```
*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 Homework#1 to determine your variable names.

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.;

*PROC FORMAT;
      *VALUE fmtwksgestn
                  0 = "term"
                  1 = "preterm"
                  ;
*RUN;

DATA Homework3;
INFILE "D:\SASProgrammingPractice\Homework3\babies.txt";
INPUT sex $1-6 prenatalcare $9-12 smokestatus $17-19 gestationwks 25-26
BWingrams 33-36
LengthinIN 41-44 @49 DOB mmddyy8.;
*FORMAT gestationwks fmtwksgestn.;

      *IF 0 <= gestationwks < 37 then preterm = 1
        ELSE IF gestationwks >= 37 then term = 0;

        IF 0 <= gestationwks < 37 then pretermbirth = 1 ;
        ELSE IF gestationwks >= 37 then pretermbirth = 0;

*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;

        *IF BWingrams < 2500 then lowBW = -1
        ELSE IF 2500 < BWingrams < 4000 then normalBW = 0
        ELSE IF BWingrams > 4000 then largeBW = 1;

       IF BWingrams < 2500 then BWlevel = -1;
        ELSE IF 2500 < BWingrams < 4000 then BWlevel = 0;
        ELSE IF BWingrams > 4000 then BWlevel = 1;
PROC PRINT DATA=Homework3;
RUN;

* c) Give a frequency distribution of all categorical variables, includng the
two new categorical variables.
```

```
Title your tables;

PROC FREQ DATA=Homework3;
TABLE sex prenatalcare smokestatus pretermbirth BWlevel;
TITLE "Frequency Distribution of Categorical Variables";
RUN;

* 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;

PROC FREQ DATA=Homework3;
TABLES pretermbirth*(BWlevel sex prenatalcare smokestatus);
TITLE "Cross Tabulation of Preterm Birth With Variables";
RUN;

*PROC CONTENTS DATA=Homework3
RUN;

* Using the permanent SAS data set named "cesd1" stored in the D2L homework3
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
Insitute 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.
;

LIBNAME cesd "D:\SASProgrammingPractice\Homework3";

DATA cesddata;
SET cesd.cesd1;

*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.;

ARRAY cesditem{4} CESD4 CESD8 CESD12 CESD16;
ARRAY cesditnew{4} cesd4new cesd8new cesd12new cesd16new;

*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.;

DO i = 1 to 4;
     cesditnew[i] = 3 - cesditem[i];
     END;
```

```
* c) Create the CES-D total score by summing upitems in the 16 non-reversed
items and the 4 reversed items;

       cesdtot = sum(of
CESD1,CESD2,CESD3,CESD5,CESD6,CESD7,CESD9,CESD10,CESD11,CESD13,CESD14,CESD15,
CESD17,CESD18,
CESD19,CESD20,cesd4new,cesd8new,cesd12new,cesd16new);
proc print data=cesddata;
SUM cesdtot;
RUN;


* d) Calculate the mean, standard deviation, minimum, maximum and median CES-
D total score;

PROC MEANS DATA=cesddata MEAN STDDEV MIN MAX MEDIAN;
VAR cesdtot;
RUN;

*PROC CONTENTS DATA=cesddata
RUN;
```