Statistics 147 LAB #5
10 pts; Summer 2020

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

This lab is designed to give the student practice performing tests of hypothesis for the difference of two population means (assuming independent samples) and **paired-t tests of hypothesis** using R and SAS.

REMINDERS:

- ♣ You reject H_0 if p-value $<$ specified of α. We'll use $\alpha = 0.05$.
- ♣ If $p-value < \alpha = 0.05 \Rightarrow$ reject $H_0 \Rightarrow$ can conclude there is a significant difference. If $p-value \not< \alpha = 0.05 \Rightarrow$ cannot reject $H_0 \Rightarrow$ cannot conclude there is a significant difference.

Data Files: You will need to download the data file **cartest1.dat** from Blackboard (under Data Files). You will also need the data file, **plant.dat** (again).

1 Using SAS:

1. (Using 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. (Name your columns: Plant A. Plant B, Plant C, and Plant D.)

Let

μ_A	=	true mean discharge effluent for *Plant A*
μ_B	=	true mean discharge effluent for *Plant B*
μ_C	=	true mean discharge effluent for *Plant C*
μ_D	=	true mean discharge effluent for *Plant D*
σ^2_A	=	population variance of the discharge for *Plant A*
σ^2_B	=	population variance of the discharge for *Plant B*
σ^2_C	=	population variance of the discharge for *Plant C*
σ^2_D	=	population variance of the discharge for *Plant D*

Do the data provide sufficient evidence to indicate a significant difference in the mean amount of effluent discharged by Plant A and Plant B? (Recall: You must perform two tests for this problem. First you must check to see if it is reasonable to assume equality of variances. Second, based on the results of the test for equality of variances, you must the test the means.)

In SAS, one can generate the information necessary to test for equality of variances and for equality of means using **proc ttest**.

Recall, in Labs #3 and #4, you read in the data. So open your SAS program file (lab4su19s.sas), save it as **lab5su19s.sas**, and add the following lines of code, **right before** the **quit** statement. **(Modify title1 to show it is Lab 5.)**

```
/* Use proc ttest to generate information
     class     name of classification variable
     var       name of variable to test */
 proc ttest;
    class plant;
    var dischrg;
run;
```

Reminder: Save your SAS file, execute it, and complete the following:

```
****************************************************************************
The TTEST Procedure
                                 Statistics
                               Lower CL            Upper CL  Lower CL
Variable  plant          N       Mean      Mean      Mean    Std Dev   Std Dev
dischrg             1   10     1.4566     1.557    1.6574    0.0965    0.1403
dischrg             2   10     1.6684     1.826    1.9836    0.1515    0.2203
dischrg   Diff (1-2)           -0.443    -0.269    -0.095    0.1396    0.1847
                           Statistics
                        Upper CL
Variable  plant          Std Dev     Std Err     Minimum     Maximum
dischrg             1    0.2562      0.0444        1.37        1.75
dischrg             2    0.4022      0.0697        1.36         2.1
dischrg   Diff (1-2)     0.2731      0.0826
                               T-Tests
Variable    Method           Variances       DF    t Value     Pr > |t|

dischrg     Pooled           Equal           18    -3.26       0.0044

dischrg     Satterthwaite    Unequal       15.3    -3.26       0.0052


                    Equality of Variances

Variable    Method      Num DF    Den DF    F Value    Pr > F

dischrg     Folded F         9         9     2.47      0.1951
****************************************************************************
```

First test for equality of variances:

NOTE: If the p-value for the F-test is $< \alpha = 0.05$, one cannot assume the variances are equal.

- ♣ $H_0 : \sigma_A^2 = \sigma_B^2$ (reasonable to assume equality of variances)
- ♣ $H_a : \sigma_A^2 \neq \sigma_B^2$ (not reasonable to assume equality of variances)
- ♣ p-value = 0.1951
- ♣ Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$.
- ♣ What is your conclusion? Can one assume equality of variances?

Conclusion: Since the p-value = 0.1951 is (less than, greater than) [circle your choice] $\alpha = 0.05$, (reject, do not reject) [circle your choice] $H_0 \rightarrow$ it (is, is not) [circle your choice] reasonable to assume equality of variances.

