**Statistics 147 Assignment #4**

**Summer 2020**

**Wesley Chang**

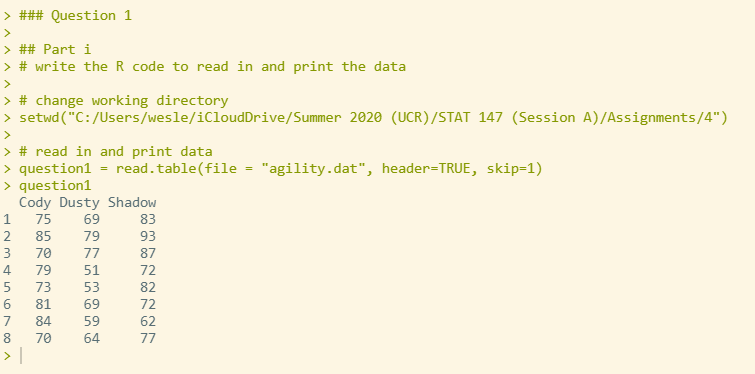
**0996**

**2 The Questions**

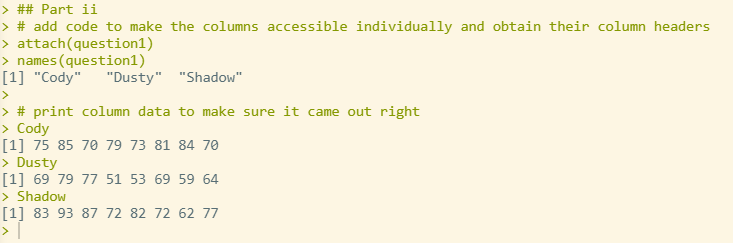
# 2.1 R

Question 1:

* 1. Write the R code to read in and print out the data. (Done as part of Assignment #2.)

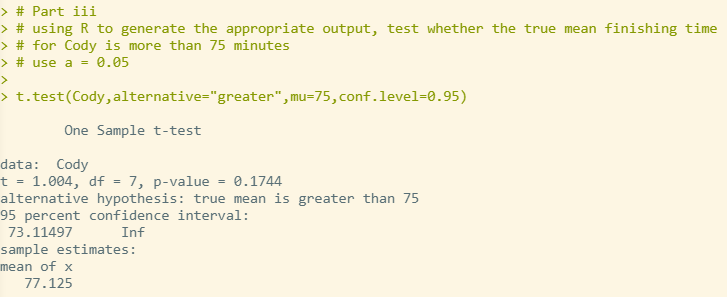


* 1. Add the appropriate lines of code to make the columns accessible individually and obtain the column headers. (Done as part of Assignment #2.)



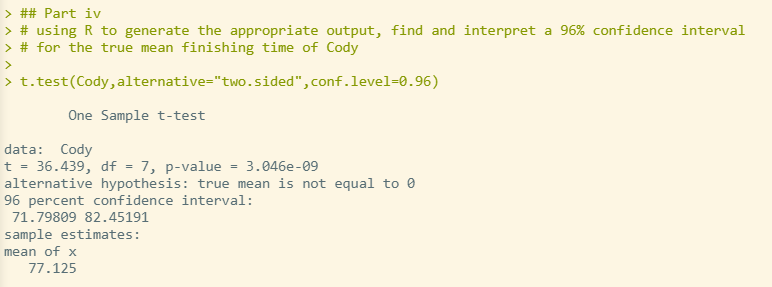
* 1. Using **R** to generate the appropriate output, test whether the true mean finishing time for **Cody** is more than 75 minutes. Use *α* = 0*.*05. (Done as part of Assignment #3.)

