**Workshop Exercise 4**

**Exercise IV: Summarizing Data**

**Dataset=hsb1.sas7bdat**

In this exercise we will use the permanent SAS dataset, hsb1.sas7bdat, which is in the sasdata2 folder. This dataset contains information on 7,185 elementary school students that were included in a larger survey of students. These students came from 160 schools (SCHOOL).

1. Get Proc Contents for the HSB1 data set. How many observations and how many variables are in this data set?
2. Calculate the following summary statistics for each SCHOOL, and save them to a new data set called MEANDAT (remember, you will need to sort by SCHOOL first):

1. Total number of students
2. Mean socioeconomic status (SES) for the school
3. Mean math achievement (MATHACH) in the school
4. Total number of females (FEMALE) in the school
5. Total number of minority students (MINORITY) in the school

How many cases are in your new data set, MEANDAT (there should be 160)?

1. Calculate some new variables. To do this you will need to use a data step. Calculate:
   1. Percent Female (PCT\_FEMALE) in each school.
   2. Percent Minority (PCT\_MINORITY) in each school.
2. Use Proc Means to get descriptive statistics for all variables in your new summary data set.
3. Use Proc Freq to get oneway frequencies for percent female and percent minority.\
4. Use Proc Univariate to get a histogram for percent female and percent minority. Describe how these histograms look.

Extra credit:

1. Fit a linear regression model with the mean of math achievement as the outcome variable and the mean of SES as the predictor. Get diagnostic plots for this model.