Now, test the means, based on your results from the test for equality of variances test:

- ♠ $H_0 : \mu_A = \mu_B$ (cannot conclude a significant difference in mean effluent discharge between Plant A and Plant B)
- ♠ $H_a : \mu_A \neq \mu_B$ (can conclude a significant difference in mean effluent discharge between Plant A and Plant B)
- ♠ p-value = 0.0044.
- ♠ Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$.
- ♠ Can one conclude there is a significant difference in the mean amount of effluent discharge between Plant A and Plant B. Justify your answer!

Conclusion: Since the p-value = 0.0044 is (less than, greater than) [circle your choice] $\alpha = 0.05$, (reject, do not reject) [circle your choice] $H_0 \rightarrow$ it (is, is not) [circle your choice] reasonable to assume a significant difference in the mean amount of effluent discharged by plant A and B.

2. Two brands of tires were being compared in life tests to determine whether there is a significant difference in average lifelength between the two brands of tires. Six cars were randomly selected and one tire of each brand placed on each car. The cars were then driven a specified number of miles over the same course (No, they did not drive the cars with only two tires - the other two tires on each car were all of a third brand of tire that was not on life test!) After completing the course, the amount of tread wear (in thousandths of an inch) were recorded as follows:

Car	1	2	3	4	5	6
Brand A tires	125	64	94	38	90	106
Brand B tires	133	65	103	37	102	115

NOTE: The data has been saved in an ordinary ASCII (text) file, *cartest1.dat*, which can be read into SAS using an infile statement.

NOTE: The actual data starts on Line 2 in the data file. Add the following lines of code to your existing SAS program file **right before the run** statement.

REMINDER: Since the samples are not independent, one must use the paired-difference t-test. **You do not test for equality of variances!**

```
/* Create temporary SAS dataset.
   Use an infile statement to open the data file.
   Use firstobs = 2 to tell SAS to start reading the data on Line 2.
   Use an input statement to input the names of the variables.
   BE SURE TO CHANGE THE PATH TO YOUR DATA FILE. */

data paired;
      infile 'c:\Luke\summer2020\su19147\datafiles\cartest1.dat' firstobs = 2;
      input  car brandA brandB;
run;
quit
```

Save and execute your file. When your new output appears on the screen, have Luke, Lauren or your neighbor initial here. AZ

REMINDER: Since the samples are not independent, one must use the paired-difference t-test. You do not test for equality of variances!

Recall: You can use proc ttest to generate the results for a paired-difference t-test. Add the following lines of code **right before** the run statement:

```
ods graphics off;
/* Using proc ttest  for paired difference t-test
Format:  proc ttest;
                paired variable1*variable2;*/
proc ttest;
       paired brandA*brandB;
```

Save and execute your program file. Complete the following:

```
The TTEST Procedure
Difference:  brandA - brandB
 N         Mean      Std Dev      Std Err      Minimum      Maximum
 6     -6.3333       5.1251       2.0923     -12.0000       1.0000
    Mean        95% CL Mean         Std Dev       95% CL Std Dev
 -6.3333    -11.7118  -0.9549       5.1251       3.1991   12.5699

    DF     t Value     Pr > |t|

     5     -3.03       0.0292
```

Let μ_D = population mean of the differences (Brand A - Brand B).

- ♠ $H_0 : \mu_D = 0$ (cannot conclude a significant difference in mean lifelength between the two brands of tires)
- ♠ $H_a : \mu_D \neq 0$, (can conclude a significant difference in mean lifelength between the two brands of tires)
- ♠ p-value = 0.0292.
- ♠ Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$.

♠ Can one conclude there is a significant difference in the mean lifelength between the two brands of tires. Justify your answer!

Conclusion: Since the p-value = 0.0292 is (less than, greater than) [circle your choice] $\alpha = 0.05$, (reject, do not reject) [circle your choice] $H_0 \rightarrow$ it (is, is not) [circle your choice] reasonable to assume a significant difference in the mean lifelength between the two brands of tires.

Exit SAS and invoke R

2 Using R

Complete the following using R.

Open and execute your **R script** from Lab #4. Save it as **lab5_su19_XX**, where XX = initials of your name. Then move your cursor to the end of the script, so you can add more code.