**The p value is 1.523e-09, which is less than the alpha = 0.05. We reject the null hypothesis, that the true mean is equal to 75 in favor of the alternative hypothesis that the true mean is greater than 75. We can conclude that the true mean finishing time for Cody is more than 75 minutes.**



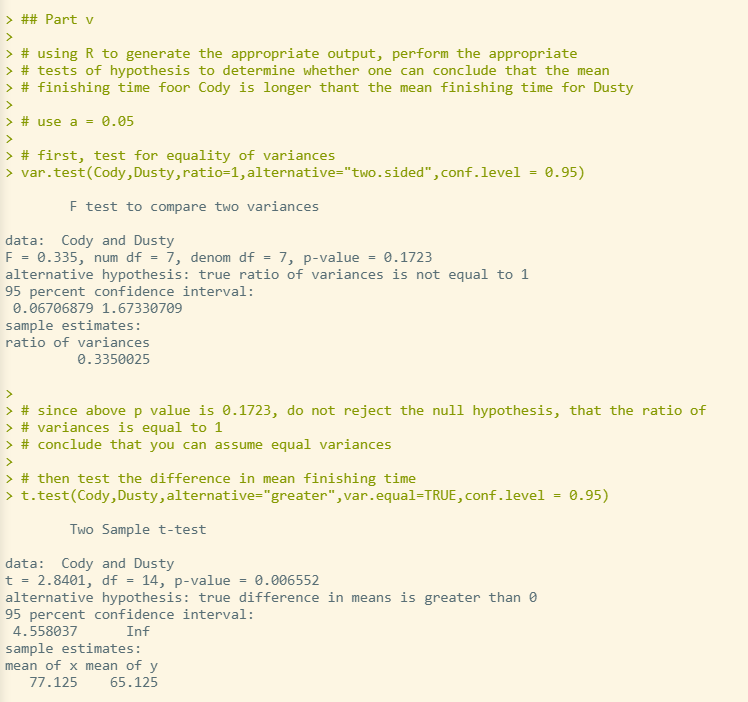
* 1. Using **R** to generate the output, find and interpret a 96% confidence interval for the true mean finishing time of **Cody**. (Done as part of Assignment #3.)

**The 96% confidence interval for the true mean finishing time for Cody is between 71.79809 and 82.45191.**



* 1. **[NEW]** Using **R** to generate the appropriate output, perform the appropriate test(s) of hypothesis to determine whether one can conclude that the mean finishing time for **Cody** is longer than the mean finishing time for **Dusty**. Use *α* = 0*.*05. (6 pts)

**The p-value of the test for difference in mean finishing time is 0.006552, so we reject the null hypothesis that the true difference in means is equal to 0 in favor of the alternative hypothesis that the true difference in means is greater than 0. We can concluded that the mean finishing time for Cody is longer than the mean finishing time for Dusty.**

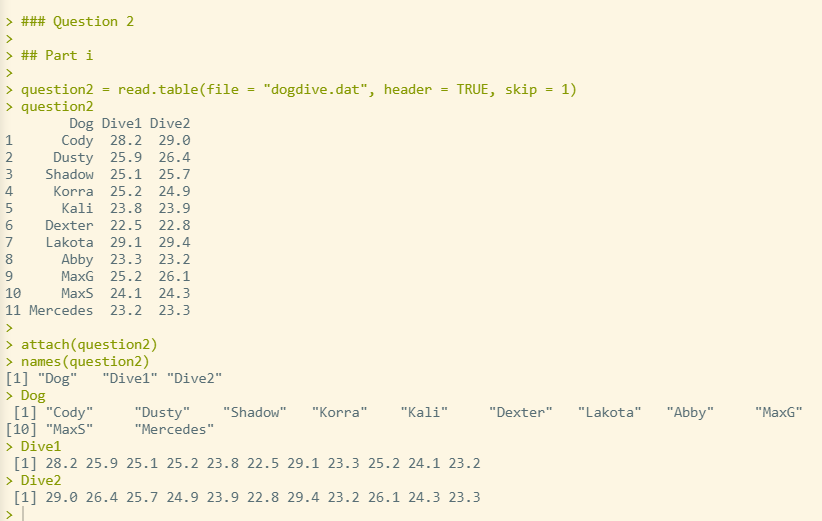


* 1. Is the above hypothesis test paired or un-paired? Please explain *why*. (3 pts)

