], na.rm = TRUE) is not more than sum(R[i, ], na.rm = TRUE). Difference: -6
```

6. Create a vector v consisting of a sample of 1,000 iid normal realizations with mean -10 and variance 10.

```
pop <- 1000
avg <- - 10
var <- 10
```

```
v <- rnorm(pop, avg, sqrt(var))
```

Find the average of v and the standard error of v.

```
sum(v)/length(v) #avg
```

```
## [1] -9.996787
```

```
sqrt(var)/sqrt(pop)#standard error
```

```
## [1] 0.1
```

Find the 5%ile of v and use the qnorm function as part of a test to ensure it is correct based on probability theory.

```
a <- quantile(v, .05)
b <- qnorm(.05, avg, sqrt(var))
#TEST
expect_equal(as.numeric(a), as.numeric(b), tol = .05)
```

Find the sample quantile corresponding to the value -7000 of v and use the pnorm function as part of a test to ensure it is correct based on probability theory.

```
inverse_quantile_obj = ecdf(v)#set inverse_quantile_obj equal to the inverse cdf function
inverse <- inverse_quantile_obj(-7000)#call function on value
pdf <- pnorm(-7000, avg, sqrt(var))#since cdf is the inverse of pdf, (cdf = inverse(pdf)), then inverse
#expect_equal(..., tol = )
expect_equal(inverse, pdf)
```

7. Create a list named my_list with keys "A", "B", ... where the entries are arrays of size 1, 2 x 2, 3 x 3 x 3, etc. Fill the array with the numbers 1, 2, 3, etc. Make 8 entries.

```
my_list = list()
my_list$A = array()
my_list$B = array()
my_list$C = array()
my_list$D = array()
my_list$E = array()
my_list$F = array()
my_list$G = array()
my_list$H = array()

for(i in 1:8) {
  my_list[[i]] = array(1:i^2, dim = c(i, i))
  if(i == 1){
    my_list[[i]] = 1
  }
}
```

Test with the following uncomprehensive tests:

```
expect_equal(my_list$A, 1)
expect_equal(my_list[[2]][, 1], 1 : 2)
expect_equal(dim(my_list[["H"]]), rep(8, 8))
```

```
## Error: dim(my_list[["H"]]) not equal to rep(8, 8).
## Lengths differ: 2 is not 8
```

Run the following code:

```
lapply(my_list, object.size)
```

```
## $A
## 48 bytes
##
## $B
## 216 bytes
##
## $C
## 248 bytes
##
## $D
## 264 bytes
##
## $E
## 328 bytes
##
## $F
## 344 bytes
##
## $G
## 400 bytes
##
## $H
## 456 bytes
```

Use `?lapply` and `?object.size` to read about what these functions do. Then explain the output you see above. For the later arrays, does it make sense given the dimensions of the arrays?

Answer here in English.

Now cleanup the namespace by deleting all stored objects and functions:

```
rm()
```

Basic Binary Classification Modeling

8. Load the famous `iris` data frame into the namespace. Provide a summary of the columns and write a few descriptive sentences about the distributions using the code below and in English.

```
iris = iris
summary(iris)
```

```
##   Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
##  Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
##  1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
##  Median :5.800   Median :3.000   Median :4.350   Median :1.300
##  Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
##  3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
##  Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
##        Species
##  setosa    :50
##  versicolor:50
##  virginica :50
##
```

```
##
##
```

```
#there are 3 specicies. length of sepal has mean=5.84, IQR = 1.3, range=3.6
#sepal width mean=3.057, IQR = 0.5, range=2.40
#petal length mean=3.758, IQR = 3.5, range=5.90
#petal width mean=1.199, IQR = 1.5, range=2.40
```

The outcome metric is `Species`. This is what we will be trying to predict. However, we have only done binary classification in class (i.e. two classes). Thus the first order of business is to drop one class. Let's drop the level "virginica" from the data frame.

```
iris = subset(iris, Species != "virginica")#creates subset of all rows except the ones where species na
```

Now create a vector `y` that is length the number of remaining rows in the data frame whose entries are 0 if "setosa" and 1 if "versicolor".

```
y = rep(0, 100)
for(i in 1:nrow(iris)) {
  y[i] = as.numeric(iris[i, 5])-1 #-1 b/c if you call as.numeric(setosa) it returns a 1 and  as.numeric
}
```

9. Fit a threshold model to `y` using the feature `Sepal.Length`. Try to write your own code to do this. What is the estimated value of the threshold parameter? What is the total number of errors this model makes?

```
X = iris$Sepal.Length
Xy = cbind(X, y)
Xy = as.data.frame(Xy)
colnames(Xy) = c("Length", "species")
n = nrow(Xy)
num_err_by_par = matrix(NA, nrow = n, ncol = 2)
colnames(num_err_by_par) = c("threshold", "num_errors")

y_check = 0

for(i in 1:n) {
  threshold = Xy$Length[i]
  num_errors = sum((Xy$Length > threshold) != y_check)
  num_err_by_par[i, ] = c(threshold, num_errors)
}

best_row = order(num_err_by_par[, "num_errors"])[1]
num_err_by_par[best_row, "threshold"]
```

```
## threshold
##         7
```

Threshold parameter ~ 7 Does this make sense given the following summaries:

```
summary(iris[iris$Species == "setosa", "Sepal.Length"])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   4.300   4.800   5.000   5.006   5.200   5.800
```

```
summary(iris[iris$Species == "virginica", "Sepal.Length"])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##
```

Write your answer here in English.

10. Fit a perceptron model explaining `y` using all three features. Try to write your own code to do this. Provide the estimated parameters (i.e. the four entries of the weight vector)? What is the total number of errors this model makes?

```
max_iterations = 1000
y = as.numeric((iris$Species))-1  #using same y from question : Now create a vector `y` that is length
w_vec = rep(0, 5)#initialize weights
feats = iris[ , 1:4]#set feats to the columns of the iris data (without species col) & with no instance
feats1 = as.matrix(cbind(1, feats))

for(iter in 1:max_iterations) {
  for(i in 1:nrow(feats1)) {
    x_i = feats1[i, ] #set vector x_1 equal to i'th row containing all its' columns
    y_i = y[i] #set y_i equal to the binary value located at postion i in the vector y
    yhat_i = ifelse(sum(x_i %*% w_vec) > 0, 1, 0)
    w_vec = w_vec + (y_i - yhat_i)*x_i
  }
}
w_vec
```

```
##            1 Sepal.Length  Sepal.Width Petal.Length  Petal.Width
##         -1.0         -1.1         -3.6          5.2          2.2
```