NOTE: Be sure you update your title to Lab #5.

NOTE: Before proceeding, be sure to load the **TeachingDemos** package. This allows you to use the **t.test** function to generate confidence intervals and test of hypotheses for a single mean or the difference of two means and the **var.test** function to test equality of variances.

The general format is

```
t.test(x, y,alternative = "what", mu = diff, var.equal = FALSE,paired = FALSE
       conf.level = level)
var.test(x, y, ratio = 1,alternative = "what",conf.level = level)
```

where

what	greater, less, two.sided
diff	hypothesized difference between two means
var.equal	TRUE or FALSE (May be omitted for non-independent samples)
paired	TRUE for non-independent sample, FALSE for independent samples (default is FALSE, so this can be omitted for independent samples)
level	confidence level (0.90, 0.95, etc.)

1. Refer to SAS Question 1. Do the data provide sufficient evidence to indicate a significant difference in the mean amount of effluent discharged by Plant A and Plant B? (Recall: You must perform two tests for this problem. First you must check to see if it is reasonable to assume equality of variances. Second, based on the results of the test for equality of variances, you must the test the means.)

   (i) Is it reasonable to assume equality of variances? Justify your answer.

   To accomplish this task, add the following lines of code to the end of your **R script**:

```
# Test for equality of variances between Plant A and Plant B
# Use var.test function:
# var.test (Pop1,Pop2,ratio = value, alternative = "two-sided",conf.level = 0.95)
var.test(PlA,PlB,ratio = 1, alternative = "two.sided", conf.level = 0.95)
```

Make sure your cursor is in the **R Editor** window. Save your script and then

▲ highlight the new text you just typed.

▲ From the main menu, select **Edit** → **Run line or selection**.

Complete the following from the R Console window.:

```
> # Test for equality of variances between Plant A and Plant B
> # Use var.test function:
> # var.test (Pop1,Pop2,ratio = value, alternative = "two-sided",conf.level = 0.95)
> var.test(PlA,PlB,ratio = 1, alternative = "two.sided", conf.level = 0.95)
        F test to compare two variances
data:  PlA and PlB
F = 0.40566, num df = 9, denom df = 9, p-value = 0.1951
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1007611  1.6331985

sample estimates:
ratio of variances
     0.4056634
```

♣ $H_0 : \sigma_A^2 = \sigma_B^2$ (reasonable to assume equality of variances)

♣ $H_a : \sigma_A^2 \neq \sigma_B^2$ (not reasonable to assume equality of variances)

♣ p-value = 0.1951

♣ Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$.

♣ What is your conclusion? Can one assume equality of variances?

Conclusion: Since the p-value = 0.1951 is (less than, greater than) [circle your choice] $\alpha = 0.05$, (reject, do not reject) [circle your choice] H_0 → it (is, is not) [circle your choice] reasonable to assume equality of variances.

(ii) Now, test the means, using the results of part (i):

To accomplish this task, add the following lines of code to the end of your script.

```
# Use t.test to test for equality of means, assuming equal variances
# Use t.test function:
# t.test(Pop1,Pop2,alternative = "two-sided",mu = value_to_be_tested,
# var.equal = TRUE OR FALSE (whichever is appropriate), conf.level = 0.95)
t.test(PlA,PlB, alternative = "two.sided", mu = 0,var.equal = TRUE,conf.level = 0.95)
```

Make sure your cursor is in the **R Editor** window. Save your script and then

▲ highlight the new text you just typed.

▲ From the main menu, select **Edit** → **Run line or selection**.

Complete the following from the R Console window:

```
> # Use t.test to test for equality of means, assuming equal variances
> # Use t.test function:
> # t.test(Pop1,Pop2,alternative = "two-sided",mu = value_to_be_tested,
> # var.equal = TRUE OR FALSE (whichever is appropriate), conf.level = 0.95)
> t.test(PlA,PlB, alternative = "two.sided", mu = 0,var.equal = TRUE,conf.level = 0.95)
        Two Sample t-test
data:  PlA and PlB

t = -3.2567, df = 18, p-value = 0.004381
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.44253639 -0.09546361
sample estimates:
mean of x mean of y
  1.557     1.826
```

♠ $H_0 : \mu_A = \mu_B$ (cannot conclude a significant difference in mean effluent discharge between Plant A and Plant B)

♠ $H_a : \mu_A \neq \mu_B$ (can conclude a significant difference in mean effluent discharge between Plant A and Plant B)

♠ p-value = 0.004381.