**The above hypothesis test is an unpaired test, as the data for each time was gathered independently using a random sampling. If it were a paired test, the times for each observation would not be independent of each other.**

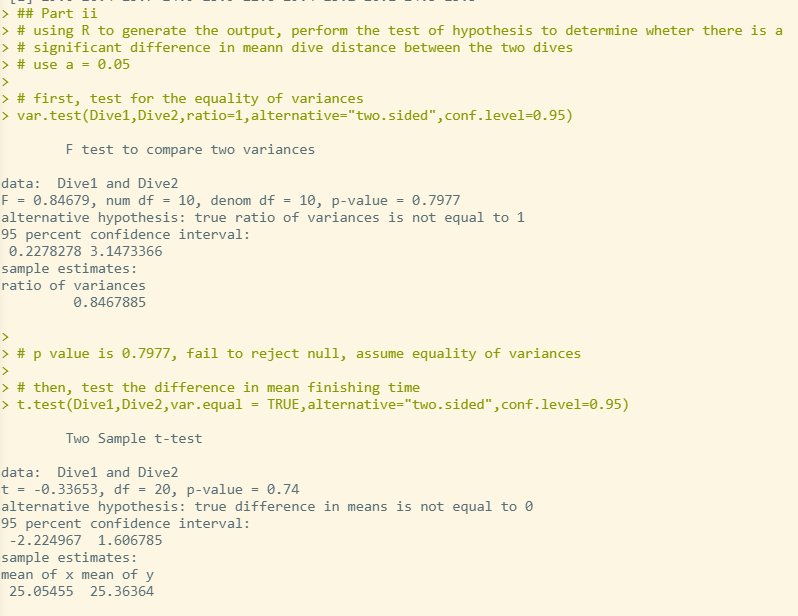
Question 2:

1. Write the R code to read in and print out the data. Be sure to add the line(s) of code to make the columns accessible individually. (2 pts)



1. Using **R** to generate the appropriate output, perform the test of hypothesis to determine whether there is a significant difference in mean dive distance between the two dives. Use *α* = 0*.*05. (5 pts)

**The p-value of the two-sample t-test is 0.74, so we fail to reject the null hypothesis. We reject the null hypothesis that the true difference in means is equal to 0 in favor of the two-sided alternative. We can conclude that there is a significant difference in mean dive distance between the two dives.**



1. Is the above hypothesis test paired or un-paired? Please explain *why*. (3 pts)

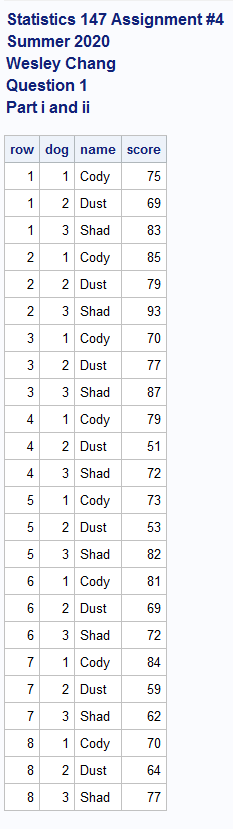
**The above hypothesis test is a paired test, as the times for each dog are gathered in order. For example, the third observation for Dive1 and Dive2 correspond to the third dog on the list. Dive1 and Dive2 are not independent, and therefore this test must be a paired test.**

# 2.2 SAS

1. **(9 pts total)** Refer to **R** Question 1.
   * 1. Write the SAS code to read in and print out the data. Use **agility1** as your temporary SAS dataset name. Use nested do loops! **(NOTE: DO NOT COPY AND PASTE THE DATA INTO YOUR SAS PROGRAM. READ THE DATA IN FROM THE EXTERNAL DATA FILE!)** (Done as part of

Assignment #2.)





