ASSIGNMENT #5 (same data as used in ASSIGNMENT #3)

#1 FORMATS - You are given a SAS data set named DEATH99. It contains observations for deaths that occurred in 1999 to residents of Albany, Schenectady, Rensselaer, and Saratoga counties. PROC CONTENTS shows:

<table>
<thead>
<tr>
<th>#</th>
<th>Variable</th>
<th>Type</th>
<th>Len</th>
<th>Pos</th>
<th>Format</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>aut</td>
<td>Char</td>
<td>1</td>
<td>15</td>
<td></td>
<td>AUTOPSY</td>
</tr>
<tr>
<td>5</td>
<td>cau</td>
<td>Char</td>
<td>4</td>
<td>4</td>
<td>MMDDYY10.</td>
<td>CAUSE OF DEATH</td>
</tr>
<tr>
<td>2</td>
<td>dob</td>
<td>Num</td>
<td>4</td>
<td>4</td>
<td>MMDDYY10.</td>
<td>DATE OF BIRTH</td>
</tr>
<tr>
<td>1</td>
<td>dod</td>
<td>Num</td>
<td>4</td>
<td>0</td>
<td>MMDDYY10.</td>
<td>DATE OF DEATH</td>
</tr>
<tr>
<td>9</td>
<td>edu</td>
<td>Char</td>
<td>2</td>
<td>16</td>
<td></td>
<td>EDUCATION</td>
</tr>
<tr>
<td>7</td>
<td>eth</td>
<td>Char</td>
<td>1</td>
<td>14</td>
<td></td>
<td>ETHNICITY</td>
</tr>
<tr>
<td>4</td>
<td>gen</td>
<td>Char</td>
<td>1</td>
<td>8</td>
<td></td>
<td>GENDER</td>
</tr>
<tr>
<td>10</td>
<td>pla</td>
<td>Char</td>
<td>1</td>
<td>18</td>
<td></td>
<td>WHERE DEATH OCCURRED</td>
</tr>
<tr>
<td>11</td>
<td>prg</td>
<td>Char</td>
<td>1</td>
<td>19</td>
<td></td>
<td>PREGNANT W/N LAST 6 MONTHS</td>
</tr>
<tr>
<td>6</td>
<td>rac</td>
<td>Char</td>
<td>1</td>
<td>13</td>
<td></td>
<td>RACE</td>
</tr>
<tr>
<td>3</td>
<td>res</td>
<td>Num</td>
<td>3</td>
<td>20</td>
<td>Z2.</td>
<td>COUNTY OF RESIDENCE</td>
</tr>
</tbody>
</table>

and the various codes used for values of variables in the file are...

- **aut**: 0 no, 1 yes, 2 refused
- **cau**: 4-digit ICD-10 cause of death
- **edu**: years of school completed
- **eth**: 0 not Hispanic, 1-4,9 Hispanic
- **gen**: 1 male, 2 female
- **pla**: A Hospital - DOA, B Hospital - ER, C Hospital - outpatient, D Hospital - inpatient, E Other institution, F Decedent's residence, G Other private home, H Other non-institution, N Not in hospital, Z Unknown or not classifiable
- **prg**: 0 no, 1 yes
- **rac**: 1 White, 2 Black, 3 Amerindian, 4 Chinese, 5 Japanese, 6 Hawaiian, 7 Filipino
- **res**: 1 Albany, 41 Rensselaer, 45 Saratoga, 46 Schenectady

Write SAS code that does the following...

A/ Use a FORMAT to create a table of deaths in the following age groups: less than 15, 15-24, 25-44, 45-64, 65-84, 85+

B/ Deaths from major cardio-vascular diseases are coded in the range I00-I78. Use a format to create a table of deaths from major cardio-vascular disease that are in the following groups:

- acute myocardial infarction: I210-I219 and I220-I229
- cerebro-vascular disease: I600-I609
- atherosclerosis: I700-I709
- heart failure: I500-I509
- hypertensive heart disease: I110-I119
- all other causes in the range I00-I78 (the OTHER = in PROC FORMAT)

HINT: create the format for only major-cardiovascular disease; use a WHERE statement in PROC FREQ to limit observations to causes that start with the letter 'I'...

WHERE CAU EQ : 'I';

C/ Use formats to create a table of deaths counted by race and education. Race groups are: white, black, all other. Education groups are: did not finish high school, <12; completed high school, 12; some college, 13-15; college+, 16+

D/ Use formats to compute the mean age at death in the following places: hospital inpatient (D); other hospital deaths (A-C); other institution (E); at home (F); all other places.
The following data step reads all the data with SAS-supplied informats:

```sas
data births;
input
ges 3.
bwt 4.
age 2.
gender $1.
eth $1.
;
label
ges = 'GESTATION IN DAYS'
bwt = 'BIRTH WEIGHT IN GRAMS'
age = 'AGE IN YEARS'
eth = 'ETHNICITY'
;
datalines;
999999915f1
260300085M2
999350025F2
330450034M9
120010017X9
;
run;
```

The allowable ranges and values for the variables are as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Range</th>
<th>Gender</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>GES</td>
<td>140-349</td>
<td>M, F</td>
<td>1, 2</td>
</tr>
<tr>
<td>BWT</td>
<td>500-7499</td>
<td>(M-MALE, F-FEMALE)</td>
<td>(1-HISPANIC, 2-NON-HISPANIC)</td>
</tr>
<tr>
<td>AGE</td>
<td>10-49</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Modify the above SAS job. Create your own INFORMATS to read the data and convert non-allowable values to MISSING (or UPPERCASE) as the raw data are read. PRINT the data set to see if what you did actually worked.