♠ Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$.

♠ Can one conclude there is a significant difference in the mean amount of effluent discharge between Plant A and Plant B. Justify your answer!

Conclusion: Since the p-value = 0.004381 is (less than, greater than) [circle your choice] $\alpha = 0.05$, (reject, do not reject) [circle your choice] $H_0 \rightarrow$ it (is, is not) [circle your choice] reasonable to assume a significant difference in the mean amount of effluent discharged by Plant A and B.

2. (Your turn!) Refer to R Question 1. Do the data provide sufficient evidence to indicate a significant difference in the mean amount of effluent discharged by Plant A and Plant C? Be sure you write the commands you used to obtain your output.

   (i) Is it reasonable to assume equality of variances? Justify your answer.

Command(s):

var.test(PlA,PlC,ratio=1,alternative="two.sided",conf.level=0.95)

```
        F test to compare two variances

data:  PlA and PlC

F = 0.60438, num df = 9, denom df = 9, p-value = 0.4648
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:

 0.1501192  2.4332268

sample estimates:
ratio of variances

    0.6043791
```

- ♣ $H_0 : \sigma_A^2 = \sigma_C^2$ (reasonable to assume equality of variances)
- ♣ $H_a : \sigma_A^2 \neq \sigma_C^2$ (not reasonable to assume equality of variances)
- ♣ p-value = 0.4648
- ♣ Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$.
- ♣ What is your conclusion? Can one assume equality of variances?

Conclusion: Since the p-value = 0.4648 is (less than, greater than) [circle your choice] $\alpha = 0.05$, (reject, do not reject) [circle your choice] $H_0 \rightarrow$ it (is, is not) [circle your choice] reasonable to assume equality of variances.

(ii) Now, test the means, using the results of part (i).

Command(s):

t.test(PIA,PIC,alternative="two-sided",mu=0,var.equal=TRUE,conf.level=0.95)

```
        Two Sample t-test
data:  PlA and PlC

t = -0.35963, df = 18, p-value = 0.7233
alternative hypothesis: true difference in means is not equal to 0


95 percent confidence interval:
-0.1778908  0.1258908
sample estimates:
mean of x mean of y

 1.557   1.583
```

♠ $H_0 : \mu_A = \mu_C$ (cannot conclude a significant difference in mean effluent discharge between Plant A and Plant C)

- ♠ $H_a : \mu_A \neq \mu_C$ (can conclude a significant difference in mean effluent discharge between Plant A and Plant C)
- ♠ p-value = 0.7233.
- ♠ Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$.
- ♠ Can one conclude there is a significant difference in the mean amount of effluent discharge between Plant A and Plant C. Justify your answer!

  Conclusion: Since the p-value = 0.7233 is (less than, greater than) [circle your choice] $\alpha = 0.05$, (reject, do not reject) [circle your choice] $H_0 \rightarrow$ it (is, is not) [circle your choice] reasonable to assume a significant difference in the mean amount of effluent discharged by Plant A and Plant C.

NOTE: One can use the **t.test** function with the **paired = TRUE** options to test the difference of mean when the samples are not independent.

REMINDER: Since the samples are not independent, one must use the paired-difference t-test. You do not test for equality of variances!

3. Refer to SAS Question 2.

   NOTE: Let diff = BrandA - BrandB

   (i) Using **R** to generate the calculations, test whether there is a significant difference, on the average, between the life length of the two brands of tires.

   NOTE: The data has been saved in a file named **cartest1.dat**. The data file includes headings in Line 1. To accomplish this task, add the following lines of code to the end of your **R script**.