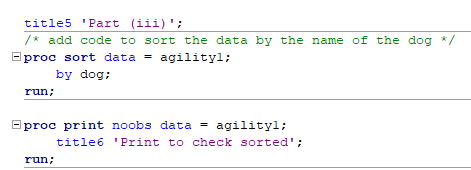
* + 1. Using if-then-else structures, name the dogs as follows: (Done as part of Assignment #2.)

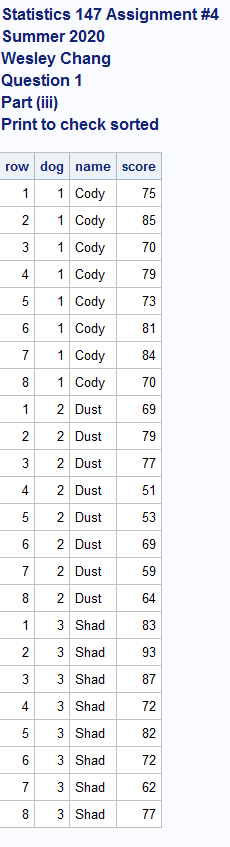
|  |  |  |  |
| --- | --- | --- | --- |
| **dog** | 1 | 2 | 3 |
| **name** | **Cody** | **Dusty** | **Shadow** |

**See above**

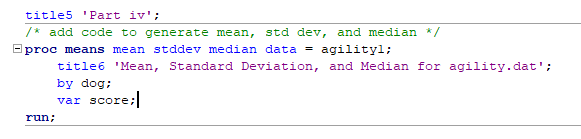
* + 1. Add the appropriate lines of code to sort the data by the **name** of the dog. (Done as part of Assignment

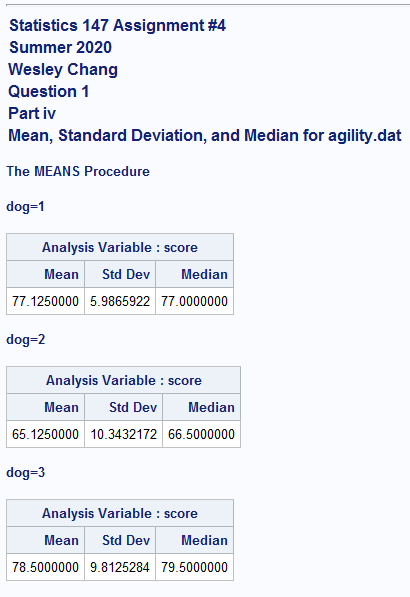
#2.)



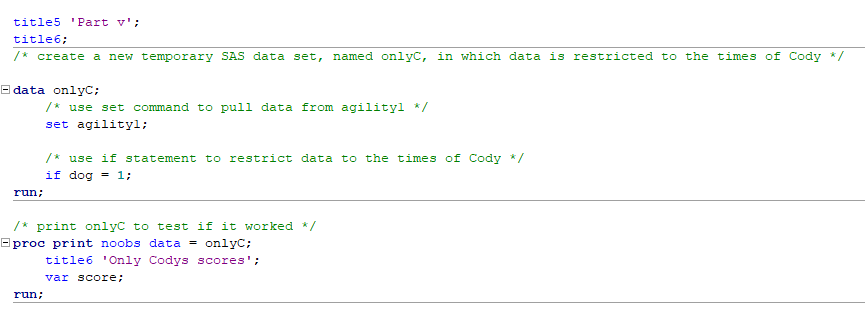


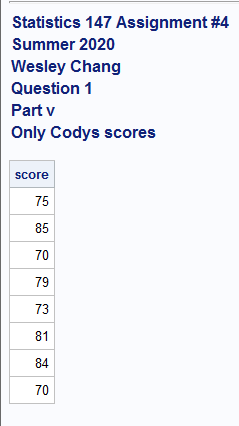
* + 1. Add the appropriate lines of code to your program to generate the mean, standard deviation and median for each of the dogs. (Done as part of Assignment #2.)





