**Homework 9**

1. Create the following SAS dataset on 5 college students:

**DATA** COLLEGE;

INPUT ID AGE GENDER $ GPA CSCORE;

DATALINES;

1 18 M 3.7 650

2 18 F 2.0 490

3 19 F 3.3 580

4 23 M 2.8 530

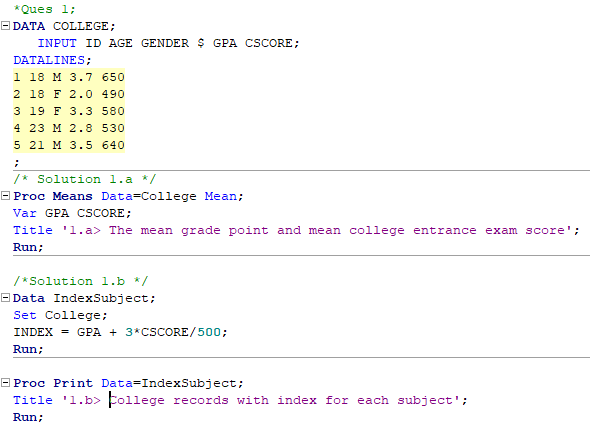
5 21 M 3.5 640

;

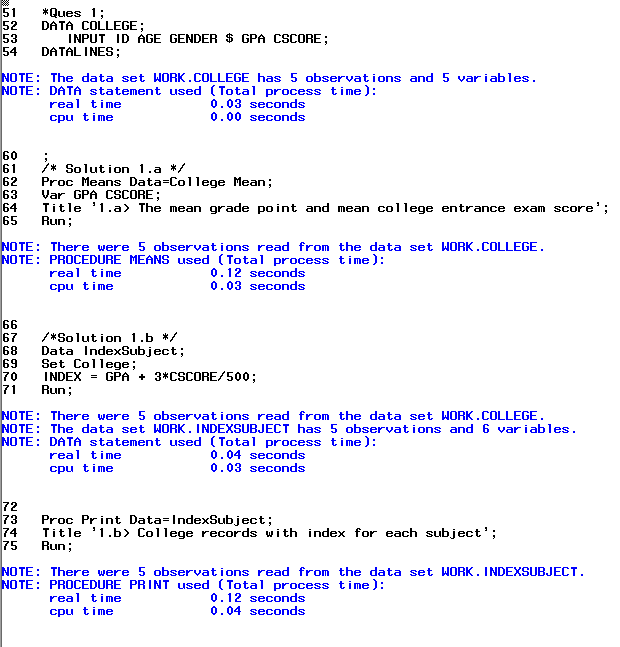
1. Add statements necessary to compute the mean grade point average and mean college entrance exam score.
2. We want to compute an index for each subject, as follows:

INDEX = GPA + 3\*CSCORE/500;

