Quiz 03 – Week 04

Problem 1.

Which of the following statements are used to read the delimited raw data file and create an SAS data set?

A. DATA and SET
B. DATA, SET and INFILE
C. DATA, SET and INPUT
D. DATA, INFILE and INPUT

Answer: D

Problem 2

Below is the data from a csv file "Emp.csv".

```
Employee id,Gender,Name,DOB,Location,Salary,ManagerEmp ID
This dataset is about company employee
101,M,John,12/1/1995,Delhi,350000,101
102,F,Sangeeta,7/4/1980,Delhi,450000,103
103,F,Mary,3/5/1973,Mumbai,500000,101
104,M,Richard,6/25/1975,Mumbai,750000,101
105,M,Fredrick,8/20/1990,Delhi,320000,101
```

And, the following code is used to read the file named EMP.

```
DATA WORK.EMP;
      INFILE 'C:\STA311\Emp.csv' DLM = ',' ;
      INPUT
            Employee_id $
            Gender $
            Name $
            DOB
            Location $
            Salary
            Manager_Emp_ID;
RUN;
```

What will be the output if we run the below SAS statements to read "emp.csv" file?

A.

	Employee_id	Gender	Name	DOB	Location	Salary	Manager_Emp_ID
1	thisdata	101	M	.	12/1/199	.	350000
2	102	F	Sangeeta	.	Delhi	450000	103
3	103	F	Mary	.	Mumbai	500000	101
4	104	M	Richard	.	Mumbai	750000	101
5	105	M	Fredrick	.	Delhi	320000	101

B.

	Employee_id	Gender	Name	DOB	Location	Salary	Manager_Emp_ID
1	102	F	Sangeeta	.	Delhi	450000	103
2	103	F	Mary	.	Mumbai	500000	101
3	104	M	Richard	.	Mumbai	750000	101
4	105	M	Fredrick	.	Delhi	320000	101

C.

	Employee_id	Gender	Name	DOB	Location	Salary	Manager_Emp_ID
1	Employee	Gender	Name	.	Location	.	.
2	This dat	101	M	.	12/1/199	.	350000
3	102	F	Sangeeta	.	Delhi	450000	103
4	103	F	Mary	.	Mumbai	500000	101
5	104	M	Richard	.	Mumbai	750000	101
6	105	M	Fredrick	.	Delhi	320000	101

D. None of the screenshots is the output from the data step.

Answer: C.

INFILE statement starts reading a file from the first line of the CSV file. It can also be header row, so we need to mention the start row explicitly.

Problem 3.

This question is based on the following data from a csv file "Emp.csv".

```
Employee id,Gender,Name,DOB,Location,Salary,ManagerEmp ID
This dataset is about company employee
101,M,John,12/1/1995,Delhi,350000,101
102,F,Sangeeta,7/4/1980,Delhi,450000,103
103,F,Mary,3/5/1973,Mumbai,500000,101
104,M,Richard,6/25/1975,Mumbai,750000,101
105,M,Fredrick,8/20/1990,Delhi,320000,101
```

Which data option will be added to INFILE statement to read the above dataset from the record with the employee name "John"? In other words, which option should be used to replace ??? in the following code?

```
DATA WORK.EMP;
      INFILE 'C:\STA311\Emp.csv' DLM = ',' ???;
      INPUT
            Employee_id $
            Gender $
            Name $
            DOB
            Location $
```

```
            Salary
            Manager_Emp_ID;
RUN;
```

A. rows=3
B. option= 3
C. firstobs=2
D. start=3
E. Start=2
F. firstobs=3

Answer: F

FIRSTOBS option can be used to explicitly mention the start row to read. In the above table, the first row is representing the header, the second row about the table, and the data set is starting with the third row.

Problem 4

Below is the csv file "class.csv" for marks of students in different subjects:

```
Name,Gender,Location,English,Maths,Hindi,Sanskrit
Mohan,M,Banglore,50,60,70,80
Ramesh,M,Banglore,45,50,65,89
John,M,Washington,68,,,88
Kathy,F,Washington,89,55,85,83
George,M,Washington,43,45,95,84
Lisa,F,Washington,76,85,,86
Venkat,M,Banglore,68,90,78,92
Srimohan,M,Banglore,59,56,80
Preet,F,Banglore,81,95,85,96
Lindsy,F,Washington,66,75,78,82
```

The following code is used to read the file class.csv into a SAS dataset table named **class.**

```
DATA WORK.class;
     INFILE 'C:\MAT311\ClassScore.csv' DML =',' FIRSTOBS = 2;
     INPUT Name $
           Gender $
           Location $
           English
           Maths
           Hindi
           Sanskrit;
RUN;
```

Above code gives the below output:

	Name	Gender	Location	English	Maths	Hindi	Sanskrit
1	Mohan	M	Banglore	50	60	70	80
2	Ramesh	M	Banglore	45	50	65	89
3	John	M	Washingt	68	88	.	.
4	George	M	Washingt	43	45	95	84
5	Lisa	F	Washingt	76	85	86	.
6	Srimohan	M	Banglore	59	56	80	.
7	Lindsy	F	Washingt	66	75	78	82

In the above output, you can see the following issues:

(1). Marks are not under the right column heading like John marks in ”Sanskrit” has shifted to “Maths”

(2). The total number of observations is 7 only.