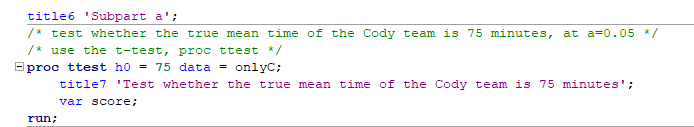
* + 1. Create a new temporary SAS data set, named **onlyC**, in which the data is restricted to the times of *Cody*. (Be sure to print the data as a check.) (Done as part of Assignment #2.)

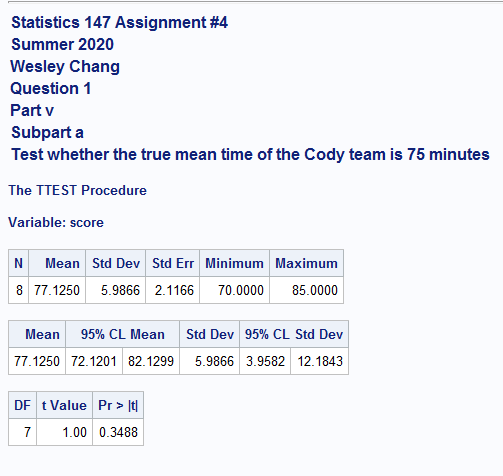




* + 1. Test whether the true mean time of the **Cody** team is 75 minutes. Use *α* = 0*.*05. (Done as part of Assignment #3.)

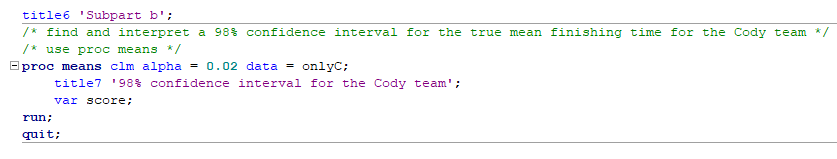
**The p-value is 0.3488, so we fail to reject the null hypothesis that the true mean time of the Cody team is 75 minutes.**

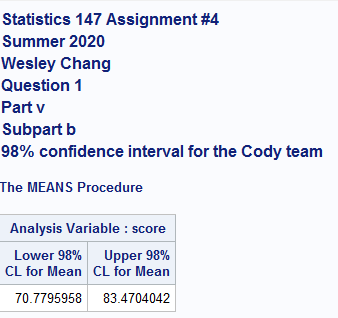




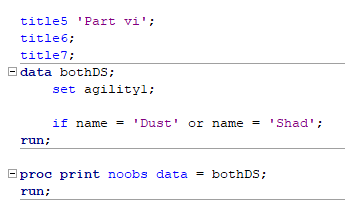
* + 1. Find and interpret a 98% confidence interval for the true mean finishing time for the *Cody* team. (Done as part of Assignment #3.)

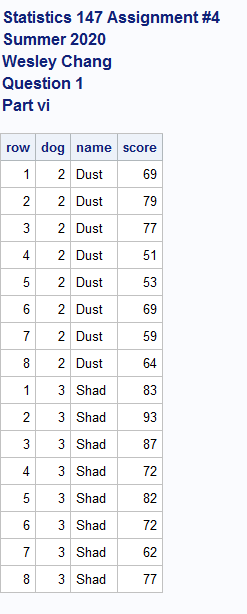
**The 98% confidence interval for the true mean finishing time for the Cody team is between 70.7779958 and 83.4704042.**





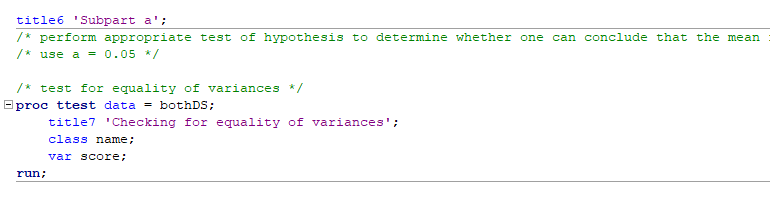
* + 1. Create a new temporary SAS data set, named **bothDS** in which the data is restricted to the times of **Dusty** and **Shadow**. (Be sure to print the data as a check.) (Done as part of Assignment #2.)