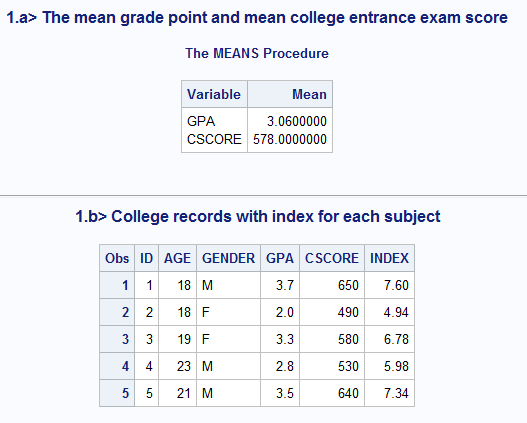
**Code:**



**Log:**

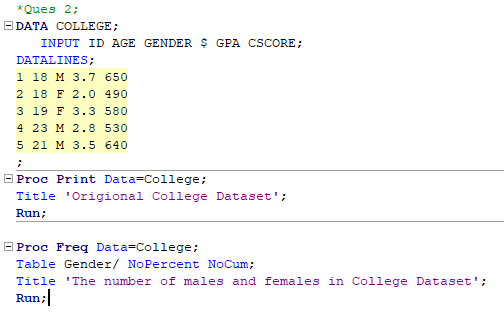


**Results:**

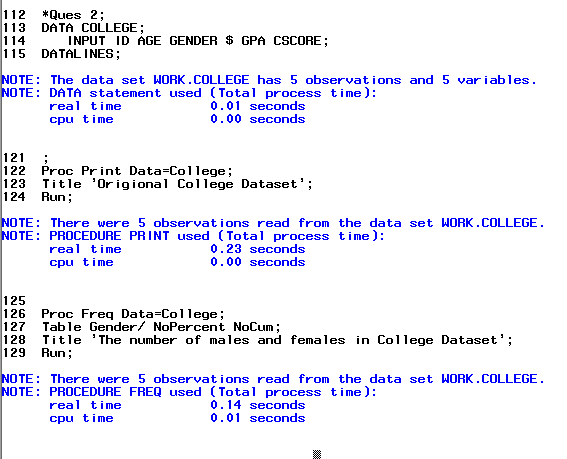


1. Add the necessary statements to compute the number of males and females in the previous problem.

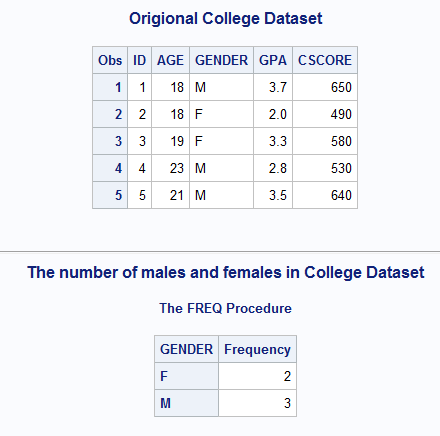
**Code:**



**Log:**



**Results:**



1. Use the data below and create a new variable (AGE\_GROUP) that has a value of 1 for ages between 0 and 35 and 2 for ages greater than 35.

**DATA** TAXPROB;

INPUT SS SALARY AGE RACE $;

FORMAT SS SSN11.;

DATALINES;

123874414 28000 35 W

646239182 29500 37 B

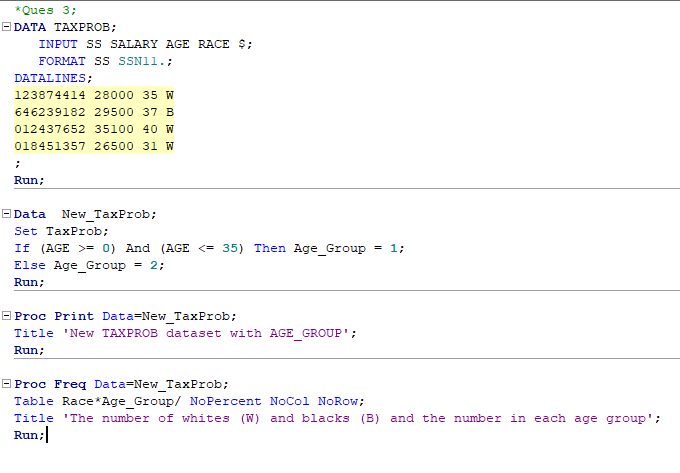
012437652 35100 40 W

018451357 26500 31 W

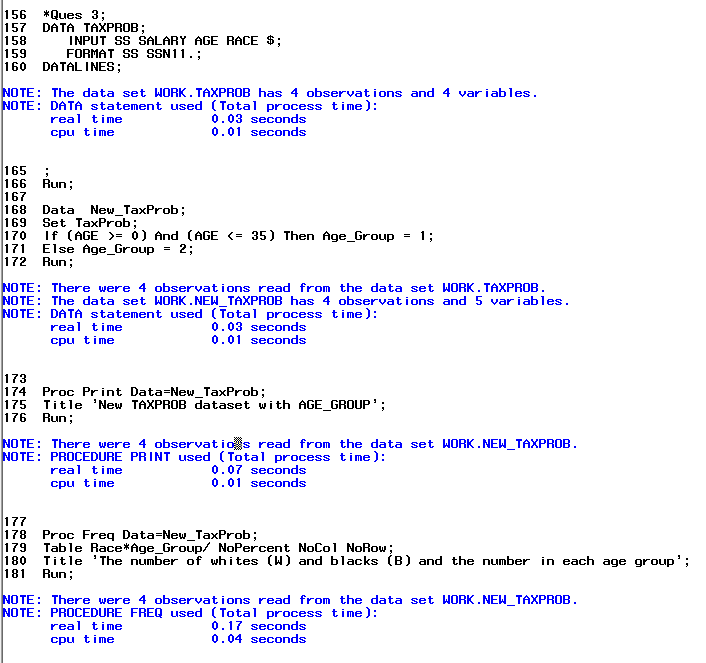
;

Compute the number of whites (W) and blacks (B) and the number in each age group. Use the appropriate option to omit cumulative statistics from the output.

