IT 433 – Final Exam
June 10, 2013

Part A: Multiple Choice Questions about SAS. Choose the most correct answer for each question. You may give an optional reason for each answer, although if the answer is correct, the reason will not be considered. Give a reason if you are not sure of your answer. 5 points each. Do all 20 questions.

1. Which are the delimiters for a comment in SAS source code?
   a. **@@ … @@** or **#** … newline   b. *** … ;** or **/* … */**
   c. *** …** newline or **/* …** newline   d. **//** … newline or **/* … */**

2. What is the value of part_number after this script executes?
```
data info;
    input id part_number $ price;
    datalines;
45678 AZQ3456DFJE 65.43
;
```
   a. 45678   b. AZQ   c. AZQ3456D   d. AZQ3456DFJE

3. How many observations are in the dataset **aaa** after this script executes?
   Contents of input file **a.dat**:
```
Bentley Dog    8   Missy   Cat 3   Simba   Cat     4
Maxwell Mouse 8   Daisy   Cat 5   Molly   Cat     1
Ginger  Dog    5   Tucker Dog 3   Jerry   Gerbil 4
```

```
data aaa;
    infile 'c:/datasets/a.dat';
    input name $ animal_type $ age @@;
```
   a. 0   b. 1   c. 3   d. 9

4. Using the input file from Problem 3, how many observations are in the dataset bbb after this script executes:
```
data bbb;
    infile 'c:/datasets/a.dat';
    input name $ animal_type $ age @;
```
   a. 0   b. 1   c. 3   d. 9

5. What is the value of date in the dataset testdate?

```
data testdate;
    date = '30dec1959'd;
```

a. -2   b. 4   c. 6   d. 364

6. The following SAS script is submitted.

```
data empinfo;
    set employees (in=inemp) salary (in=insal);
    by empid;
    if inemp and insal;
```

If employees has 7 observations and salary has 8 observations, how many observations does empinfo have?

a. 0   b. Between 1 and 7   c. 8   d. 15

7. What are the values in the column x of the dataset testloops after this script executes?

```
data testloops;
    do i = 1 to 3;
        do j = 1 to 4;
            x = i + j;
        end;
        output;
    end;
```

a. 2 3 4 5 3 4 5 6 4 5 6 7   b. 5 6 7   c. 6 7 8   d. 8

8. What does the SAS **retain** statement do?

a. Causes an observation to be recorded in the current dataset.

b. Causes missing values to be converted to numeric values.

c. Causes values of an array to be retained in the current dataset.

d. Prevents a variable from being set to missing when the flow of control passes back to the beginning of the dataset to process the next input line.

9. Which infile statement option will specify tab delimiters?

a. **dlm='09'x**   b. **dlm='0A'x**   c. **dlm='\t'**   d. **dlm='tab'**

10. The following SAS script is submitted. The value for jobcode is **`'grade2'`**.

```
data test;
    set chemists;
    if jobcode = 'Grade2' then desc = 'Senior Chemist';
    else desc = 'Entry Level Chemist';
```

What is the value of the variable desc?

a. Entry Level Chemist   b. grade2   c. Grade2   d. Senior Chemist

11. The following SAS script is submitted:

```
proc sort data=employee;
    by descending fname;
proc sort data=salary;
    by descending fname;
data empdata;
    merge employee salary;
    by fname;
run;
```

Why does the script fail to execute?

a. The sort procs contain invalid syntax.

b. The merged datasets are not permanent SAS datasets.

c. A run statement is needed after each proc sort.

d. The datasets were not merged in the order by which they were sorted.

12. The following script fails to execute due to syntax errors:

```
data totalsales;
    set monthlysales;
    array month_sales{12};
    do i = 1 to 12;
        month_sales{i} = sales;
    end;
    keep year product sales month_sales;
```

What is the cause of the syntax error(s)?

a. An array can only be defined in a permanent SAS dataset.

b. An array cannot be referenced in a keep statement.

c. The keep statement must be placed inside the do loop.

d. The variable **`month_sales`** does not exist.

13. Which option can be added to the **`infile`** statement to insure that delimiters within quotes are treated as data?

a. **`csv`**   b. **`dlmstr`**   c. **`dsd`**   d. **`eov`**

14. What does **#3** do in a SAS data step?

    a. Moves the column pointer 3 columns to the right.

    b. Moves the column pointer to column 3.

    c. Reads data from the line 2 lines after the current line.

    d. Reads data from the line 3 lines after the current line.

15. The following SAS script is submitted:
    ```
    data snckfood.jalepeno;
       set sashelp.snacks;
       if Product = 'Jalepeno sticks';
    ```
    What statements must be submitted before this script for it to execute successfully?

    a. A libname statement must be submitted to define snckfood.

    b. A libname statement must be submitted to define sashelp.

    c. Libname statements must be submitted to define both snckfood and sashelp.

    d. No libname statements need to be submitted.

