

**CS 698 R Programming and Data Analytics**  
**Midterm Examination**  
**Monday, March 21, 2016**

- This midterm examination consists of 10 pages (including this one), 7 questions, and 100 points. Please check to make sure you have all the pages.
- This is a closed-book/notes exam, but an A4-sized cheating sheet is allowed.
- Write your answers on the given space for each question.
- Hand in both your answer sheets and this midterm examination.
- You have 120 minutes to complete the exam.

On my honor, I pledge that I have not violated the provisions of the NJIT Academic Honor Code. I understand that if I do not include this pledge, as required by the instructor, with my signature, the instructor will not grade my work.

Last Name (printed): \_\_\_\_\_

First Name (printed): \_\_\_\_\_

Sign : \_\_\_\_\_

Question	Score	Possible
1		10
2		20
3		10
4		10
5		10
6		20
7		20
Total		100

**Problem 1 (10 pts):** Please give the results of following commands (Note that Q4 has two print commands):

- 1) `rep(1:3, each=3)`
- 2) `seq(1, 10, by=2)`
- 3) `order(10:1, decreasing=F)`
- 4) `print((1:4)>2&(1:4)%2==0) ;`  
`print((1:4)>2&&(1:4)%2==0);`
- 5) `x = c(1:5, NA);`  
`print( mean(x) );`

**Problem 2 (20 pts):** Let `sex = c(1,1,1,1,1,1,2,2,2,2)`; `graduate = c(1,0,1,0,1,0,0,0,0,1)`; `score=c(9:1, NA)`. A data frame is constructed as `zz = data.frame(sex, graduate, score)`. Give the results of these R commands:

- 1) `table(zz[, "sex"])`

```
> zz
  sex graduate score
1   1         1     9
2   1         0     8
3   1         1     7
4   1         0     6
5   1         1     5
6   1         0     4
7   2         0     3
8   2         0     2
9   2         0     1
10  2         1    NA
```

- 2) `apply(zz[-1,], 2, max)`

3) `zz[zz[,3]>7,]`

4) `which.max(zz$score)`

5) `zz[order(zz["graduate"],zz["score"]),]`

6) `subset(zz, zz["sex"]==1)`

7) `tapply(zz$score, zz$graduate, mean, na.rm=T)`

8) `apply(zz[-10, ], 1, function(x){ sum(x) })`

**Problem 3 (10 pts):** Let `mylist=list(sex = c(1,1,1,1,2,2,2,2,2), smoking = c(1,0,1,0,1,0,0,0,1,1), age=c(21:30))`. Give the results of these R commands:

1) `length(mylist)`

2) `lapply(mylist, function(kk){ kk*3 })`

3) `sapply(mylist, max)`

4) `(mylist$sex-mylist$smoking)^3`

**Problem 4 (10 pts):** Define `zz=matrix(c(c(1,2,NA), c(3,4,5), 6:9), nrow = 2, ncol = 5, byrow = TRUE)`. Use `apply()` function to compute the sum of each **row** and the average of each **column(ignore any missing value)**.

**Problem 5 (10 pts):** Define `geneExpr=data.frame(gene=LETTERS[1:10], expr1=c(2.1, 4.5, 6.8, 7.9, 8.1, 5.0, 4.6, 3.2, 3.5, 7.8), expr2=c(6.1, 4.2, 2.8, 0.9, 0.1, 3.0, 2.6, 8.2, 3.4, 6.8))`. Give the results of the following R commands:

a) `counter<-0`

```
for (i in 1:length(geneExpr[,1])) {
  if (geneExpr[i,2]<=7){
    counter<-counter+1
  }
}
print(counter)
```

```
> geneExpr
  gene expr1 expr2
1    A   2.1   6.1
2    B   4.5   4.2
3    C   6.8   2.8
4    D   7.9   0.9
5    E   8.1   0.1
6    F   5.0   3.0
7    G   4.6   2.6
8    H   3.2   8.2
9    I   3.5   3.4
10   J   7.8   6.8
```

b) `myMin<-function(x, cut) { min(x[x>cut]) }`  
`apply(geneExpr[,-1],2, myMin, cut=3)`

**Problem 6 (20 pts):** Please write R functions to finish following questions:

a) Please write a function, named `sd`, to compute the standard deviation. Given parameter vector `X`, it returns the standard deviation of the vector `X`:

$$SD(X) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}.$$

It should deal with some special cases. For example, if vector X contains only one element, it should return NA. Testing code and expected outputs are shown below:

```
> sd(1)
[1] NA
> sd(1:10)
[1] 3.02765
> sd(c(1,1,1,1,1,1))
[1] 0
```

b) Pi can be computed by adding the following terms

$$\pi = 4 \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} = \frac{4}{1} - \frac{4}{3} + \frac{4}{5} - \frac{4}{7} + \frac{4}{9} - \frac{4}{11} \cdots$$

Please write a function Pi(M) to return the first M digits of the Pi.

For example, Pi(1) will return 3, Pi(2) will return 3.1,....., Pi(5) will return 3.1415



**Problem 7 (20 pts):** Please answer following questions:

- a) Please provide a regular expression for which the following sentences either match or not match

Match:

bad habit

good, because there is nothing worse than a man in kinky underwear

Badcop, it is because people want to use drugs

Good Monday Holiday

Is it really good?

Yes, it is really Good.

But NOT match:

It is not that bad or good, it is fine.

habit

because there is nothing worse than a man in kinky underwear

it is because people want to use drugs

Monday Holiday

Is it real?

- b) Please provide a regular expression to match all the following telephone numbers saved in the following variable:

```
phones <- c("2197338965", "+1 219 733 8965", " 219 733 8965", "329-293-8753 ",  
"595 794 7569", "387 287 6718", "233.398.9187 ", "482 952 3315", "Work: 579-499-  
7527", "Home:543.355.3679")
```

c) Suppose **state.name** is a vector variable storing the state names. Please write an R command to get the state names, which are ended with “a”. The command should return the state names (values) rather than the indexes.

d) Please underscore the characters, if any, that match the pattern specified by the regular expression **[Gg]eorge( [Ww].)? [Bb]ush** in the following sentences,

i bet i can spell better than you and george bush combined

I met with George W Bush yesterday

BBC reported that President George W. Bush claimed God told him to invade Iraq

a bird in the hand is worth two george bushes

Another man is named George Ww. Bush