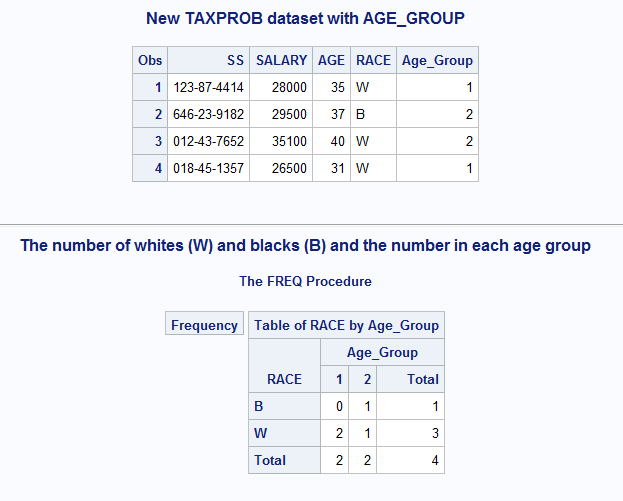
**Code:**



**Log:**



**Results:**



1. Use this data and PROC UNIVARIATE to produce histograms, normal probability plots, and boxplots and test the distributions for normality. Do this for variables like REACT, LIVER\_WT, and SPLEEN, first for all subjects and then separately for each of the two DOSES.

**DATA** LIVER;

INPUT SUBJ DOSE REACT LIVER\_WT SPLEEN;

DATALINES;