16. The following SAS script is executed:
    ```
    data lentest;
        input s $ 1-8;
    datalines;
    ab       *

    ;
    ```
    What are the values of **length(s)** and **lengthc(s)**?

    a. 2 and 2   b. 2 and 8   c. 7 and 8   d. 8 and 8

17. The SAS dataset patients has the variables patient_id, patient_name, gender, age, room_number, and doctor_id. The SAS dataset doctors has the variables doc_id, doc_name, and phone_number. Which **proc sql** statements find the names and phone numbers of the doctors of the patients in the room number 354?

    a. ```
       select doc_name, phone_number from patients
       where docid=doc_id and room_number=354;
       ```
    b. ```
       select doc_name, phone_number from patients, doctors
       where docid=doc_id and room_number=354;
       ```
    c. ```
       select doc_name, phone_number from patients
       where doctor_id=doc_id and room_number=354;
       ```
    d. ```
       select doc_name, phone_number from patients, doctors
       where doctor_id=doc_id and room_number=354;
       ```

18. For the SAS dataset patients of Problem 17, which `proc sql` statements show the average ages by gender for all the patients of Doctor Rayel?

    a. `select gender, mean(age) from patients, doctors;`
       `where doctor_id=doc_id and doc_name='Rayel'`
       `group by gender;`

    b. `select gender, mean of from patients, doctors`
       `where doctor_id=doc_id and doc_name=Rayel`
       `group by gender;`

    c. `select gender, mean(age) from patients, doctors`
       `where doctor_id=doc_id and doc_name='Rayel'`
       `group by gender;`

    d. `select gender, mean of from patients, doctors`
       `where doctor_id=doc_id and doc_name=Rayel`
       `group by gender;`

19. Using the SAS dataset patients from Problem 17, the format AGE. Is defined for age:
```
proc format;
    value AGE 0-20="Minor" 21-130="Adult";
```
Which `proc freq` statement creates this table:

```
      The FREQ Procedure
       Table of age by gender
          age        gender
```

Frequency	F	M	Total
Minor	235	335	570
Adult	516	678	1194
Total	751	1013	1764

    a. `table age gender norow nocol nopercent;`
    b. `table age gender / norow nocol nopercent;`
    c. `table age by gender / norow nocol nopercent;`
    d. `table age*gender / norow nocol nopercent;`

20. Which of the following defines a SAS MACRO variable named `&state`?

    a. `&state = "Illinois"`   b. `&LET &state = "Illinois"`
    c. `%LET state = Illinois`   d. `%LET &state = "Illinois"`

Part B: Multiple Choice Questions about R. Choose the most correct answer for each question. You may give an optional reason for each answer, although if the answer is correct, the reason will not be considered. Give a reason if you are not sure of your answer. 5 points each. Do all 20 questions.

1. What does the **?** operator do?
   a. Displays the help webpage for an R function.
   b. Returns the mode of a functions return value.
   c. Returns the mode of its operand.
   d. Returns the number of bytes of storage space of its operand.

2. What is the R symbol for integer division?
   a. `%%`   b. `%/%`   c. `/`   d. `\`

3. What is the output?
   ```
   print(apply(cbind(matrix(1, 3, 2), matrix(2, 3, 3)), 2, sum))
   ```
   a. `[1] 3 3 6 6 6`   b. `[1] 8 8 8`   c. `[1] 8 8 8 8 8`   d. `[1] 24`

4. Which of these R functions does NOT return a dataframe?
   a. `read.csv`   b. `read.fwf`   c. `readLines`   d. `read.table`

5. What is the output?
   ```
   print(system.time(rep(1, 1000))[3])
   ```
   a. `elapsed`   b. `user   system elapsed`
   `          0`   `    0        0        0`
   c. `elapsed`   d. `[1] "2013-06-10 17:45:00 CDT"`
   `    0.1223`

6. What is printed from this R script?
   ```
   print(diag(matrix(1:16, 4, 4)))
   ```
   a. `[1] 30`   b. `[1] 1   4   9   16`
   c. `[1] 1   6   11   16`   d.
   ```
           [,1] [,2] [,3] [,4]
   [1,]     1    0    0    0
   [1,]     0    4    0    0
   [1,]     0    0    9    0
   [1,]     0    0    0   16
   ```

7. What is the output?
   ```
   cat((function(x) return(sum(abs(x))))(c(-7, 3, -4)), "\n")
   ```
   a. `-8`   b. `7 3 4`   c. `14`   d. `function(x) (sum(abs(x)))`

8. What is output?
   ```
   g <- function(f) { return(f(121)) }
   print(g(sqrt))
   ```
   a. `[1] 11`   b. `Error: unexpected '{'`
   c. `sqrt(121)`   d. `Error: illegal closure`

9. If `lst` represents an R list with component names **"name", "height",** and
   "weight", which statement removes the weight component from the list:
   a. `lst <- lst$(-weight)`   b. `lst$weight <- NULL`
   c. `lst <- lst$[[-2]]`   d. `lst[[-2]]`

10. Which is the correct R designation for the complex number $\sqrt{-1}$
    a. `i`   b. `1*i`   c. `0+1i`   d. `sqrt(-1)`

11. Which of these R function calls will translate the character vector
    ```
    states <- c("mississippi", "missouri", "illinois") to
    "massassappa" "massoura" "allanoas"?
    ```
    a. `chartr('a', 'i', states)`   b. `chartr('i', 'a', states)`
    c. `chartr(states , 'a', 'i')`   d. `chartr(states, 'i', 'a')`