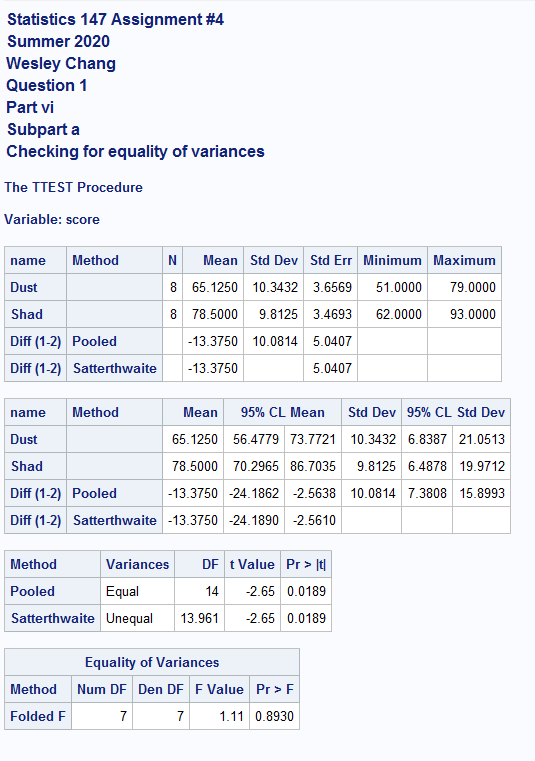


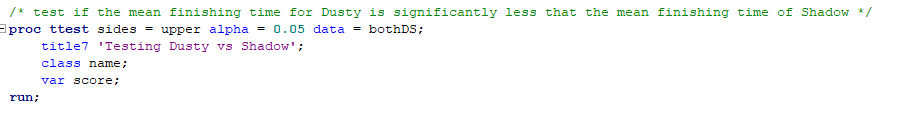


* + 1. **[NEW]** Perform the appropriate test(s) of hypothesis to determine whether one can conclude the mean finishing time of **Dusty** is significantly less than the mean finishing time of **Shadow**. Use *α* = 0*.*05. (6 pts)

**The p value for the equality of variances test is 0.8930, so we fail to reject the null hypothesis that the variances are equal. Assuming equal variance, we use the p-value for the pooled test, which is 0.9905, so we fail to reject the null hypothesis that the mean finishing time of Dusty is less than the mean finishing time of Shadow.**