   NOTE: You will have to change the path to the location of your data file. If you saved it in your **RSpace** directory, you do not need to include the path.

```
# Use the read.table command to read in the data, use header = TRUE
# since the data starts on Line 2
cartest = read.table(file = "c:/Luke/summer2020/su19147/datafiles/cartest1.dat",header = TRUE)
# Print the data as a check
cartest
# Use the attach() function for accessibility of individual columns
attach(cartest)
Use the names() function to obtain column names
names(cartest)
```

   Make sure your cursor is in the **R Editor** window. Save your script and then

   ▲ highlight the new text you just typed.

   ▲ From the main menu, select **Edit** → **Run line or selection**.

   When your new output appears on the screen, have Luke, Lauren or your neighbor initial here. AZ

   (ii) Using **R** to generate the calculations, test whether there is a significant difference, on the average, between the life length of the two brands of tires.

   To accomplish this task, add the following lines of code to the end of your script.

```
# Use t.test with paired = TRUE to test difference of means
# Format: t.test(Variable1,Variable2,alternative = "you choice",
# mu = value of mean difference to be tested,
# paired = TRUE, conf.level = your_choice)
t.test(BrandA,BrandB, alternative = "two.sided", mu = 0, paired = TRUE, conf.level = 0.95)
```

Make sure your cursor is in the **R Editor** window. Save your script and then

▲ highlight the new text you just typed.

▲ From the main menu, select **Edit** → **Run line or selection**.

Complete the following from the R Console window.

```
> # Use t.test with paired = TRUE to test difference of means
> # Format: t.test(Variable1,Variable2,alternative = "your choice",
> # mu = value of mean difference to be tested,
> # paired = TRUE, conf.level = your_choice)
> t.test(BrandA,BrandB, alternative = "two.sided", mu = 0, paired = TRUE, conf.level = 0.95)
        Paired t-test
data:  BrandA and BrandB

t = -3.027, df = 5, p-value = 0.02918
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-11.711798  -0.954869

sample estimates:
mean of the differences
    -6.333333
```

Let μ_D = population mean of the differences (Brand A - Brand B).

♠ $H_0 : \mu_D = 0$ (cannot conclude a significant difference in mean lifelength between the two brands of tires)

♠ $H_a : \mu_D \neq 0$, (can conclude a significant difference in mean lifelength between the two brands of tires)

♠ p-value = 0.02918.

♠ Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$.

♠ Can one conclude there is a significant difference in the mean lifelength between the two brands of tires. Justify your answer!

Conclusion: Since the p-value = 0.02918 is (less than, greater than) [circle your choice] $\alpha = 0.05$, (reject, do not reject) [circle your choice] $H_0 \rightarrow$ it (is, is not) [circle your choice] reasonable to assume a significant difference in the mean lifelength between the two brands of tires.

(iii) (Your turn!) Using **R** to generate the calculations, test whether the lifelength of BrandA is significantly less than the lifelength of BrandB.

Command:

t.test(BrandA,BrandB,alternative="less",mu=0,paired=TRUE,conf.level=0.95)

Complete the following from the R Console window.

```
Paired t-test
data:  BrandA and BrandB

t = -3.027, df = 5, p-value = 0.01459
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
   -Inf   -2.117219
sample estimates:
mean of the differences
  -6.333333
```

Let μ_D = population mean of the differences (Brand A - Brand B).

- ♠ $H_0 : \mu_D = 0$ (cannot conclude the lifelength of BrandA is significantly less than the lifelength of BrandB)
- ♠ $H_a : \mu_D < 0$, (can conclude the lifelength of BrandA is significantly less than the lifelength of BrandB)
- ♠ p-value = 0.01459.
- ♠ Rejection Region: Reject H_0 if p-value $< \alpha = 0.05$.
- ♠ Can one conclude there is a significant difference in the mean lifelength between the two brands of tires. Justify your answer!

Conclusion: Since the p-value = 0.01459 is (less than, greater than) [circle your choice] $\alpha = 0.05$, (reject, do not reject) [circle your choice] $H_0 \rightarrow$ it (is, is not) [circle your choice] reasonable to assume the lifelength of BrandA is significantly less than the lifelength of BrandB.

NOTE: Be sure to save your script!

You have now successfully completed Lab #5. Please make sure your work area is neat and clean, log off your account and turn in your worksheet. Don't forget your flash drive, if you used one. Have a good day!

Luke & Ruihan