12. Which R functions redirect input from standard in and output to standard out, respectively.
    a. `getwd; setwd`   b. `redirectIn; redirectOut`
    c. `source; sink`   d. `STDIN; STDOUT`

13. What is the result of evaluating this expression: `0 * Inf`
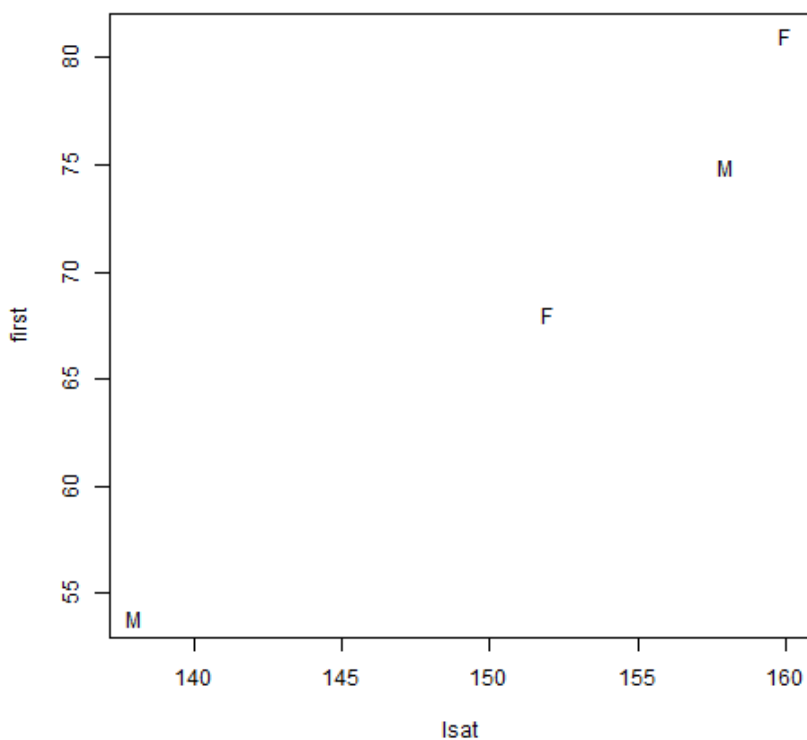   a. `Inf`   b. `NA`   c. `NaN`   d. `NULL`

14. The following dataframe named **law_stud** shows admissions data from law school students: **lsat** = Law School Admission Test score, first = First year law school average score.

```
   lsat   first gender
1 152     68      F
2 138     54      M
3 160     81      F
4 158     75      M
```

Also, symb is defined as
symb = as.character(law_stud[,3])
Which R statement produces this plot?

a. **plot(law_stud[,1], law_stud[,2], symb=symb, type="p")**
b. **plot(law_stud[,1], law_stud[,2], symb=symb, type="b")**
c. **plot(law_stud[,-3], pch=symb, type="p")**
d. **plot(law_stud[,c(1, 2)], pch=symb, type="b")**

15. Which argument in the plot function causes a plot to occur on top of the previous plot?
   a. **ask=TRUE**   b. **new=TRUE**   c. **overlay=TRUE**   d. **points=c(x, y)**

16. Which function returns the dataframe df with its ith row removed?

    a. `removeRow <- function(df, i) { return(df[-1, i]) }`

    b. `removeRow <- function(df, i) { return(-df[1, -i]) }`

    c. `removeRow <- function(df, i) { return(df[-i, ]) }`

    d. `removeRow <- function(df, i) { return(df[, -i]) }`

17. A birdwatcher observed the following birds in Illinois. The name of this dataframe is bird_data:

    ```
         species    count
    1 Bluejay        5
    2 Cardinal       8
    3 Nuthatch       3
    4 Robin         12
    5 Woodpecker     5
    ```

    Which function produces a barplot of this data?

    a. `barplot(bird_data[,1], labels=bird_data[,2])`

    b. `barplot(bird_data[,2], labels=bird_data[,1])`

    c. `barplot(bird_data[,1], names.arg=bird_data[,2])`

    d. `barplot(bird_data[,2], names.arg=bird_data[,1])`

18. Using the bird_data dataframe in Problem 17, which function call produces a pie chart of the data?

    a. `pie(bird_data[,1], labels=bird_data[,2])`

    b. `pie(bird_data[,2], labels=bird_data[,1])`

    c. `pie(bird_data[,2], names.arg=bird_data[,1])`

    d. `piechart(bird_data[,1], names.arg=bird_data[,2])`

19. Which output is the most plausible for this R statement?
    `print(mean(runif(5000, min=1, max=3)))`

    a. `[1] 0.5752651`   b. `[1] 2`   c. `[1] 2.00176`   d. `[1] 2.67654`

20. What is printed when this statement is executed?
    `print(pnorm(1) - pnorm(-1))`

    a. `[1] 0.3413447`   b. `[1] 0.6826895`

    c. `[1] 0.8413447`   d. `[1] 2`