Midterm One

Instructions: This is an in-class and open book test, but internet surfing is not allowed except the class web page. Give concise but detailed answers for full credit.

Question One [16 points]

A group of patients were followed up and their data were collected. The data items collected are:

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Label</th>
<th>Starting Columns</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>Patient ID</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>DOB</td>
<td>Date of Birth</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>ADMIT</td>
<td>Admission Date</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>DISCHRG</td>
<td>Discharge Date</td>
<td>24</td>
<td>10</td>
</tr>
<tr>
<td>DX</td>
<td>Diagnosis Code</td>
<td>34</td>
<td>1</td>
</tr>
<tr>
<td>FEE</td>
<td>Total Fee</td>
<td>35</td>
<td>5</td>
</tr>
</tbody>
</table>

The data are stored in hospital.txt in the class page: www.sci.csueastbay.edu/~sfan. Use this data set to answer the following questions.

(a) [4 points] Use the command “INFILE” to read in this data including the labels of variables. You need to use the SAS date and dollar formats for date and fee variables. Print out your data set to verify your code. Write your SAS code and the outputs of the columns of ID, DOB and FEE here:

```
DATA HOSPITAL;
INFILE 'C:\st6250\datafiles\hospital.txt';
INPUT @1 ID $3.
@4 DOB MMDDYY10.
@14 ADMIT MMDDYY10.
@24 DISCHRG MMDDYY10.
@34 DX 1.
@35 FEE 5.;
label DOB='date of birth';
label ADMIT='admission date';
label DISCHRG='discharge date';
label DX='diagnosis code';
FORMAT DOB ADMIT DISCHRG MMDDYY10.;
format fee dollar7.;
proc print data=hospital;
run;
```
(b) [4 points] Use the command “Set” to create a new data set “hosptial1” including all variables and the following three new variables:
   a. Stay_days, the number of days the patients stayed in the hospital
   b. Long_stay, the indicator function of patients staying for more than 14 days, i.e. Long_stay = 1 if the patient stayed in the hospital for more than 14 days and zero otherwise
   c. Age (in days), the age of the patient when admitted

Print out your data set to verify your code. Write your SAS code and the outputs of the columns of ID, Stay_days, Long_stay, and Age here:

```sas
data hospital1;
set hospital;
STAY_days = DISCHRG-ADMIT + 1;
  if stay_days > 14 then long_stay=1;
  else long_stay=0;
*age = intck('day',DOB,ADMIT);
  age=ADMIT-DOB+1;
run;

proc print data=hospital1;
run;
```

<table>
<thead>
<tr>
<th>Obs</th>
<th>ID</th>
<th>DOB</th>
<th>ADMIT</th>
<th>DISCHRG</th>
<th>DX</th>
<th>FEE</th>
<th>STAY_days</th>
<th>Long_stay</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>001</td>
<td>10/21/1946</td>
<td>12/12/2004</td>
<td>12/14/2004</td>
<td>8</td>
<td>$8,000</td>
<td>3</td>
<td>0</td>
<td>21238</td>
</tr>
<tr>
<td>2</td>
<td>002</td>
<td>05/01/1980</td>
<td>07/08/2004</td>
<td>08/08/2004</td>
<td>4</td>
<td>$12,000</td>
<td>32</td>
<td>1</td>
<td>8835</td>
</tr>
<tr>
<td>3</td>
<td>003</td>
<td>01/01/1960</td>
<td>01/01/2004</td>
<td>01/04/2004</td>
<td>3</td>
<td>$9,000</td>
<td>4</td>
<td>0</td>
<td>16072</td>
</tr>
</tbody>
</table>
(c) [2 points] Write a SAS code to PRINT only the data of those patients staying the hospital for at least a week or spending at least $10,000. Write your SAS code and the outputs of the columns of ID, Stay_days and Fee here:

```sas
proc print data=hospital1;
where stay_days >= 7 or fee >= 10000;
run;
```

<table>
<thead>
<tr>
<th>Obs</th>
<th>ID</th>
<th>DOB</th>
<th>ADMIT</th>
<th>DISCHRG</th>
<th>DX</th>
<th>FEE</th>
<th>days</th>
<th>stay_days</th>
<th>long_age</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>002</td>
<td>05/01/1980</td>
<td>07/08/2004</td>
<td>08/08/2004</td>
<td>4</td>
<td>$12,000</td>
<td>32</td>
<td>1</td>
<td>8835</td>
</tr>
</tbody>
</table>

(d) [3 points] Use SAS to draw a graph to study the relationship between the length of stay (Stay_days) and age when admitted (Age). From the graph, do you see a positive or negative association between them? Write your SAS code and sketch the graph here.

```sas
proc gplot data=hospital1;
title 'the scatterplot of length of stay vs. age';
plot stay_days* age;
run;
```

It shows a negative association.
(e) [3 points] Write a SAS code to sort the data ‘hospital1’ by Long_stay first and then by Fee, and print ONLY the columns of ID, Long_stay and Fee in the sorted data file. Write your SAS code and outputs here:

```sas
proc sort data=hospital1;
by long_stay fee;
run;

proc print data=hospital1 noobs;
var ID long_stay fee;
run;
```

<table>
<thead>
<tr>
<th>ID</th>
<th>Long_stay</th>
<th>Fee</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>0</td>
<td>$8,000</td>
</tr>
<tr>
<td>003</td>
<td>0</td>
<td>$9,000</td>
</tr>
<tr>
<td>002</td>
<td>1</td>
<td>$12,000</td>
</tr>
<tr>
<td>004</td>
<td>1</td>
<td>$15,123</td>
</tr>
</tbody>
</table>

**Question Two** [4 points]

Write a SAS code to create a permanent format “age” in a library of your choice such that those with age 0 to < 21 are ‘under 21’; those with 21 to < 60 are ‘middle age’; those with age ≥ 60 are ‘seniors’; and those with unknown age are ‘missing’.

What statements do you need to make in order to use these formats in a new SAS session? What is the statement to see the contents of these formats?

```sas
libname myfmts 'C:\Documents and Settings\kfan\Desktop\academic\teach\CSU\st6250\sas_data';
proc format libname=myfmts;
value age
    0-<21 = 'under 21'
    21-<60 = 'middle age'
    60-high = 'seniors'
    other = 'missing';
run;

** in a new SAS session;**
libname myfmts 'C:\Documents and Settings\kfan\Desktop\academic\teach\CSU\st6250\sas_data';
options fmtsearch=(myfmts);
proc format library=myfmts fmtlib;
run;
```