1 1 5.4 10.2 8.9

2 1 5.9 9.8 7.3

3 1 4.8 12.2 9.1

4 1 6.9 11.8 8.8

5 1 15.8 10.9 9.0

6 2 4.9 13.8 6.6

7 2 5.0 12.0 7.9

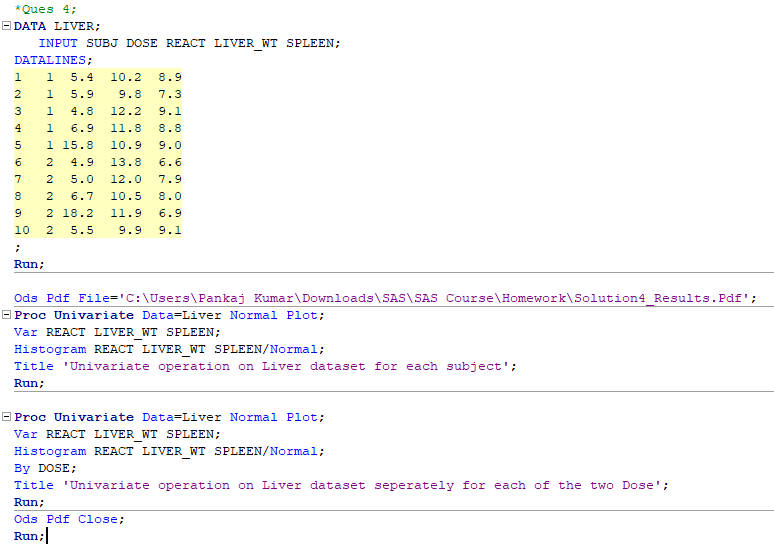
8 2 6.7 10.5 8.0

9 2 18.2 11.9 6.9

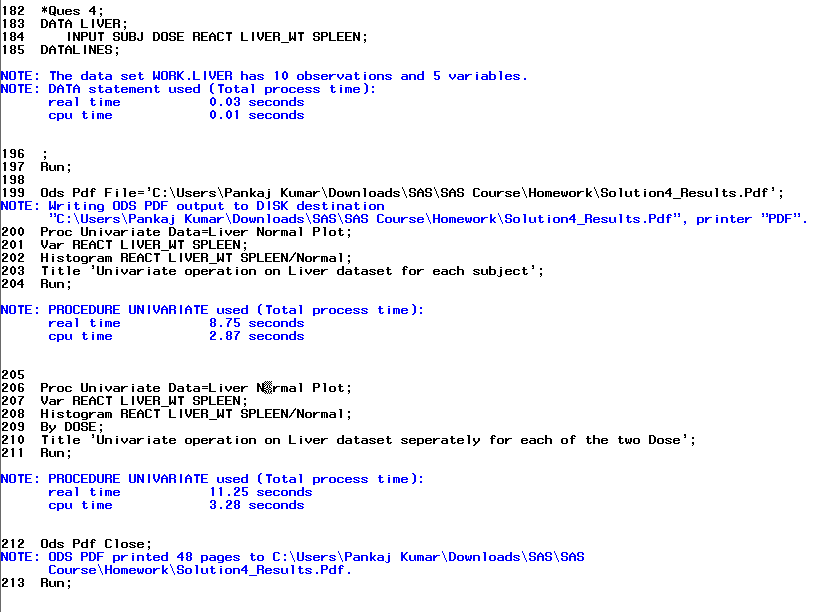
10 2 5.5 9.9 9.1

;

**Code:**



**Log:**



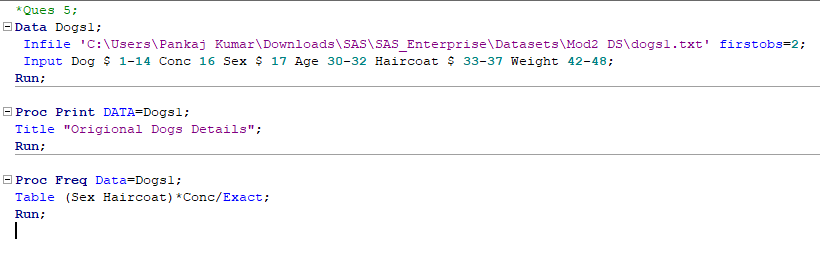
**Results:**

Results saved in the pdf file is attached below:

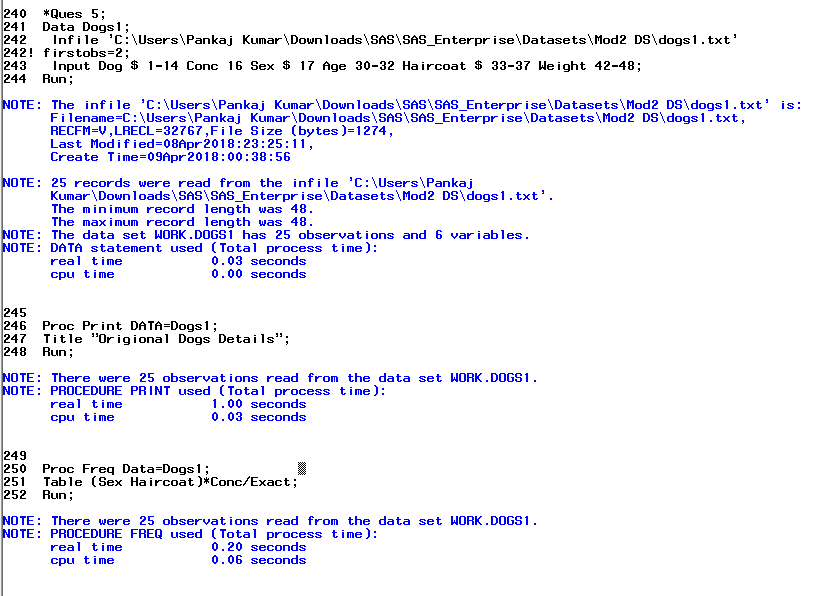


1. Refer to DOGS1 data. The investigator may want to ensure that the dogs allocated to each treatment group were of similar compositions with respect to gender and hair coat. Use PROC FREQ to conduct Fisher’s exact test to see if the concentration of the drug received was statistically independent of the gender of the dog. Likewise, see if the length of the coat and the drug treatment were statistically independent with Fisher’s exact test. Write your interpretation of the results of these tests.

