STATISTICS 479
Exam I (100 points)

Questions 1 through 12 are worth 2 points each.

1. To create a SAS data set from (external) raw data, an \texttt{input} \underline{statement} is necessary in the data step. (Not the datalines; statements.)

2. When raw data is available in a file (say, as a text file), an \texttt{infil\underline{e}} statement in the SAS program tells SAS location of that file and the name of the file.

3. To create a new SAS data set using observations from an existing SAS data set, a \underline{data} statement must be followed by a \texttt{set} statement.

4. Complete the following data option needed to access a SAS data set named scores available in a folder named in a libname with the libref \texttt{stat479}. \underline{data=\texttt{stat479.scores}}

5. Give a SAS format to print a numeric value using a maximum of 12 print positions that includes 4 decimal places: \underline{12.4}

6. Give the symbols used on the input statement to move the pointer to the 13th column: \underline{@@13}

7. Give a symbol used at the end of an input statement to hold the current data line in order to read more data from the same line: \underline{@@ or @}

8. Write a statement in a SAS data step to indicate that variables Height and Height are to be omitted the SAS data set created? \underline{drop \texttt{Weight Height;}}

9. The \underline{output} statement in a SAS data step tells SAS to write the current observation from the PDV to the SAS data set at that point in the data step.

10. To use a do loop in a data step to compute the square root of several numeric variables, these variables must first be declared in an \underline{array} statement.

11. Write a statement to tell SAS not to write an observation with a missing value for the numeric variable Income, to the data set: \underline{if income = . then delete;}

12. Use SAS statements (using a do-end loop) to set the variables Mileage equal to 'Good' and Style equal to 'Sedan', if the variable Mpg has a value greater than 35.

\begin{verbatim}
if Mpg > 35 then do;
Mileage = 'Good';
Style = 'Sedan';
end;
\end{verbatim}
13. Consider the following SAS program and answer questions below (3 points per each part). (Note: The line numbers are for reference only).

```
data wtclub;
input Idno Sex $ Team $ Strut Wtnow ;
datalines;
1023 M red 189 165
1039 F yellow 145 124
1219 M red 210 192
1248 M yellow 194 177
1078 F red 127 118
1221 M yellow 220 200
1157 M green 155 141
1331 M blue 187 172
1067 F green 135 122
1261 M blue 181 166
1339 F green 141 129
1192 F yellow 152 139
1352 M green 156 137
1252 M blue 196 180
1067 F red 148 138
1124 M green 158 142
1197 F red 138 128
1133 F blue 130 117
1036 F green 135 123
1087 M yellow 146 132
1328 M red 155 142
;
proc means;
class Team Sax;
var Strut Wtnow;
output out=wttstats mean=Mean1 Mean2 std=SE1 SE2;
run;
proc print data=wttstats;
run;
```

(a) If the datalines were available in the text file `weight.txt` in the folder `c:sarah\wtclub`, write a statement to add to the above program in order to read the data from this file (instead of including the data instream):

```
infile 'c:sarah\wtclub\weight.txt';
```

(b) Is the SAS dataset named `wtclub` temporary or permanent? How many variables and observations are in this data set?

Temporary 5 variables 21 observations

(c) Show a sketch of the output from the `proc means` step (no numbers required).

<table>
<thead>
<tr>
<th>Team</th>
<th>Sax</th>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>StdDev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>blue</td>
<td>F</td>
<td>Strut</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>Wtnow</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>green</td>
<td>F</td>
<td>Strut</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>Wtnow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>red</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yellow</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(d) Is the SAS data set named wtstats temporary or permanent? How many observations are in this dataset? temporary 15

(e) Following is a line extracted from the output of proc print. Describe the statistics printed under last 4 variable names: Mean1, Mean2, SE1, SE2

<table>
<thead>
<tr>
<th>Obs</th>
<th>Team</th>
<th>Sex</th>
<th><em>TYPE</em></th>
<th><em>FREQ</em></th>
<th>Mean1</th>
<th>Mean2</th>
<th>SE1</th>
<th>SE2</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>red</td>
<td>2</td>
<td>6</td>
<td>61.167</td>
<td>46.167</td>
<td>31.858</td>
<td>27.7086</td>
<td></td>
</tr>
</tbody>
</table>

For all members of the red team:

Average of the variables & Standard deviations of the variables &
Statist & Winnow & Statist & Winnow

14. A SAS data set is to be created using the following input statement:

```
input State $ City $ Pop1994 Income Housing Electric;
```

Write SAS program statements (with correct syntax) to be included in the data step to accomplish the tasks in (a) to (d) and answer (e) to (g). (3 points each part)

(a) Eliminate observations with Pop1994 values less than or equal to 50,000 from the data set.

```
if Pop1994 <= 50000 then delete;
```

(b) To be able to store 12 characters for variable City.

```
length City $ 12;
```

(c) Form a new category variable named Housgrp that will have values of 1, 2 or 3 depending on whether Housing values are less than or equal to 25,000, over 25,000 but less than or equal to 45,000, or over 45,000, respectively.

```
if Housing <= 25000 then Housgrp = 1;
else if 25000 < Housing <= 45000 then Housgrp = 2;
else Housgrp = 3;
```

(d) To create a new numeric variable Percent containing values for Housing as a percentage (not a fraction) of Income.

```
Percent = 100 * Housing / Income;
```

(e) To specify appropriate printing formats for the money values Income, Housing, and Electric.

```
format Income Housing Electric dollar12.2;
```