Which of the following command can be used with “INFILE” statement to remove these errors?

A. MISSING
B. MISSOVER
C. DSD
D. Both A and C
E. Both B and C

Answer: E

Whenever a read a delimited file using INFILE statement and if the file has two or more delimiter together (n value between them) or the last column data is missing then it takes the next possible value as an input for that column. And, the next possible value can be other column data of the same row or next line also.

Now, to avoid these reading issues, we use DSD to prevent reading from the next column of the same row and MISSOVER for the next line or observation.

Problem 5

Which statement is true regarding PROC IMPORT?

A. By default, PROC IMPORT overwrites an existing SAS data set.

B. PROC IMPORT writes SAS data to a CSV file.

C. The DBMS= option identifies the type of data to import.

D. Dates are imported as character values.

Answer: C

Problem 6

Which of the following statements are used to read an SAS data set and create an NEW SAS data set?

A. DATA and SET
B. DATA, SET, and INFILE
C. DATA, SET, and INPUT
D. DATA, INFILE, and INPUT

Answer: A

Problem 7.

Consider the following data in csv format. You can copy the data block to Notepad or similar text editor and save it as csv format in your local drive with filename **employeecsv**

```
Employee id,Gender,Name,DOB,Location,Salary,ManagerEmp ID
This dataset is about company employee
101,M,John,12/1/1995,Delhi,350000,101
102,F,Sangeeta,7/4/1980,Delhi,450000,103
103,F,Mary,3/5/1973,Mumbai,500000,101
104,M,Richard,6/25/1975,Mumbai,750000,101
105,M,Fredrick,8/20/1990,Delhi,320000,101
```

We use the following PROC IMPORT to load this csv data to SAS

```
PROC IMPORT OUT= WORK.employ_csv
            DATAFILE= "C:\STA311\employeecsv.csv"
            DBMS=CSV    /* identifier of database management system */
            REPLACE;    /* (1). end of the long statement; (2). replace
                           existing SAS data set with same name
                           employ_scv before create the new sas
                           data set                                  */
            GETNAMES=YES;  /* tell SAS whether the variable names in
                              the csv file */

            DATAROW=2;     /* the row number of the first record to be
                              read in SAS. This allows you jump to
                           certain row to read records. it is similar
                           to FIRSTOBS = in INFILE-INPUT statements
                           in the DATA STEP.                         */
RUN;
```

The above code generated an SAS data set in the following

Obs	Employee_id	Gender	Name	DOB	Location	Salary	ManagerEmp_ID
1	This dataset is about company employee			.		.	.
2	101	M	John	12/01/1995	Delhi	350000	101
3	102	F	Sangeeta	07/04/1980	Delhi	450000	103
4	103	F	Mary	03/05/1973	Mumbai	500000	101
5	104	M	Richard	06/25/1975	Mumbai	750000	101
6	105	M	Fredrick	08/20/1990	Delhi	320000	101

The first row is not correct. Which of the following modification of statements will fix that problem.

A. DATAROW = 2; DBMS = TAB
B. DATAROW = 3; DBMS= TAB
C. DATAROW = 3; DBMS = CSV
D. DATAROW = 3, DBMS = EXCEL
E. DATAROW=2; DMBS = CSV

Answer: C.

Problem 8

Which of the following Data Step DOES NOT create the following output (hint: if necessary, you should run the program to check your choice):

Obs	Employee_id	Gender	Name	Location	Salary	ManagerEmp_ID
1	101	M	John	Delhi	350000	101
2	102	F	Sangeeta	Delhi	450000	103
3	103	F	Mary	Mumbai	500000	101
4	104	M	Richard	Mumbai	750000	101
5	105	M	Fredrick	Delhi	320000	101

A

```
DATA EMPOLYEE01;
INFILE DATALINES;
INPUT Employee_id $
      Gender $
      Name $
      Location $
      Salary : comma7.2.
      ManagerEmp_ID;
DATALINES;
101 M John Delhi 350,000 101
102 F Sangeeta Delhi 450,000 103
103 F Mary Mumbai 500,000 101
```

```
104 M Richard Mumbai 750,000 101
105 M Fredrick Delhi 320,000 101
;
RUN;

PROC PRINT DATA = EMPOLYEE01;
RUN;
```

