Questions 1 through 12 are worth 2 points each.

1. The SAS statement that must follow a data statement that describes how to read data values from a data line when using raw data to create a SAS data set is __input_.

2. Give a SAS format to print a numeric value using a maximum of 10 print positions that includes 3 decimal places: __10.3__

3. If you want to take square roots of 20 variables using a do loop, the 20 variables must first be assigned a single name in an __array__ statement.

4. Write a statement in a SAS data step to indicate that variables Sales and Percent are not to be included in the SAS data set created? __drop Sales Percent;__

5. To create a new SAS data set seniors using observations from an existing SAS data set named bio101, a data seniors; statement must be followed by the __set bio101;__ statement.

6. Name two SAS statements that can be used to specify or change new variable attributes (i.e., things like formats and labels) in both a data step or a proc step: __format; label__

7. To process subsets of observations in a data set by a SAS procedure, use a __by__ statement in the proc step.

8. Give the symbols used on the input statement to move the pointer to the 18th column: __lobs__

9. The __output__ statement allows the user to write an observation to the SAS data set being created at the current point in the data step.

10. Describe the number stored in the variable DoB when the date 08/24/98 is input to DoB using mmdy8.? __#of days from 1/1/60 to 8/24/98__

11. Suppose a SAS data set named biology was previously created and saved in some folder (e.g., your own folder in the C: drive). In a SAS program, this folder is referenced in a libname statement as mylib. Complete the following option needed to obtain a listing of this dataset: __proc print data=mylib.biology__

12. Write statements in SAS code (that involve a do-end loop) to increase the value of the variable Score by 50% and subtract 20 from the value of the variable Rating if the value of the variable Level is equal to “High”.

   ```sas
   if Level="High" then do;
   Score = Score * 1.5;
   Rating = Rating - 20;
   end;
   ```
13. Study the following SAS program and answer the questions given below (4 points per each part). Note that the line numbers are for reference only.

```sas
data jrhgih;
input Name $ Sex $ Height Weight School $;
datalines:
Alfred M 69.0 122.5 BJH
Alicia F 56.5 84.0 BJH
Bencia F 65.3 98.0 BJH
Bennett F 63.2 96.2 A1H
Carol P 62.8 102.5 BJH
Carlos M 63.7 102.9 A1H
Hanny M 63.5 121.5 A1H
Jaime F 57.3 86.0 BJH
Janet F 59.8 84.5 A1H
Jean M 68.2 113.4 A1H
Joyce M 51.3 50.5 BJH
Luc M 66.3 77.0 A1H
Marie F 66.5 112.0 BJH
Medford M 64.9 114.0 A1H
Philip M 69.0 115.0 A1H
Robert M 64.8 128.0 BJH
Thomas M 57.5 85.0 A1H
Wakana F 61.3 99.0 A1H
William M 66.5 112.0 BJH;
run;
```

(a) If you know that the data were available in the text file `school.txt` in the folder `c:\statdept\project`, write a statement needed in the program in order to read the data from this file (instead of including the data in-stream).

```sas
infile "c:\statdept\project\school.txt" ;
```

(b) Is the SAS dataset named `jrhgih` temporary or permanent? How many variables and observations are in this data set? Temporary | 5 | 19

(c) Show a sketch of the output from `proc means` in line 6. (Actual values are not needed; but show what statistics are computed, where they are printed, and for which variables)

- School: AJH, BJH
- Sex: M, F
- Variable: Height, Weight
- N: x, x
- Mean: x, x
- Std Dev: x, x
- Min: x, x
- Max: x, x

(d) Is the SAS data set named `stats1` temporary or permanent? How many observations are in this data set? Temporary | 2
(e) Following is a line extracted from the output of proc print. Describe what the numbers printed under the 4 variable names: Mean_Ht, Mean_Wt, SD_Ht, SD_Wt represent.

<table>
<thead>
<tr>
<th>Obs</th>
<th>School</th>
<th>Sex</th>
<th><em>TYPE</em></th>
<th><em>FREQ</em></th>
<th>Mean_Ht</th>
<th>Mean_Wt</th>
<th>SD_Ht</th>
<th>SD_Wt</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>BJH</td>
<td>2</td>
<td>8</td>
<td></td>
<td>61.3750</td>
<td>96.625</td>
<td>5.65149</td>
<td>23.5838</td>
</tr>
</tbody>
</table>

\[\text{Average g. Height for Male} \quad \text{for Female} \quad \text{in school BJH} \quad \text{stand. dev. for Male} \quad \text{and Female} \quad \text{in BJH}\]

14. Show the values for the variable Time that will be stored in the SAS dataset reaction for each of the 5 observations (5 points):

```
data reaction;
  input Time 5.2;
  datalines;
  743
  743
  743
  7.43
  7.43
;```

<table>
<thead>
<tr>
<th>Obs</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.43</td>
</tr>
<tr>
<td>2</td>
<td>7.43</td>
</tr>
<tr>
<td>3</td>
<td>7.43</td>
</tr>
<tr>
<td>4</td>
<td>7.43</td>
</tr>
<tr>
<td>5</td>
<td>7.43</td>
</tr>
</tbody>
</table>

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

```
input State $ City $ Pop2004 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) To omit observations with Pop2004 values less than or equal to 50,000 from the data set.

```
if Pop2004 le 50000 then delete;
```

(b) To be able to store 12 characters (bytes) as city name in the 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 le 25000 then Housgrp = 1 ;
else if 25000 < Housing le 45000 then Housgrp = 2 ;
else Housgrp = 3 ;
```

(d) To create a new numeric variable Percent containing values for Housing as a percentage (ratio multiplied by 100) of Income.

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

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

