Statistics 147 In Class Exercise #4

Summer 2020; 10 pts

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

GOAL: (In class practice) Getting more acquainted with 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): **dograces.dat**

1 SAS

Invoke **SAS**.

1. Luke, Rachel and Bentley 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 Since there is a lack of snow in Southern California, each have obtained a sled in which the runners have been replaced by wheels.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 Bentley 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 3rd In Class Exercise #3, **inclass3_147_su20**. Save it as **inclass4_147_su20**. Change **title1** to

```
title1 'Statistics 147 In Class Exercise #4';
```

(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) **NEW** Use **proc means** to generate a **98%** confidence for **Trusty Dusty**.

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

```
/* Create new temporary SAS dataset to restrict attention to Trusty Dusty */
data onlyDusty;
    /* Use set command to bring in all the data from the SAS dataset alldogs */
    set alldogs;

    /* Use if statement to restrict the data to team 1 (Trusty Dusty) */
     if leader = 'Trusty Dusty ';
       /* Or can use  if team = 1; */
run;

/* Print the results */
proc print;
     /* Revise title5 */
     title5 'Part (iv)';
run;
/* ****************************************
      NEW CODE BEGINS HERE
     **************************************** */
/* Use proc means to generate confidence interval
   Specify value of alpha to use: 98% -> alpha = 0.02 */
proc means n mean stddev clm alpha = 0.02;
   title5  'Part (vi) 98% CI using proc means';
   var time;
run;
/* ****************************************
    END NEW CODE
    ****************************************  */
```

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

```
Statistics 147 In Class Exercise #4
Summer 2020
Your name goes here
SAS Question 1
Part (vi) 98% CI using proc means

The MEANS Procedure
```

```
                    Analysis Variable : time

                                                  Lower 98%       Upper 98%
 N            Mean          Std Dev      CL for Mean      CL for Mean
 20    59.9750000    9.9711624    51.3129183       62.637
```

Confidence Intervals Limits: 51.32, 62.64

Interpretation:

We can be confident based on the data that 98% of observations will fall into the interval between 51.32 and 62.64

(vii) **NEW** Use **proc ttest** to generate the output needed to teat whether the mean time for **Trusty Dusty** is less than 60 minutes.

To accomplish this task, add the following lines of code, right after the new code you just added:

```
/* Use proc ttest to test single mean; specify value to test: h0 = value_to_test
   Use sides = lower to generate test for mu_Dusty < 60 */
proc ttest h0 = 60 sides = lower;
   title5 'Part (iv) Using proc ttest to test mean = 60 vs mean < 60';
   var time;
run;
```

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

```
Statistics 147 In Class Exercise #4
Summer 2020
Your name goes here
SAS Question 1
Part (iv) Using proc ttest to test mean = 60 vs mean < 60

The TTEST Procedure
Variable:  time

  N        Mean     Std Dev     Std Err     Minimum     Maximum
 10     51.9400      9.6545      3.0530     38.4000     68.8000

    Mean       95% CL Mean        Std Dev      95% CL Std Dev
 51.9400    -Infty    57.5365      9.6545      6.6407  17.6254

    DF      t Value      Pr < t
     9        -2.64      0.0135
```

Recall: μ_1 = true mean of **Trusty Dusty**.

- ♠ $H_0 : \mu_1 = 60$
- ♠ $H_a : \mu_1 > 60$

♠ p-value = 0.0135.

♠ RR: Reject H_0 if p-value $< \alpha = 0.05$

♠ What is your conclusion? (Be sure to justify your answer!)

Conclusion: Since the p-value = 0.0135 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 true mean time for **Trusty Dusty** is significantly less than 60 minutes.

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 amke the columns individually accessible.

   Let's use an *R script* to enter our commands. Open **R**. From the main menu select **File → New script**. The **R Editor** window will open. (It will say *untitled* until you save the script.)

   ★ Move the cursor to the **R Editor** window and type in the following:

```
# Statistics 147 InClass Exercise #4 Summer 2020
# Your name goes here
#
# R Question 1
# Use the read.table command to read in the data
# The data starts on Line 3, so use skip = 1
# format: read.table(file = "filename including path",header = TRUE, skip = 1)
# Be sure to change the path to your data file.
# Type the following two lines all on one line
dog_data = read.table(file = "c:/Luke/summer2020/su20147/datafiles/
dograces.dat",header= TRUE, skip = 1)
# Print the data as a check
dog_data
# Use attach command to get access to individual columns
attach(dog_data)
# Use the names() function to obtain column names
names(dog_data)
```

   Make sure your cursor is in the **R Editor** window.

   ▲ To **save** your script,from the main menu, select **File → Save As**. Select the location where you would like to save your script and type **inclass4_R_147_su20_XX**, where XX = initials of your name.

   ▲ To execute your script, from the main menu, select **Edit → All**.

You should see everything you typed, plus the data, in the **R Console** window. When you see the data, have Ruihan, Luke or your neighbor initial here. _____

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

Add the following lines of code to your script.

```
# Generate 97% CI for Shadow
# Use t.test
# Format: t.test(name_of_variable,alternative = appropriate option,
# conf.level = confidence-level-in-decimal-format)
t.test(Shadow,alternative="two.sided",conf.level= 0.97)
```

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.

```
> t.test(Shadow,alternative="two.sided",conf.level= 0.97)
        One Sample t-test
data:  Shadow
t = 16.258, df = 9, p-value = 5.593e-08
alternative hypothesis: true mean is not equal to 0
97 percent confidence interval:
```

43.64182 60.05818

```
sample estimates:
mean of x
```

51.85

Interpretation:

We can be confident based on the data that 97% of observations will fall into the interval between 43.64 and 60.06

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

Add the following lines of code to your script.

```
# Test mu(Lakota) > 60
# Use t.test
# Format: t.test(name_of_variable,alternative = appropriate option,
# conf.level = confidence-level-in-decimal-format)
t.test(Lakota,alternative="greater",mu = 60, 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.

```
        One Sample t-test
data:  Lakota
t = 0.94082, df = 9, p-value = 0.1857
alternative hypothesis: true mean is greater than 60
```

Recall: μ_3 = true mean time for **Lakota**

- ♠ H_0 : $\mu_3 = 60$
- ♠ H_a : $\mu_3 > 60$
- ♠ p-value = 0.1857.
- ♠ RR: Reject H_0 if p-value $< \alpha = 0.05$
- ♠ What is your conclusion? (Be sure to justify your answer!)

  Conclusion: Since the p-value = 0.1857 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 true mean time for **Lakota** is significantly greater than 60 minutes.

You have now successfully completed **In Class SAS Practice #4**. 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