Statistics 147 LAB #3
10 pts; Summer 2020

NAME: Wesley Chang **ID: (last 4 #s only)** 0996

This lab is designed to give the student practice with DO loops and IF-THEN-ELSE structures in SAS.

Data File: You will need to download the data file **plant.dat** from Blackboard (under Data Files).

Note: The data starts on Line 2.

Invoke SAS.

Four chemical plants, producing the same product and owned by the same company, discharge effluent into streams in the vicinity of their locations. To check the extent of the pollution created by the effluents and to determine whether the amount of polluting effluents varies from plant to plant, the company collected random samples of liquid waste from each of the four plants. The data, in pounds per gallon of waste, is given in the table below.

PlantA	PlantB	PlantC	PlantD
1.65	1.70	1.40	1.58
1.72	1.85	1.75	1.77
1.50	1.36	1.58	1.48
1.37	2.05	1.65	1.69
1.60	1.80	1.55	1.65
1.40	2.10	1.45	1.65
1.75	1.95	1.66	1.79
1.38	1.65	1.70	1.58
1.65	1.80	1.85	1.77
1.55	2.00	1.24	1.60

Let Plant A be sample 1, Plant B be sample 2, Plant C be sample 3, and Plant D be sample 4.

1. When we use SAS to generate two-sample confidence intervals or perform tests of hypothesis, the data needs to be read in using DO loops so that one can identify the sample associated with each of the observations. Thus, we will need a variable to identify the sample to which the observation belongs and then a common variable to represent the actual data values.

   (i) Read in and print out the data.

   This can be accomplished using as follows:

```
/* Set up format of the output */
options nocenter ps = 55 nocenter ls = 78 nodate nonumber formdlim='*';

/* Use DM to clear all windows except the editor window */
DM log "odsresults; clear; out; clear; log; clear;";
ods graphics off;

/* Create titles */
title1 'Statistics 147 Lab #3, Summer 2020';
title2 'Your name goes here';

/* Create temporary SAS dataset named lab3su20 */
```

```
data lab3su20;
    /* Open data file plant.dat. Be sure to specify the path indicating where you have
       saved the data file.  The actual data starts on Line 2.*/
    infile '/PATH/TO/YOUR/DATAFILES/plant.dat' firstobs = 2;

    /* Create nested do loops to read in the data
       NOTE: There are 10 rows and four columns of data
       First, Do loop for the rows*/
    do row = 1 to 10;
       /* Do Loop for the columns */
       do plant = 1 to 4;
           /* Use If-Then-Else Structure to name the plants */
           if      plant = 1 then name = 'Plant A';
           else if plant = 2 then name = 'Plant B';
           else if plant = 3 then name = 'Plant C';
           else                   name = 'Plant D';

           /* Input response (data values) */
           input dischrg @@;

           /* Output the data */
           output;
       /* Close the plant Do-loops */
       end;
   /* Close the row Do-loops */
   end;
run;

/* Print the results */
proc print noobs data = lab3su20;
   /* Create titles for output*/
   title3 'Question 1';
   title4 'Part (i) Read in and Print data';
run;

quit;
```

Save your file as **lab3su20s.sas**. Execute the program and when you have the data printed on the screen in your output window, have Luke, Ruihan, or your neighbor initial here: _____

(ii) Suppose we would like to generate the mean and variance for the effluent discharge for each plant. We can accomplish this by adding the following lines to our existing program right before the **run** statement:

```
/* First sort the data according to the variable plant */
proc sort data = lab3su20;
   /* Modify title4 */
   title4 'Part (ii): Descriptive Statistics';
   /* Specify the variable to be used to sort */
   by plant;
run;

/* Use proc means to generate the mean and variance for each plant
```

```
    n              number of observations
    mean           sample mean
    var            sample variance
    by plant       group the data according to the plant from which the observation
                   was selected
    var discharge  generate mean and variance of the variable dischrg (for each plant) */

proc means n mean var data = lab3su20;
   by plant;
   var dischrg;
run;
```

Save and execute your program. Complete the following table.

Plant	n	Mean	Variance
Plant A	10	1.557	0.020
Plant B	10	1.826	0.049
Plant C	10	1.583	0.033
Plant D	10	1.656	0.010

(iii) Now suppose you want to generate descriptive statistics for the mean discharge effluent for Plant B only. You must first restrict the data to Plant B. This can be accomplished by adding the following lines of code to the bottom of your SAS program (right before the **run** statement):

```
/* Create new temporary SAS dataset which only contains the
   observations from Plant B */
data onlyB;
    /* Use set command to bring in all the data */
    set lab3su20;

    /* Use if statement to restrict the data to Plant B, i.e., Plant 2*/
    if plant = 2;    /* Could also use: if name = 'Plant B'; */
run;

proc print noobs data = onlyB;
    /* Modify title4 */
    title4 'Part (iii): Plant B only';
run;

proc means n mean stddev data = onlyB;
    tite4 'Descriptive Statistics for Plant B only';
    var dischrg;
run;
```

Save and run (execute) your program and complete the following:

Plant	n	Mean	Std Deviation
Plant B	10	1.826	0.220

(iv) Now suppose you would like to use the data from both Plants A and B. This can be accomplished by adding the following code to the end of your program (**right before the run statement**):

```
/* Create new SAS dataset which only contains the
   observations from Plants A and  B */
data bothAB;
    /* Use set command to bring in all the data from the lab3su20 dataset*/
    set lab3su20;

    /* Use if statement to restrict the observations to Plants A or B, i.e., Plants 1 or 2 */
    if plant = 1 or plant = 2;
    /* Could also use: if name = 'Plant A' or name = 'Plant B'; */
run;


proc print data = bothAB;
    /* Modify title4 */
    title4 'Part (iv): Plant A and Plant B with Descriptive Statistics ';
run;

/* Use proc means to generate some descriptive statistics.
   Use the by statement to generate statistics for each plant. */
proc means n mean stddev data = bothAB;
    by plant;
    var dischrg;
run;
```

Save and run (execute) your program and complete the following:

Plant	n	Mean	Std Deviation
Plant A	10	1.557	0.140
Plant B	10	1.826	0.220

2. Now let's try some DO loop calculations! Using nested DO loops, calculate $y = 2x^2\sqrt[3]{t}$, $x = 1, 2$ and $t = 3, 6, 9$.

   We can accomplish this by adding the following lines of code to the end of our existing program (right before the **run** statement).

```
/* Set up temporary SAS data set called looptry */
data looptry;
    /* Set up loop for x values */
    do x = 1 to 2;
        /* Set up loop for t values.
           Do t = start to stop by increment */
        do t = 3 to 9 by 3;
            /* Set up equation*/
            /*Don't forget, x^2 is x**2 in SAS*/
            y = 2*x**2*(t**(1.0/3.0));

            /*Output the value each time through loop*/
            output;
        /* Close the t loop */
        end;
    /* Close the x loop */
    end;
run;

/* Print the results */
proc print noobs data = looptry;
    /* Modify title3 */
    title3 'Question 2: Do Loops';
    /* Clear title4 */
    title4 ' ';
    var x t y;
run;
```

   Save and run (execute) your program and complete the following:

x	t	y
1	3	2.8845
1	6	3.6342
1	9	4.1602
2	3	11.5380
2	6	14.5370
2	9	16.6407

You have now successfully completed Lab #3. Thanks and have a good day!

Luke & Ruihan