Midterm Three

Instructions: This is an in-class and open book midterm. You must write your answers on the provided spaces. Give concise but detailed answers for full credit.

**Question One** (8 points)

Consider the two data sets:

```sas
data employee;
  length ID $ 3 Name $ 12;
  input ID Name;
datalines;
  1 Smith
  2 Schneider
  4 Gregory
;
data hours;
  length ID $ 3 JobClass $ 1;
  input ID
    JobClass
    Hours;
datalines;
  1 A 39
  4 B 44
  9 B 57
;
```

Write SAS codes for the following parts.

(a) [3 points] Use PROC SQL to merge these two data sets by ID and call it combine.

(b) [1 point] Use PROC PRINT to print only the first two observations of combine.

```sas
*(a);
proc sql;
create table combine as
select e.id as e_id, h.id as h_id, name, jobclass, hours
from employee as e full join
  hours as h
on h.id eq e.id;
quit;

*(b);
proc print data=combine(obs=2);
run;
```
(c) [2 points] Use ODS to find the name of the output objective “summary statistics” of the code:

```sas
proc means data=combine;
  var hours;
run;
ods trace on;
proc means data=combine;
  var hours;
run;
ods trace off;

Name is Summary as shown in the log window.
```

(d) [2 points] Use ODS to save the output objective “summary statistics” as a new SAS dataset mean.

```sas
ods output summary=mean;
proc means data=combine;
  var hours;
run;
```

**Question Two (4 points)**

The data set scholarship contains three variables: gender (female, male), scholarship (yes, no) and GPA (student’s current GPA). Analyze the following SAS code and draw the output table WITHOUT RUNNING THE SAS CODE. Indicate what the cell values should be.

```sas
proc tabulate data=scholarship;
title "proc tabulate example";
class gender scholarship;
var GPA;
table gender*scholarship, GPA*(n mean min max);
keylabel
  n='number'
  mean='average'
  min='minimum'
  max='maximum';
run;
```

<table>
<thead>
<tr>
<th>Gender</th>
<th>Scholarship</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>number</td>
</tr>
<tr>
<td>F</td>
<td>N</td>
<td>46.00</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>10.00</td>
</tr>
<tr>
<td>M</td>
<td>N</td>
<td>35.00</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>3.00</td>
</tr>
</tbody>
</table>

An example of cell value: The first column –last row cell is the # of male scholarship recipients.
**Question Three** (8 points)

Run the following data set to create a SAS dataset called `scores`, which contains four test scores for each student in the class (plus an ID number and his/her gender).

```sas
Data scores;
input id $ gender $ score1-score4;
datalines;
001 F 50 70 62 78
002 F 70 89 76 83
003 M 40 50 98 70
004 M 90 78 86 76;
```

Use this data set to answer the following parts.

(a) [1 points] Use SET to add the mean score of four tests (called AVE) into the dataset `scores`.

```sas
   data scores;
   set scores;
   ave=mean(of score1-score4);
   run;
```

(b) [4 points] Use PROC REPORT to create the following report:

The CSUEB Statistics Department

The mean of four tests for each student
---------------------------------------

<table>
<thead>
<tr>
<th>Student ID</th>
<th>Mean Score</th>
<th>Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>65.0</td>
<td>no</td>
</tr>
<tr>
<td>002</td>
<td>79.5</td>
<td>yes</td>
</tr>
<tr>
<td>003</td>
<td>64.5</td>
<td>no</td>
</tr>
<tr>
<td>004</td>
<td>82.5</td>
<td>yes</td>
</tr>
</tbody>
</table>

Make sure you do the following things:
1) Use OPTIONS to align your titles to the left
2) Use FORMAT to print the first digit after decimal point for AVE
3) Use COMPUTE inside PROC REPORT to create a variable PASS which is ‘yes’ if the mean score is at least 70 and ‘no’ otherwise.

```sas
   options nocenter;
   title1 "The CSUEB Statistics Department";
   title3 "The mean of four tests for each student";
   title4 "---------------------------------------";
   proc report data=scores nowd;
   column id ave pass;
   define id/ display "Student ID" width=7;
   define ave/ display "mean score" width=7 format=7.1;
   define pass/ computed "mean >= 70?" width=10;
   compute pass/ character length=10;
   if ave ge 70 then pass='yes';
   else pass='no';
   endcomp;
   run;
```
(c) [3 points] To see the relationship between test 1 scores and test 4 scores, draw a scatterplot as follows:

```
proc sort data=scores out=scores1;
by score1;
run;

goptions reset=all
   colors=(black);
title "scatterplot of test 4 vs. test 1";
symbol value=dot interpol=join width=2;
proc gplot data=scores1;
plot score4*score1;
run;
quit;
```