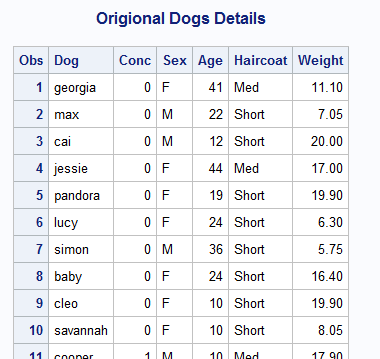
**Code:**

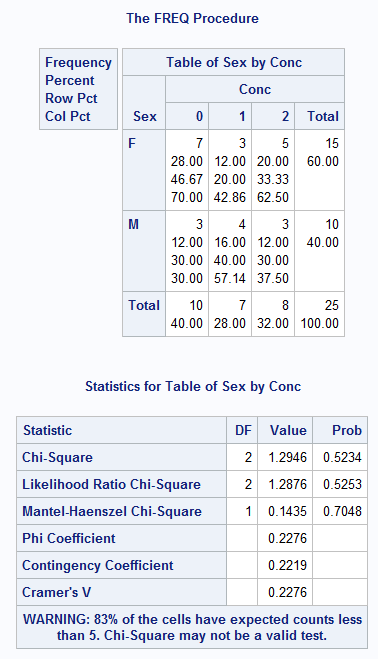


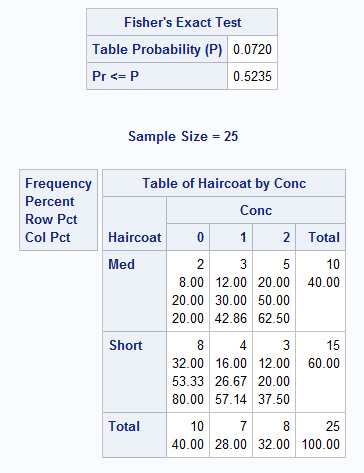
**Log:**

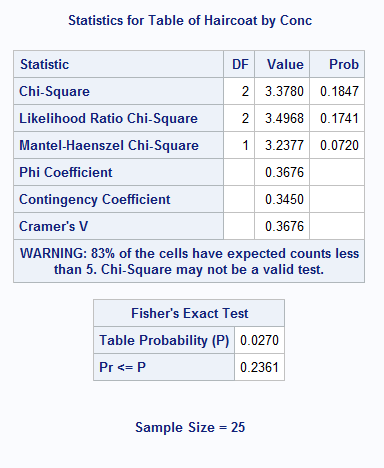


**Results:**



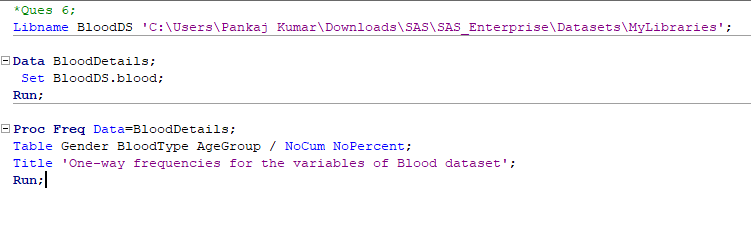




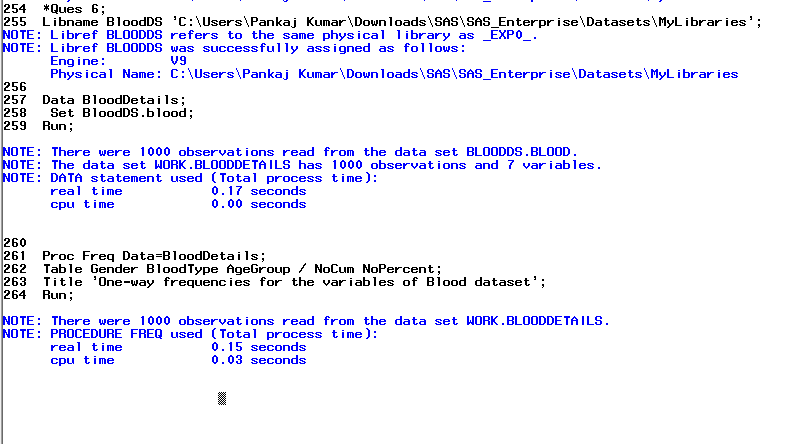


1. Using SAS data set Blood, generate one-way frequencies for the variables Gender, BloodType, and AgeGroup. Use the appropriate options to omit the cumulative statistics and percentages.

