STAT 385 Practice Midterm

STAT 385, Summer 2016	Version A	Name (Print):_____
Midterm Exam - 07/07/16		NetID:_____
Time Limit: 2 Hours		Instructor: James Balamuta

Instructions

This exam contains 4 pages (including this cover page) and 3 sections. Check to see if any pages are missing. Enter all requested information on the top of this page. The inclass section of the midterm has exam has 90 points. Do not write in the table below.

Section	Earned
True or False	/ 30
Multiple Choice	/ 20
Free Response	/ 40
Take-Home	/ 35
Total:	/ 125

The following rules apply:

- Be sure to show all your work; your partial credit might depend on it
- **No credit will be given without supporting work**
- You are not allowed to consult your notes or textbook during this test.
- You may use a non-graphing scientific and/or financial calculator.
- You may not use a cell phone or a computer as a calculator. All cell phones must be turned off.
- Anyone found copying another students' work, or cheating in any other way, will be given an "F" for the course.
- Turn in all scratch paper with your exam.

Good Luck!

Bonne chance!

Buena Suerte!

Academic Integrity

The University statement on your obligation to maintain academic integrity is:

> If you engage in an act of academic dishonesty, you become liable to severe disciplinary action. Such acts include cheating; falsification or invention of information or citation in an academic endeavor; helping or attempting to help others commit academic anfractions; plagarism; offering bribes, favors, or threats; academic interference; computer related infractions; and failure to comply with research regulations.

Rule 33 of the Code of Policies and Regulations Applying to All Students gives complete details of rules governing academic integrity for all students. You are responsible for knowing and abiding by these rules.

(30 Points) True or False with explanation

Please answer the following questions by circling either **TRUE** (**T**) or **FALSE** (**F**) and then provide justification for your answer. Each question is worth **2 points**. Answers without justification will receive 0 points.

1. **T** or **F**: The value of `sqrt(x)*sqrt(x) == x` is always `TRUE`.
2. **T** or **F**: The *only* reason to use functions is to reduce the amount of lines of code.
3. **T** or **F**: *R* operates primarily on a pass by copy paradigm regarding basic objects as they enter into functions.
4. **T** or **F**: `seq_len(num)` is equivalent in all cases to `1:num`, when `num` is an integer.
5. **T** or **F**: Everything in *R* should be considered an *object*.
6. **T** or **F**: Running the command `runif(10)` *twice* should yield equivalent results. (e.g. `all.equal(runif(10), runif(10))`)
7. **T** or **F**: If A is a matrix with dimensions $n \times m$, then A^T should also be a matrix with dimensions $n \times m$
8. **T** or **F**: The correct regex pattern for detecting `a` followed by one or more `b`'s ending with a `c` is `[ab+c]`
9. **T** or **F**: The definition of tidy data is that there are no strange symbols in the data.
10. **T** or **F**: If `x = 1:10` and `y = list(x,x)`, then the memory size of `y` is equivalent to two times the memory size of `x`.

(20 Points) Multiple Choice

Please answer the following questions by circling **one** letter. No justification is required.

1. Let `x = 5` and `y = 0` what would happen if `y != 0 && x/y`?
    a. Receive `Inf`
    b. Division by 0 Error
    c. `FALSE`
    d. `TRUE`
2. What would be the output from the code snippet?
    a. `val = 1` and `result = 10`
    b. `val = 10` and `result = 10`
    c. `val = 1` and `result = 1`
    d. `val = 10` and `result = 1`
    e. The function would error since `val` is not defined.

```
val = 1
modify_10 = function() {
  val = val * 10
  return(val)
}
result = modify_10()
```

3. If `weights` is a vector of 20 observations that contains how much everyone weighs, then the amount of people greater than 150 pounds is given by:

    a. `sum(weights[150])`
    b. `sum(weights =< 150)`
    c. `sum(weights[weights =< 150])`
    d. `sum(weights >= 150)`
    e. `sum(weights[weights >= 150])`
    f. `sum(weights[weights == 150])`

4. The least squares estimator is given by:

    a. $\hat{\beta} = \hat{\sigma^2}(X^T X)^{-1}$
    b. $E[\hat{\beta}] = \beta$
    c. $\hat{\beta} = (X^T W X)^{-1} X^T W y$ for $W \neq I_N$.
    d. $\hat{\beta} = (X^T X)^{-1} X^T y$

5. The matrix given by `mat = matrix( c(1,2,3), nrow = 2, ncol = 3)` is equal to:

    a.

```
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    1    2    3
```

`b.`

```
     [,1] [,2]
[1,]    1    1
[2,]    2    2
[3,]    3    3
```

`c.`

```
     [,1] [,2]
[1,]    1    2
[2,]    3    1
[3,]    2    3
```

`d.`

```
     [,1] [,2] [,3]
[1,]    1    3    2
[2,]    2    1    3
```

6. To read in a file with a header and separated by `,`, you should use:

    a. `read_sas()`
    b. `read_csv()`
    c. `read_fwf()`
    d. `read_dta()`

7. `git` is a:

    a. Popular song

b. Version Control
c. GitHub
d. New statistical technique

8. For a recursive function, one way to speed it up is to:

   a. Convert it to a `loop`
   b. Write it to use *external* memoization
   c. Write it to use *internal* memoization
   d. Parallelize the computation.

9. Calling `length()` on a `matrix()` returns:

   a. Number of Columns in Matrix
   b. Number of Rows in Matrix
   c. Number of Elements in Matrix
   d. Error

10. The X in $Y = X\beta + \varepsilon$ represents the:

    a. Design matrix
    b. Response vector
    c. Unobserved disturbance
    d. Coefficients

(40 Points) Free Response

1. (10 Points) Consider a function given within the following code snippet.

   a. What kind of function is this?
   b. If you were to run the function via `2*3`, what kind of output would you receive? If the output is not what a user would expect, suggest how you would fix it.

```
'*' = function(x,y) {
  cat("Multiplying",x,"and",y,"\n")
  return(x*y)
}
```

2. (15 Points) Data Structures

   a. Describe the basic building blocks of R data structures.
   b. List all of the different data structures in *R*.
   c. When should you use each data structure?

3. (15 Points) S3 Objects

   a. Describe what a generic function is and when it should be used.
   b. Consider two classes called `cat` and `dog`. Both classes store `name` and `species`. The `cat` has a random behavioral component while the `dog` class has a `true/false` property. How could you create a common class between the two?
   c. Explain what would be required in order to call the generic function `print()` to see the properties of the classes.

(35 Points) Take Home Portion

A surveyor from the field has sent you a list of coordinates given by (X, Y). Given a new pair of coordinates (x_0, y_0), you want to find the top p coordinates that are the closest to the previously established list of coordinates. The closest distance is given by euclidean distance:

$$D = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Write a set of functions that:

1. Reads in a list of coordinates

```
x,y
140.26,18.76
103.88,51.4
74.62,6.66
118.59,44.76
102.64,49.9
116.33,24.48
112.58,63.9
92.17,96.96
131.98,58.61
116.64,51.19
```

2. Calculates the distance of all points to a supplied coordinate (x_0, y_0)
3. Provides a summary output displaying the top p coordinates in addition to a summary of the distribution of differences.
4. Plots **all** points and highlights the top p coordinates.