ASSIGNMENT #7 *** ANSWERS ***

#1 (NUMERIC FUNCTIONS) Here is a data step that creates the data set QUESTIONS...

data question;
  input id : $3. q1-q8 gender : $1. test : mmddyy.;
  format test mmddyy10.;
  datalines;
    101 2 3 . 2 3 3 3 M 011098
    102 4 4 1 . 4 . 3 M 101198
    103 4 3 4 2 2 3 3 F 111598
    104 2 . 3 2 3 . 3 F 120498
    105 2 3 3 2 3 4 3 . M 020399
    106 4 2 . 4 4 4 4 4 M 040699
    107 4 4 . 2 3 4 3 3 F 051599
    108 3 . . 1 3 3 . M 082099
    109 2 . . 2 1 2 . F 101199
    110 2 4 . 3 3 3 3 M 011000
  ;
run;

Variables q1 through q8 are scores on eight questions asked to each of ten people.

A/ Use the data set QUESTIONS and add three new variables: mean score on the eight questions; total score on the eight questions; number of questions answered (non-missing values).

data questions;
  set questions;
  *** use functions to create new variables;
  ms = mean (of q1-q8);
  ts = sum (of q1-q8);
  ns = n (of q1-q8);
  label ms = 'MEAN SCORE'
       ts = 'TOTAL SCORE'
       ns = '# ANSWERED';
run;

B/ What are the lowest and highest mean scores (mean of q1-q8 for each person) for only those people who answered at least seven questions?

proc means data=questions n min max maxdec=1;
  var ms;
  where ns ge 7;
run;

C/ Find the mean score for each of the eight questions for those people who took the test before 1/1/1999 and for those who took the test on or after 1/1/1999.

proc format;
  value when low -< '01jan1999'd = 'PRIOR TO 1/1/1999'
             '01jan1999'd - high = 'ON/AFTER 1/1/1999';
run;

proc means data=questions mean maxdec=1;
  var q1-q8;
  class test;
  format test when.;
run;
#2  You have the following data...

12345,10,20,30,10,15,45,60,75  
23456,09,05,31,23,12,99,20,54  
34567,10,10,10,54,88,10,10,43  
45678,12,47,14,87,43,12,34,23  
56789,88,88,88,65,12,15,23,22

There's an ID number plus 8 numbers in each record. Write a SAS job that will...

A  read the data and create a SAS data set
B  find the mean of the 8 numbers for each person
C  find the person with the highest and lowest mean score for the eight numbers
D  find the difference for each person between the mean of the first two numbers and the mean of the last two numbers, then find the mean of that difference

```
data test;  
infile datalines dsd;  
input id x1-x8;  
mean_x = mean (of x1-x8);  
diff_x = mean (of x1-x2) - mean(of x7-x8);  
min_x  = min (of x1-x8);  
max_x  = max (of x1-x8);  
datalines;  
12345,10,20,30,10,15,45,60,75  
23456,09,05,31,23,12,99,20,54  
34567,10,10,10,54,88,10,10,43  
45678,12,47,14,87,43,12,34,23  
56789,88,88,88,65,12,15,23,22  
;  
run;
```

E  find the lowest and highest number in the data set

```
proc means data=test min max maxdec=0;  
var min_x max_x;  
run;
```

or ...

```
proc transpose data=test out=new (keep=col1 rename=(col1=score));  
var x1-x8;  
by id;  
run;
```

```
proc means data=new min max maxdec=0;  
var score;  
run;
```
The following creates a SAS data set with four variables:

```
data births;
  infile datalines dsd;
  input bwt : $4.
ges : $3.
gender zip
;
label bwt = 'BIRTH WEIGHT (GRAMS)'
ges = 'GESTATION (DAYS)'
;
datalines;
  3000,280,1,12203
  1500,230,2,
  2800,,,12345
;
run;
```

Write a SAS job that changes the variable types in data set BIRTHS: change BWT and GES to NUMERIC; change GENDER and ZIP to CHARACTER;

NOTE: TRY TO KEEP THE SAME VARIABLE NAMES as in the original data set BIRTHS.

```
* change character-to-numeric, numeric-to-character;
* create new variables;
  data birthb;
  set birtha;
  bwt_n = input(bwt,best.);
ges_n = input(ges,best.);
gen_c = put(gender,$1.);
zip_c = put(zip,$5.);
run;
  proc contents data=birthb;
  run;
  proc print data=birthb;
  run;
  or...

* change character-to-numeric, numeric-to-character;
* keep old variable names;
  options missing=' ';  
  data birthc;
  set birtha (rename=(bwt=_bwt ges=_ges gender=_gen zip=_zip));
  bwt = input(_bwt,best.);
ges = input(_ges,best.);
gen = put(_gen,$1.);
zip = put(_zip,$5.);
drop _: ;
  run;
  options missing='.';
```