(f) Give the name of a SAS procedure that may be used to examine the distribution of numeric variables such as Housing or Income. Name a statistic or a plot printed by this procedure useful for studying the shape of the distribution.

```
Proc Univariate Box Plot, Normal Prob Plot
```

(g) What SAS procedure would you use to create a SAS data set containing sample statistics for numeric variables Housing and Electric for subgroups of City or Housgrp? Name the statement used to create the SAS data set in this proc step.

```
Means, Output
```
15. Ms. Anderson wants to use a SAS program to compute the total score, assign letter grades, and compute summary statistics for her college Stat 101 class. A subset of the data looks like:

<table>
<thead>
<tr>
<th>Id</th>
<th>Major</th>
<th>Year</th>
<th>Quiz</th>
<th>Exam1</th>
<th>Exam2</th>
<th>Lab</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>5109</td>
<td>Psych</td>
<td>3</td>
<td>75</td>
<td>78</td>
<td>90</td>
<td>87</td>
<td>87</td>
</tr>
<tr>
<td>7391</td>
<td>Stat</td>
<td>1</td>
<td>87</td>
<td>75</td>
<td>80</td>
<td>85</td>
<td>75</td>
</tr>
<tr>
<td>2962</td>
<td>Econ</td>
<td>2</td>
<td>93</td>
<td>68</td>
<td>60</td>
<td>75</td>
<td>93</td>
</tr>
</tbody>
</table>

Note: Year records the year in school. Quiz and Lab are the totals for 10 quizzes and 10 labs, respectively.

(a) Write SAS statements to create a SAS dataset named stat101. Name your variables same as given above and assume that data will be included instream with data values entered delimited by blanks. (5 points)

```sas
data stat101;
input Id Major Year Quiz Exam1 Exam2 Lab Final;
datalines;
5109 Psych 3 75 78 90 87 87
```

(b) A maximum of 100 points each could be earned for the quizzes, each midterm exam, the labs, and the final exam. Write a SAS statement to be added to the above data step to include a new variable Total containing values of the course percentage, calculated weighting the points obtained for the quizzes by 10%, each of the two midterms and the labs by 20%, and the final by 30%. (5 points)

```
Total = .1*Quiz + .2*(Exam1 + Exam2 + Lab) + .3*Final;
```

(c) Write SAS statement(s) to be added to the above data step to create a new variable Grade containing letter grades A, B, C, D, and F, using 90, 80, 70, 60 percent cutoffs, respectively. You may use the variable Total from part b) in your statements. (5 points)

```
if Total >= 90 then Grade = 'A';
else if 80 <= Total < 90 then Grade = 'B';
else if 70 <= Total < 80 then Grade = 'C';
else if 60 <= Total < 70 then Grade = 'D';
else Grade = 'F';
```
(d) Ms. Anderson also would like to use a SAS procedure to compute (and print) summary statistics (mean, standard deviation, maximum, minimum, standard error of the mean) of the scores for the quizzes, labs, exams, the final and the course percentage, separately for each major. Write SAS statements to do this. (5 points)

```
proc means data=stat101 mean std min max stderr;
  var Quiz Exam1 Exam2 Lab Final Total;
class Major;
run;
```

16. Display the printed output resulting from executing the following SAS program. (6 points)

```
data test;
  input Score @@;
  if Score > 60 then do;
    Grade = 'P';
    Extra = (Score - 60)*10;
  end;
  else if Score < 60 then do;
    Grade = 'F';
  end;
datalines;
57 59 60 62 65.
;
proc print data=test; run;
```

<table>
<thead>
<tr>
<th>Score</th>
<th>Grade</th>
<th>Extra</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>F</td>
<td>.</td>
</tr>
<tr>
<td>59</td>
<td>F</td>
<td>.</td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>P</td>
<td>20</td>
</tr>
<tr>
<td>65</td>
<td>P</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td></td>
</tr>
</tbody>
</table>
17. Display the printed output produced by executing the following SAS program. Show what is in the program data vector at the point the first observation is written to the SAS data set. (7 points)

data one;
input X1-X3 @@;
X3=X1**2-3*X2;
X4=sqrt(X2)+1;
drop X1 X2;
datalines;
3 4 5 6 9 3 -2 1 4 16 8;
;
proc print data=one;
run;

<p>|PDV|
|---|---|---|---|---|---|---|</p>
<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th><em>N</em></th>
<th><em>Error</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>-3</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

18. Display the printed output produced by executing the following SAS program. Show what is in the program data vector immediately after processing the first line of data. (7 points)

data rating;
input Team1-Team5;
array Tm99{5} Team1-Team5;
do I=1 to 5;
   if Tm99{I}=. then Tm99{I}=0;
   Tm99{I}= Tm99{I}**2;
end;
drop I; datalines;
4 6 0 -1 .
3 -2 8 -9 12
5 . -4 7 6
7 5 10 4 5
;
proc print data=rating;
run;

<p>|PDV|
|---|---|---|---|---|---|---|---|</p>
<table>
<thead>
<tr>
<th></th>
<th>Team1</th>
<th>Team2</th>
<th>Team3</th>
<th>Team4</th>
<th>Team5</th>
<th><em>N</em></th>
<th><em>Error</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16</td>
<td>36</td>
<td>0</td>
<td>1</td>
<td>49</td>
<td>36</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>4</td>
<td>64</td>
<td>81</td>
<td>144</td>
<td>16</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Team1</th>
<th>Team2</th>
<th>Team3</th>
<th>Team4</th>
<th>Team5</th>
<th>I</th>
<th><em>N</em></th>
<th><em>Error</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16</td>
<td>36</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>