```
format Income Housing Electric dollar12.2 ;
```
16. The staff of an Olympic swimming center collects timing data for the 200m medley relay event where participants receive training in all styles (backstroke, breaststroke, butterfly, and freestyle) and timings (in sec) were recorded at the beginning of the study and again after 3 months of training.

<table>
<thead>
<tr>
<th>IdNo</th>
<th>Gender</th>
<th>Back1</th>
<th>Back3</th>
<th>Breast1</th>
<th>Breast3</th>
<th>Fly1</th>
<th>Fly3</th>
<th>Free1</th>
<th>Free3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1041</td>
<td>F</td>
<td>35.1</td>
<td>34.7</td>
<td>36.7</td>
<td>36.4</td>
<td>28.3</td>
<td>29.4</td>
<td>36.1</td>
<td>35.9</td>
</tr>
<tr>
<td>2415</td>
<td>F</td>
<td>34.6</td>
<td>33.8</td>
<td>32.6</td>
<td>32.2</td>
<td>26.9</td>
<td>26.8</td>
<td>26.2</td>
<td>26.0</td>
</tr>
<tr>
<td>1798</td>
<td>M</td>
<td>27.2</td>
<td>26.7</td>
<td>33.8</td>
<td>32.5</td>
<td>25.2</td>
<td>24.6</td>
<td>24.1</td>
<td>24.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5812</td>
<td>M</td>
<td>27.0</td>
<td>26.5</td>
<td>29.2</td>
<td>28.2</td>
<td>23.0</td>
<td>22.4</td>
<td>21.9</td>
<td>20.5</td>
</tr>
</tbody>
</table>

(a) Write SAS statements to create a SAS dataset named medley. Use variable names as given above and assume that data will be included instream with data values entered separated by blanks. (4 points)
```
data medley;
input IdNo $ Gender $ Back1 Back3 Breast1 Breast3 Fly1 Fly3 Free1 Free3;
data lines;
1041 F 35.1 34.7 .... 35.9
```

(b) Write SAS statement(s) to be added to the above data step to create a new variable called Improve that indicates the level of improvement for each participant as the average decrease in timings for all four events (5 points)
```
Improve = ((Back1-Back3) + (Breast1-Breast3) + (Fly1-Fly3)
           + (Free1-Free3)) / 4;
```

(c) Write SAS statement(s) to be added to the above data step to create a new category variable ImpGrp containing the values “Small”, “Moderate”, or “Large” according as the average decrease in timings is less than .4, greater than or equal to .4 but less than .5, or greater than or equal to .5. You may use the variable Improve from part (b) in your statement(s) and note that the character-string sizes are different. (5 points)
```
-length ImpGrp $ &;
if Improve < .4 then ImpGrp = "Small";
else if .4 <= Improve < .5 then ImpGrp = "Moderate";
else ImpGrp = "Large";
```
(d) The center statistician would like to use a SAS procedure to compute the summary statistics mean, standard deviation, standard error of the mean, and the maximum, of the timings for the variables Back1, Back3, Breast1, Breast3, Fly1, Fly3, Free1, and Free3, computed separately for each gender within each improvement category "Small", "Moderate", or "Large". Write SAS statements necessary to accomplish this. (7 points)

```
proc means data=medley mean std stderr max;
class ImpGrp Gender;
var Back1 Back3 Breast1 Breast3 Fly1 Fly3 Free1 Free3;
```

17. Display the printed output produced from executing the following SAS program. Show what is in the program data vector immediately after processing the first line of data. (6 points)

```
data sportscars;
  input Model $ Year Make $ Seats Color $;
  Status="current";
  if Year <1975 then Status= "classic";
  if Model = "Corvette"|Model= "Camaro" then Make= "Chevy";
  if Model= "Miata" then do;
      Make= "Mazda";
      Seats=2;
  end;
datalines;
Corvette 1955 . 2 black
XJ6 1985 Jaguar 2 teal
Mustang 1966 Ford 4 red
Miata 1992 . silver
CRX 1991 Honda 2 blue
Camaro 1990 . 4 red
;
proc print; run;
```

```
Model Year Make Seats Color Status
Corvette 1955 Chevy 2 black classic
XJ6 1985 Jaguar 2 teal current
Mustang 1966 Ford 4 red classic
Miata 1992 Mazda 2 silver current
CRX 1991 Honda 2 blue current
Camaro 1990 Chevy 4 red current
```
18. 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 to be written to the SAS data set. (5 points)

```sas
data bd21;
input Y1 Y2 @@;
Y3=Y2**2-5.0;
Y4=sqrt(Y1)/2+1;
drop Y1 Y2;
datalines;
4 -3 0 2 9 . 16 5 1 12
;
proc print data=bd21;
run;
```

<table>
<thead>
<tr>
<th>PDV</th>
<th>Y1</th>
<th>Y2</th>
<th>Y3</th>
<th>Y4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>-3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

19. 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. (5 points)

```sas
data gym;
input Id $ Wt1-Wt5;
array Wts(6) Wt1-Wt5;
drop Week;
Wtloss=0;
do Week=1 to 5;
   if Wts(Week)=. then Wts(Week)=0;
   Wtloss=Wts(Week);
end;
datalines;
PD12 4 6 7 -1 2
HJ23 5 -2 3 4
;
proc print data=gym;
run;
```

<table>
<thead>
<tr>
<th>PDV</th>
<th>Id</th>
<th>Wt1</th>
<th>Wt2</th>
<th>Wt3</th>
<th>Wt4</th>
<th>Wt5</th>
<th>Wtloss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PD12</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>-1</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>HJ23</td>
<td>5</td>
<td>-2</td>
<td>3</td>
<td>4</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>