This exercise uses the IVH data set that you saved in the last homework assignment. If you don't have the data set handy, you may copy mine, g:\shared\bio113\ivh.sas7bdat. This exercise may be done as a *single long program* (Is it possible to do the recoding and assignment tasks in a single data step?) or as a *series of small ones*.

You printed 10 observations from the data set in homework 1 to see if the variables were as promised by the description of the variables in the data dictionary (page 3 of Homework 1). A much better way to examine the variables for all the observations (you don't want to print out all 566 observations and scrutinize each one) involves using SAS PROCS. Summarize the continuous variables (variables taking on, say, more than 10 values) using PROC MEANS (e.g., proc means; var <varlist>;) (Notes page 56, LSB3 page 218) and the discrete ones with PROC FREQ (e.g., proc freq; tables <varlist>;) (Notes page 53, LSB3 page 220). In your program, write SAS statements to perform the two PROCs ***without*** creating a SAS dataset.

Examine the output to determine which variables need to have values assigned to SAS's missing value code (.). The values that need to be reassigned are listed in the 'Missing code' column of the data dictionary. As you continue your program, use assignment statements and ARRAYs (when it makes sense) to perform the recodes.

Save this recoded data set as a new ***permanent*** SAS data set named **ivh1** on your p drive. Turn in the SAS program(s) and log(s) where you run the PROCs and where you perform the recodes and save the data set. ***Do not*** turn in the SAS output--keep it as a reference for future homework assignments. By the way, you may want to repeat the two PROCs to confirm that your recodes worked.

Use your new permanent SAS data set, **ivh1**, for the remaining questions. Add the variables described below to the new permanent SAS data set (I.e., accumulate the new variables into the existing permanent data set). You will use this data set in future homework assignments. ***Do*** turn in the SAS programs, logs and output for these questions.

Create a variable that divides birth weight into 4 categories: < 750 grams, 750 to < 1000 grams, 1000 to < 1250 grams, and 1250 or more grams.

Create a variable that divides duration of labor into 3 categories: 0 hours, > 0 hours to ≤ 12 hours, and > 12 hours.

Create a variable that divides duration of membrane rupture into 2 categories: < 1 hour, 1 or more hours.

Create a last variable with 6 categories based on both gestational age and birth weight: (1) ≤ 1000 grams and < 26 weeks, (2) ≤ 1000 grams and 26-28 weeks, (3) ≤ 1000 grams and > 28 weeks (4) > 1000 grams and < 26 week, (5) > 1000 grams and 26-28 weeks, and (6) > 1000 grams and > 28 weeks.

Use PROC FREQ to display ***1-way*** tables of each of the 4 categorical variables created above. Then use PROC FREQ to display ***2-way*** tables of the categorical variable for duration of membrane rupture with the other 4 categorical variables (3 more tables).

Use SAS ARRAYs to create cc1, cc2, cc3, and cc4, the weight adjusted sum of colloids and crystalloids, in cc's per kilogram, i.e., ((daily colloid) + daily crystalloid))/daily weight/1000). Take care that you do a calculation when the baby got no colloid (this happens fairly often) but did get crystalloid and when the baby got colloid but no crystalloid (rare).

Use SAS ARRAYs to create variables pctwt1, pctwt2, pctwt3 and pctwt4, the percent change from birth weight of each daily weight. The calculation is: 100*(birth weight - daily weight)/birth weight. ***Round*** the four variables to the nearest tenth (0.1).

Use SAS functions to calculate the average of the mean blood pressure (MAP), the average pCO_2, the lowest pCO_2, and the highest pCO_2 of the 4 daily measures for each baby (4 variables).

Write an arithmetic expression to calculate the average of the 4 variables for the mean blood pressure (MAP) and the pCO_2 for each baby (2 variables).

Create an Apgar score ratio variable, the ratio of the 1 minute Apgar and the 5 minute Apgar scores (i.e., apgar1/apgar5) and ***round*** it to the nearest thousandth (0.001).

Use PROC MEANS to display the mean, min, max, n, and nmiss of the ***15*** new variables (from cc1 to the Apgar ratio variable).

Use SAS ARRAYs to create a variable that ***counts*** the number of adverse events each baby had. The events are 5 minute Apgar < 5, IVH, death, and receipt of dopamine and occurrence of PDA or pneumothorax on days 1 through 4 (15 possible events). Hint: this task and the next one are easy if each event takes on the values 0 and 1.

Calculate the number of events, again, (create a second variable) using a SAS ***function***.

Use PROC FREQ to see the values taken on by these two variables.