B

```
DATA EMPOLYEE01;
INFILE DATALINES;
INPUT Employee_id $
      Gender $
      Name $
      Location $
      Salary comma7.2.
      ManagerEmp_ID;
DATALINES;
101 M John Delhi 350,000 101
102 F Sangeeta Delhi 450,000 103
103 F Mary Mumbai 500,000 101
104 M Richard Mumbai 750,000 101
105 M Fredrick Delhi 320,000 101
;
RUN;

PROC PRINT DATA = EMPOLYEE01;
RUN;
```

C

```
DATA EMPOLYEE01;
INFILE DATALINES;
INPUT Employee_id $
      Gender $
      Name $
      Location $
      Salary comma.
      ManagerEmp_ID;
DATALINES;
101 M John Delhi 350,000 101
102 F Sangeeta Delhi 450,000 103
103 F Mary Mumbai 500,000 101
104 M Richard Mumbai 750,000 101
105 M Fredrick Delhi 320,000 101
;
RUN;

PROC PRINT DATA = EMPOLYEE01;
RUN;
```

D

```
DATA EMPOLYEE01;
INFILE DATALINES;
INPUT Employee_id $
      Gender $
      Name $
      Location $
      Salary : comma.
      ManagerEmp_ID;
DATALINES;
101 M John Delhi 350,000 101
102 F Sangeeta Delhi 450,000 103
103 F Mary Mumbai 500,000 101
104 M Richard Mumbai 750,000 101
105 M Fredrick Delhi 320,000 101
;
RUN;

PROC PRINT DATA = EMPOLYEE01;
RUN;
```

E

None of the program can create the output data set.

Answer: D

This is the applications colon (:) modifier that input variable has general informat. This this particular data, the general informat is comma (,) – the value of the numeric variable in each record was separated by a comma. In the future week, we will introduce some special numeric informat.

Problem 9.

Which of the provided SAS data step produces the following output?

Obs	name	year
1	George Washington	1789
2	John Adams	1797
3	Thomas Jefferson	1801

A

```
DATA BirthDay;
LENGTH name $ 25;
INPUT name $ year;
```

```
DATALINES;
George Washington  1789
John Adams  1797
Thomas Jefferson  1801
;
RUN;

PROC PRINT DATA = BirthDay;
RUN;
```

B

```
DATA BirthDay;
LENGTH name $ 25;
INPUT name & $ year;
DATALINES;
George Washington  1789
John Adams  1797
Thomas Jefferson  1801
;
RUN;

PROC PRINT DATA = BirthDay;
RUN;
```

C

```
DATA BirthDay;
LENGTH name $ 25;
INPUT name & $ year;
DATALINES;
George Washington 1789
John Adams 1797
Thomas Jefferson 1801
;
RUN;

PROC PRINT DATA = BirthDay;
RUN;
```

D

```
DATA BirthDay;
LENGTH name $ 25;
INPUT name & $ year;
DATALINES;
George  Washington 1789
John  Adams 1797
Thomas  Jefferson 1801
;
RUN;
```

```
PROC PRINT DATA = BirthDay;
RUN;
```

E

None of the provided code can produce the output data.

Answer: B.

The ampersand (&) modifier tells SAS to use **TWO** white space characters **to signal the end of a character variable**, allowing embedded blanks to be read using list input.

Problem 10

Assume that an SAS data set **gradebook.sas7bdat** in the temporary library has three variables: name, midterm, and final. [hint: You are encouraged to create an SAS data set with 2 records and place it in the temporary library with name **gradebook.sas7bdat.** You can then test the SAS data steps to find the correct answer]

Which of the following code reads in SAS data **gradebook.sas7bdat** correctly?

A

```
PROC IMPORT DATAFILE=GradeBook
            OUT = newGradeBook
            DBMS = SAS replace;
            GETNAMES = YES;
RUN;

PROC PRINT DATA = newGradeBook;
RUN;
```

B

```
DATA newGradeBook;
INFILE gradebook;
INPUT name $ midterm final;
RUN;

PROC PRINT DATA = newGradeBook;
RUN;
```

C

```
DATA newGradeBook;
SET gradebook;
RUN;

PROC PRINT DATA = newGradeBook;
RUN;
```

D

```
DATA newGradeBook;
INFILE DATALINES DSD MISSOVER;
SET gradebook;
RUN;

PROC PRINT DATA = newGradeBook;
RUN;
```

E

None of the provided code will correctly read in the SAS data set.

Answer: C