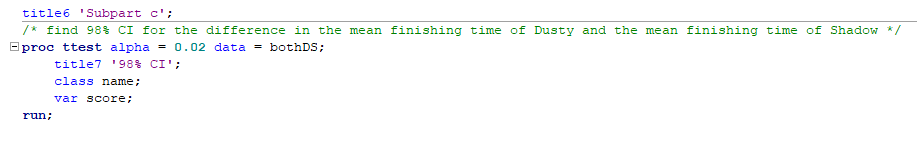


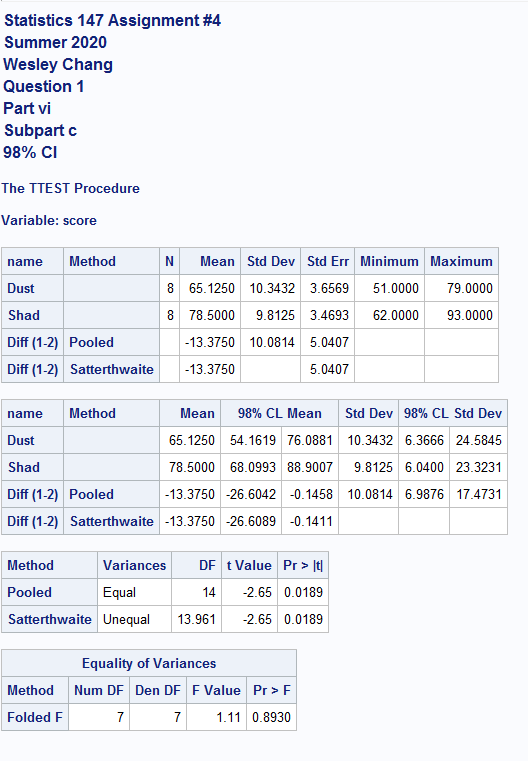


* + 1. **[NEW]** Is the above hypothesis test paired or un-paired? Please explain *why*. (2 pts)

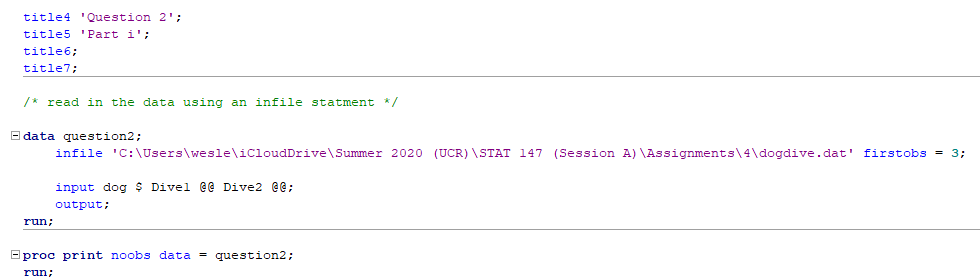
**The above hypothesis test is an unpaired test, as the data for each time was gathered independently using a random sampling. If it were a paired test, the times for each observation would not be independent of each other.**

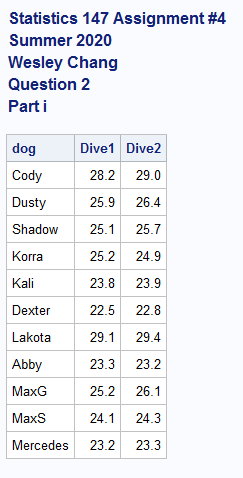
* + 1. **[NEW]** Find a 98% confidence interval for the difference in the mean finishing time of **Dusty** and the mean finishing time of **Shadow**. Based on this interval, can one conclude there is a significant difference between the mean finishing time of Dusty and the mean finishing time of **Shadow**? Justify your answer! (3 pts)



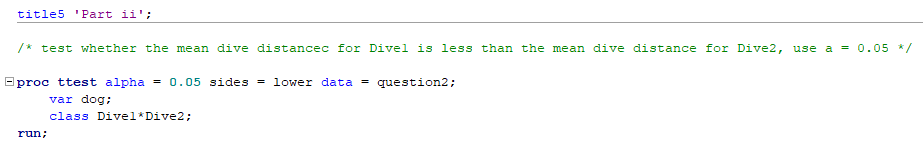


1. **[NEW] (6 pts total)** Refer to R Question 2.
   1. Read in the data using an infile statement. (Do not use DO loops! Be sure to print your data as a check!) (4 pts)





* 1. Perform the test of hypothesis to determine whether the mean dive distance for **Dive1** is **less** than the mean dive distance for **Dive2**. Use *α* = 0*.*05. (5 pts)



* 1. Is the above hypothesis test paired or un-paired? Please explain *why*. (1 pts)

**The above hypothesis test is a paired test, as the times for each dog are gathered in order. For example, the third observation for Dive1 and Dive2 correspond to the third dog on the list. Dive1 and Dive2 are not independent, and therefore this test must be a paired test.**

R code:

# Statistics 147 Assignment #4

# Summer 2020

# Wesley Chang

### Question 1

## Part i

# write the R code to read in and print the data

# change working directory

setwd("C:/Users/wesle/iCloudDrive/Summer 2020 (UCR)/STAT 147 (Session A)/Assignments/4")

