Statistics 147 In Class Exercise #5

Summer 2020; 10 pts

NAME: _____ **ID: (last 4 #s only)** _____

GOAL: (In class practice) Getting more acquainted with 2-sample confidence intervals and tests of hypothesis in **SAS** and **R**.

NOTE: You will need the following data file, which you should have downloaded for In Class Exercise #3 (July 3, 2020) and In Class Exercise #4 (July 10, 2020): **dograces.dat**

1 SAS

Invoke **SAS**.

1. Luke, Rachel and Luke have taken up dog-sled racing in hopes that someday they can enter The Iditarod Dog Sled Race in Alaska with *Trusty Dusty*, *White Shadow* and *Lakota Dakota* as their respective lead dogs. Since there is a lack of snow in Southern California, each have obtained a sled in which the runners have been replaced by wheels. After a significant number of practice runs, Luke, Rachel and Luke race every day (not necessarily together) for 3 months and record their time to finish the course (in minutes). Ruihan does not believe there is a significant difference in mean finishing times between the three teams. To test this claim, Ruihan takes three independent random samples of 10 days times for each of the three teams, yielding the following data:

```
Filename: dograces.dat
Dusty Shadow Lakota
45.5  43.6  64.9
59.2  59.9  66.2
38.4  39.8  56.5
68.8  70.4  75.7
51.9  50.2  55.2
47.4  48.9  61.5
41.6  40.3  49.9
58.9  58.0  63.5
60.7  60.9  65.0
47.0  46.5  62.6
```

NOTE:

- ♠ The data is located in a datafile named **dograces.dat**.
- ♠ The headings are included in the data file. The actual data begins on line 3.
- ♠ **Assume** the *Trusty Dusty* is dog team leader 1, *White Shadow* is dog team leader 2, *Lakota Dakota* is dog team leader 3.
- ♠ μ_i = true average finishing time for dog team i and σ_i^2 = variance of the finishing time for dog team i, $i = 1, 2, 3$

Open your SAS program file from the July 10th In Class Exercise #4, **inclass4_147_su19**. Save it as **inclass5_147_su19**. Change **title1** to

```
title1 'Statistics 147 In Class Exercise #5';
```

(i) Read in and print out the data using nested Do loops. (Be sure to give your columns the appropriate dog team leader names!) (Done on July 3rd.)

(ii) Sort the teams by the dog leader names. (Done on July 3rd.)

(iii) Generate the mean, standard deviation and variance for each of the three dog teams. (Done on July 3rd.)

(iv) Create a new SAS dataset, called **onlyDusty**, and bring in the data using the SET command. Use the appropriate IF structure to restrict the data to the **Trusty Dusty** dog team. Be sure to print the data! (Done on July 3rd.)

(v) Create a new SAS dataset, called **bothShadCody** and bring in the data using the SET command. Use the appropriate IF structure to restrict the data to the **White Shadow** and **Lakota Dakota** dog teams. Be sure to print the data! (Done on July 3rd.)

(vi) Use **proc means** to generate a **98%** confidence for **Trusty Dusty**. (Done on July 10th.)

(vi) Use **proc ttest** to generate the output needed to test whether the mean time for **Trusty Dusty** is less than 60 minutes. (Done on July 10th.)

(vii) **NEW** Refer to Part (v). Consider the **White Shadow** and **Lakota Dakota** teams.

(a) Is it reasonable to assume equality of variances?

To accomplish this goal, scroll down through your code until you are in the section where you created the new SAS dataset **bothShadCody**. Add the new code as follows:

```
/* Create new temporary SAS dataset to restrict attention to White Shadow and Lakota Dakota */
data bothShadCody;
/* Use set command to bring in all the data from the SAS dataset alldogs */
     set alldogs;
/* Revise title5 */
     title5 'Part (v)';
/* Use if statement to restrict the data to team 2 (White Shadow) and team 3 (Lakota Dakota) */
     if leader = 'White Shadow ' or leader = 'Lakota Dakota';
/* Or can use
     if team  = 2 or team = 3;  */
/* Print the results */
proc print;
/* ****************************************
    NEW CODE STARTS HERE
    ****************************************
*/
/* Use proc ttest  to generate test for equality
of variances and to test mu(LakotaDakota) not= mu(WhiteShadow)
   class classification variable
   var    response variable*/
proc ttest;
    title5 'Part (vii) (a) and (b)';
    class leader;
    var time;
```

Save and execute your program. When your output appears on the screen, complete the following:

```
Statistics 147 In Class Exercise #5
Summer 2020
Your name goes here
SAS Question 1
Part (vii) (a) and (b)
              Equality of Variances
Method      Num DF    Den DF    F Value    Pr > F


Folded F         9         9       2.04    ___________
```

▲ $H_0 : \sigma_2^2 = \sigma_3^2$ $(\sigma^2_{WhiteShadow} = \sigma^2_{LakotaDakota})$ ▲ $H_a : \sigma_2^2 \neq \sigma_3^2$ $(\sigma^2_{WhiteShadow} \neq \sigma^2_{LakotaDakota})$

▲ p-value = _____

▲ Rejection Region: Reject H_0 if p-value $< \alpha$.