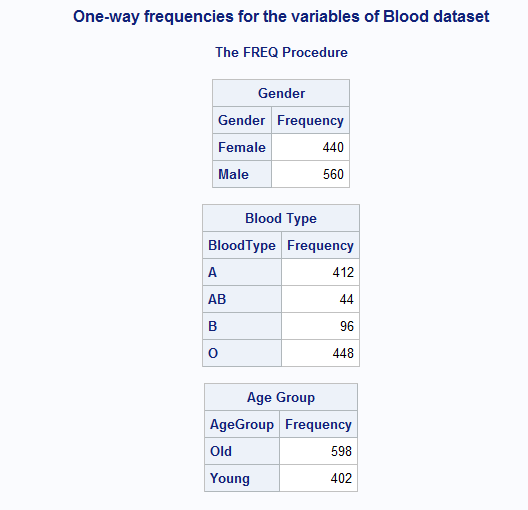
**Code:**



**Log:**

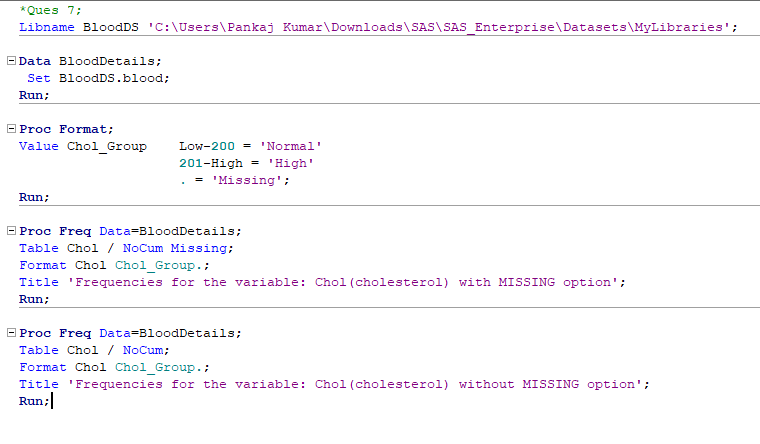


**Results:**

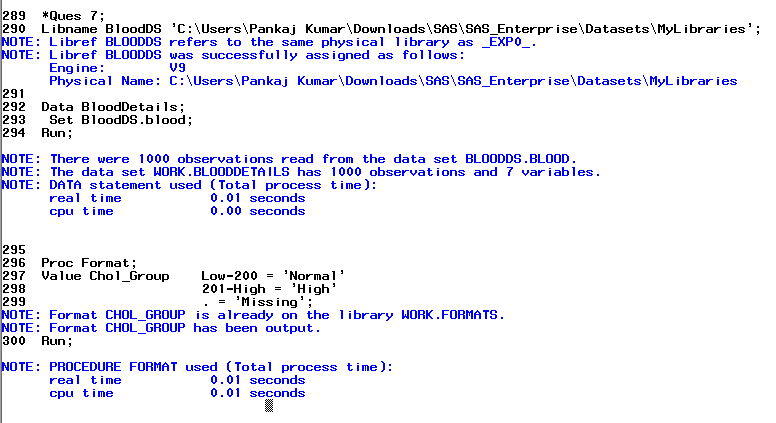


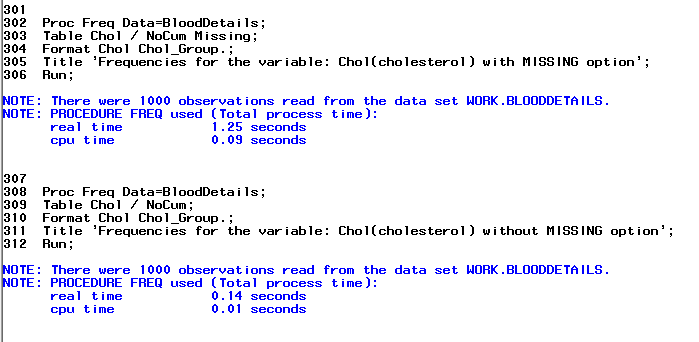
1. Using data set Blood, produce frequencies for the variable Chol (cholesterol). Use a format to group the frequencies into three groups: low to 200 (normal), 201 and higher (high) and missing. Run PROC FREQ twice, once using the MISSING option and once without. Compare the percentages in both listings.

