**STAT 8510**
**Homework #3**
**Due Sept. 23, 2021 (SAS program and output)**

1. Using the permanent SAS data set that you created in Homework #1 of the 200 babies, you will manipulate the data using IF-THEN-ELSE IF statements and create new variables. The variables that were listed are given below. You will need to refer back to Homework #1 to determine your variable names.

| Variable | Column | Variable Type |
|---|---|---|
| Sex | 1-6 | Character |
| Prenatal care | 9-12 | Character |
| Smoking status | 17-19 | Character |
| Weeks gestation | 25-26 | Numeric |
| Birth weight in grams | 33-36 | Numeric |
| Length in inches | 41-44 | Numeric |
| Date of birth | 49-58 | Date in MMDDYY format |

   a. Create a new variable for preterm birth that has two levels using the weeks gestation variable with values of 1 for preterm and 0 for term. Any baby who had a gestation of strictly less than 37 weeks is considered preterm. Those greater than or equal to 37 weeks are considered term.
   b. Using the birth weight in grams variable create a variable that has three levels that indicates low, normal, and large birth weights, assigning values of -1 to low, 0 to normal, and 1 to large. Any baby who has a birth weight less than 2500 grams is considered having a low birth weight, any baby who has a birth weight between 2500 and 4000 grams inclusive is considered having a normal birth weight, and any baby greater than 4000 grams is considered having a large birth weight.
   c. Give a frequency distribution of all categorical variables, including the two new categorical variables. Title your tables.
   d. Give a cross tabulation of preterm birth with the following variables: the new birth weight group, sex, prenatal care, and smoking status. Title your tables.

2. Using the permanent SAS data set named "cesd1" stored in In the D2L homework 3 folder. You will be creating the Centers for Epidemiologic Studies Depression (CES-D) score. The CES-D is a 20-item instrument that was developed by the National Institute of Mental Health to detect major or clinical depression in adolescents and adults. Each item is scored on a 0-3 scale with 0 indicating rarely or none, 1 indicating some or a little of the time, 2 indicating occasionally or moderate amount of time, and 3 indicating most or all of the time. The variables are ID, and CESD1-CESD20.

   a. CES-D items 4, 8, 12, and 16 are worded positively while the other items are worded negatively. Create an ARRAY that contains these 4 items that specifies there are 4 items in the array. Create another ARRAY that will contain 4 reversed variables

cesd4new, cesd8new, cesd12new, cesd16new and specify that the array contains 4 items.

b. Using a DO loop with an index that goes from 1 to 4, reverse these items in the ARRAY so that a value of 0 becomes a value of 3, a value of 1 becomes a value of 2, a value of 2 becomes a value of 1, and a value of 3 becomes a value of 0.
c. Create the CES-D total score by summing up items in the 16 non-reversed items and the 4 reversed items.
d. Calculate the mean, standard deviation, minimum, maximum and median CES-D total score.