▲ **Conclusion:** Since the p-value = _____ 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 (homogeneity) of variances.

(b) Assuming the results of subpart (a), perform the appropriate test of hypothesis to determine whether the mean time for the *White Shadow* team is significantly different from the mean time for the *Lakota Dakota* team.

NOTE: No new code is needed for this part.

Complete the following partilal output.

```
Method           Variances        DF    t Value       Pr > |t|

Pooled           Equal            18       2.63      _________

Satterthwaite    Unequal      16.111       2.63      _________
```

★ $H_0 :\ \mu_2 = \mu_3 \quad (\mu_{WhiteShadow} = \mu_{LakotaDakota})$

★ $H_a :\ \mu_2 \neq \mu_3 \quad (\mu_{WhiteShadow} \neq \mu_{LakotaDakota})$

★ p-value = _____

★ Rejection Region: Reject H_0 if p-value $< \alpha$.

★ **Conclusion:** Since the p-value = _____ 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 that the true mean time for the *Lakota Dakota* team is significantly different from the true mean time for the *White Shadow* team.

Exit SAS

2 R

Invoke R and complete the following.

1. Refer to SAS Question 1.

   (i) Read in and print out the data file, **dograces.dat**. Include the code to obtain the column names and to make the columns individually accessible. (Done on July 10th.)

   Open your R Script from the July 10th In Class Exercise #4, **inclass4_R_147_su19_XX**, where XX = initials of your name. Save your script as **inclass5_R_147_su19_XX**, where XX = initials of your name.

   Change the first title to

```
# Statistics 147 InClass Exercise #5 Summer 2020
```

   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.

   (ii) Using **R**, find and interpret a **97%** confidence interval for the true mean time for **Shadow**. (Done on July 10th.)

   (iii) Using **R** to complete the calculations, test the hypothesis that the true mean time for Lakota is significantly **greater than** 60 minutes. (Done on July 10th.)

   (iv) Consider the **White Shadow** and **Lakota Dakota** teams.

   (a) Is it reasonable to assume equality of variances?

   Add the following lines of code to the end of your script.

```
# Use var.test to test equal variances with ratio = 1 and alternative = "two.sided"
var.test(Shadow,Lakota,ratio = 1,alternative = "two.sided", conf.level = 0.95)
```

   Make sure your cursor is in the **R Editor** window. Save and then execute your script (from the main menu, select **Edit** $\rightarrow$ **All**).

   Complete the following from the R Console window.

```
> # Use var.test to test equal variances with ratio = 1 and alternative = "two.sided"
> var.test(Shadow,Lakota,ratio = 1,alternative = "two.sided", conf.level = 0.95)

        F test to compare two variances

data:  Shadow and Lakota

F = 2.0414, num df = 9, denom df = 9, p-value = ____________
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5070445 8.2184960
sample estimates:
ratio of variances
          2.041358
```

Complete the following test.

▲ $H_0 : \sigma_2^2 = \sigma_3^2$ ($\sigma^2_{WhiteShadow} = \sigma^2_{LakotaDakota}$) ▲ $H_a : \sigma_2^2 \neq \sigma_3^2$ ($\sigma^2_{WhiteShadow} \neq \sigma^2_{LakotaDakota}$)

▲ p-value = _____

▲ Rejection Region: Reject H_0 if p-value $< \alpha$.

▲ **Conclusion:** Since the p-value = _____ 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 (homogeneity) of variances.

(b) Assuming the results of subpart (a), perform the appropriate test of hypothesis to determine whether the mean time for the *Lakota Dakota* team is significantly **greater** than the mean time for the *White Shadow* team.

Add the following lines of code to the end of your script.

```
# ####################################################################
# Use t.test(variable1,variable2, alternative = "greater", var.equal=TRUE ,
# conf.level = 0.95) to generate
# output for testing mu(Lakota) > mu(Shadow)
t.test(Lakota,Shadow,alternative = "greater", 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(variable1,variable2, alternative = "greater", var.equal=TRUE ,
> # conf.level = 0.95) to generate
> # output for testing mu(Lakota) > mu(Shadow)
> t.test(Lakota,Shadow,alternative = "greater", var.equal = TRUE,conf.level = 0.95)

        Two Sample t-test

data:  Lakota and Shadow
t = 2.6332, df = 18, p-value = _________________
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 3.499898      Inf
sample estimates:
mean of x mean of y
    62.10     51.85
> # ####################################################################
```

Complete the following test.

★ $H_0: \ \mu_3 = \mu_2 \quad (\mu_{LakotaDakota} = \mu_{WhiteShadow})$

★ $H_a: \ \mu_3 > \mu_2 \quad (\mu_{LakotaDakota} > \mu_{WhiteShadow})$

★ p-value = _____

★ Rejection Region: Reject H_0 if p-value $< \alpha$.

★ **Conclusion:** Since the p-value = _____ 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 that the true mean time for the *Lakota Dakota* team is significantly greater than the true mean time for the *White Shadow* team.

You have now successfully completed **In Class Exercise #5**. Please turn in this worksheet. Be sure to log off your account, take your flash drive (if you used one) and make sure that your work area is neat and clean. Have a nice day!

Luke & Ruihan