**Code:**

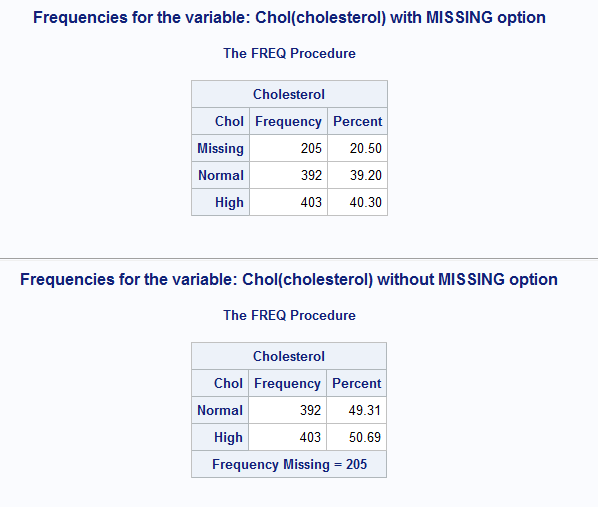


**Log:**





**Results:**



1. What’s wrong with this program?

[You have to specify an error (if there is one) in each line (line numbers are given for your convenience). If you don't see any error, you can write "No Error". You are also welcome to write how you could improve the code, if you feel so.]

Your answer will be in the format,

Line 1 : <Your reason/observation>

Line 2:...

...

Line 11: ..

**1** DATA **123**;

**2** INPUT AGE STATUS PROGNOSIS DOCTOR GENDER STATUS2

**3** STATUS3;

**4** (data lines)

;

**5** PROC CHART DATA=**123** BY GENDER;

**6** VBAR STATUS

**7** VBAR PROGNOSIS;

**8** RUN;

**9** PROC PLOT DATA=**123**;

**10** DOCTOR BY PROGNOSIS;

**11** RUN;

**Solution:**

**Line 1:** The name of the dataset is invalid as it starts with an integer.

**Line 2-3:** Assuming the data type of each variable as Integer makes the Input statement valid. But, Variables like Status, Doctor and Gender are usually Character type which can make the Input statement invalid. But since the data is not provided under Data lines, we cannot say anything about the validity of the Input statement.

**Line 4:** The 2 words: **Data**  and **Lines** should be a single word as **DataLines** and it should not be enclosed within parenthesis.

**Line 5:** The semicolon is missing after the dataset name and before the By keyword. Also, if we are using this By statement for chart, we need to sort the data first.

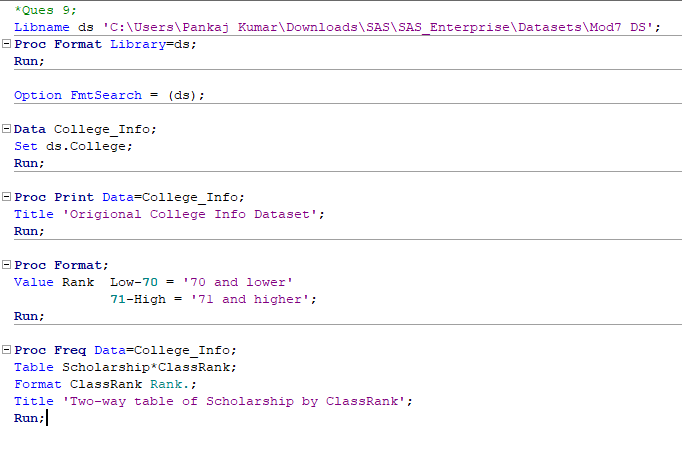
**Line 6:** Missing semicolon at the end of the statement.

**Line 7:** We could have used single Vbar statement with 2 variables.

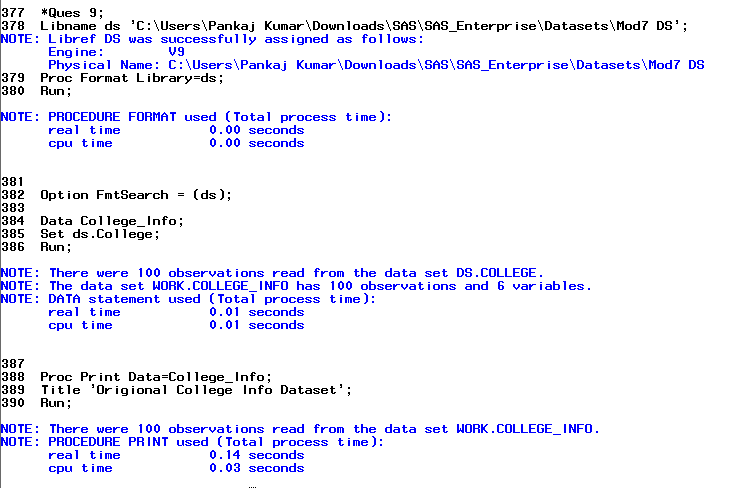
**Line 10:** Plot keyword missing and By keyword should not be there.

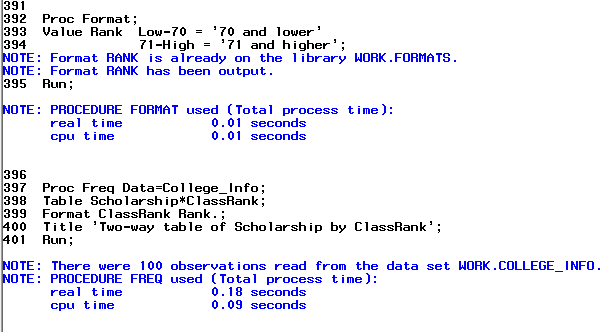
9. Using SAS data set COLLEGE, create two-way table of Scholarship (rows) by ClassRank (columns). Use a user-defined format to group class rank into two groups: 70 and lower, and 71 and higher.

**Code:**

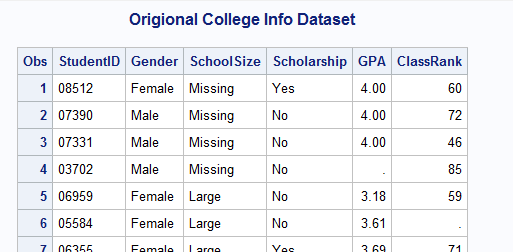


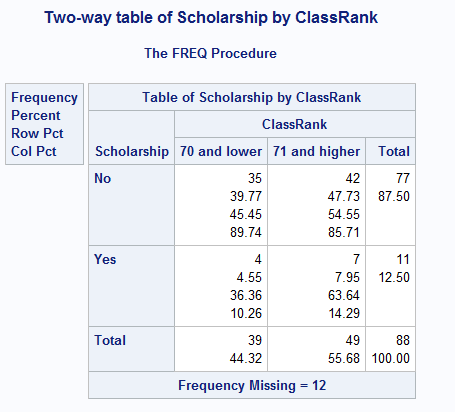
**Log:**





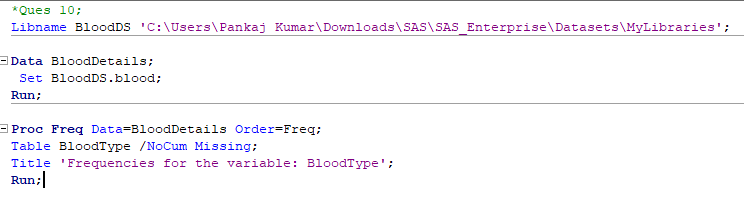
**Results:**



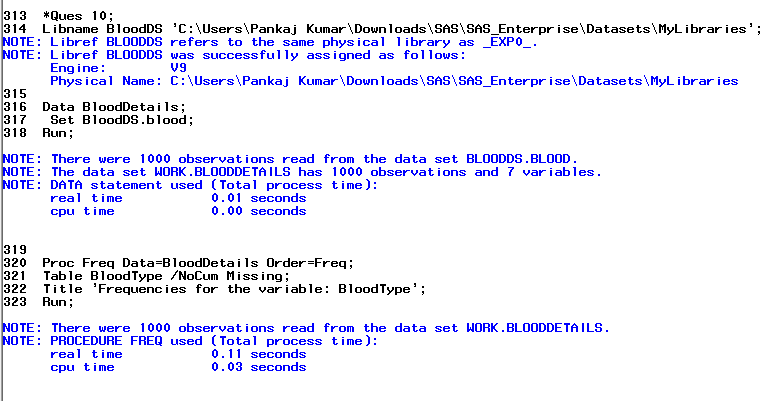


10. Using SAS data set blood, produce a table of frequencies for BloodType, in frequency order.

**Code:**



**Log:**



**Results:**