# read in and print data

question1 = read.table(file = "agility.dat", header=TRUE, skip=1)

question1

## Part ii

# add code to make the columns accessible individually and obtain their column headers

attach(question1)

names(question1)

# print column data to make sure it came out right

Cody

Dusty

Shadow

# Part iii

# using R to generate the appropriate output, test whether the true mean finishing time

# for Cody is more than 75 minutes

# use a = 0.05

t.test(Cody,alternative="greater",mu=75,conf.level=0.95)

## Part iv

# using R to generate the appropriate output, find and interpret a 96% confidence interval

# for the true mean finishing time of Cody

t.test(Cody,alternative="two.sided",conf.level=0.96)

## Part v

# using R to generate the appropriate output, perform the appropriate

# tests of hypothesis to determine whether one can conclude that the mean

# finishing time foor Cody is longer thant the mean finishing time for Dusty

# use a = 0.05

# first, test for equality of variances

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

# since above p value is 0.1723, do not reject the null hypothesis, that the ratio of

# variances is equal to 1

# conclude that you can assume equal variances

# then test the difference in mean finishing time

t.test(Cody,Dusty,alternative="greater",var.equal=TRUE,conf.level = 0.95)

### Question 2

## Part i

question2 = read.table(file = "dogdive.dat", header = TRUE, skip = 1)

question2

attach(question2)

names(question2)

Dog

Dive1

Dive2

## Part ii

# using R to generate the output, perform the test of hypothesis to determine wheter there is a

# significant difference in meann dive distance between the two dives

# use a = 0.05

# first, test for the equality of variances

var.test(Dive1,Dive2,ratio=1,alternative="two.sided",conf.level=0.95)

# p value is 0.7977, fail to reject null, assume equality of variances

# then, test the difference in mean finishing time

t.test(Dive1,Dive2,var.equal = TRUE,alternative="two.sided",conf.level=0.95)

SAS Code:

title1 'Statistics 147 Assignment #4';

title2 'Summer 2020';

title3 'Wesley Chang';

options nocenter ps = **55** nocenter ls = **78** nodate nonumber formdlim='\*';

DM log "odsresults; clear; out; clear; log; clear;";

ods graphics off;

title4 'Question 1';

title5 'Part i and ii';

**data** agility1;

infile 'C:\Users\wesle\iCloudDrive\Summer 2020 (UCR)\STAT 147 (Session A)\Assignments\4\agility.dat' firstobs = **3**;

/\* do loop for rows \*/

do row = **1** to **8**;

/\* do loop for columns \*/

do dog = **1** to **3**;

/\* if then structure to name dogs \*/

if dog = **1** then name = 'Cody';

else if dog = **2** then name = 'Dusty';

else name = 'Shadow';

/\* input statement \*/

input score @@;

output;

end;

end;

**run**;

/\* print out results \*/

**proc** **print** noobs data = agility1;

**run**;

title5 'Part (iii)';

/\* add code to sort the data by the name of the dog \*/

**proc** **sort** data = agility1;

by dog;

**run**;

**proc** **print** noobs data = agility1;

title6 'Print to check sorted';

**run**;

title5 'Part iv';

/\* add code to generate mean, std dev, and median \*/

**proc** **means** mean stddev median data = agility1;

title6 'Mean, Standard Deviation, and Median for agility.dat';

by dog;

var score;

**run**;

title5 'Part v';

title6;

/\* create a new temporary SAS data set, named onlyC, in which data is restricted to the times of Cody \*/

**data** onlyC;

/\* use set command to pull data from agility1 \*/

set agility1;

/\* use if statement to restrict data to the times of Cody \*/

if dog = **1**;

**run**;

/\* print onlyC to test if it worked \*/

**proc** **print** noobs data = onlyC;

title6 'Only Codys scores';

var score;

**run**;

title6 'Subpart a';

/\* test whether the true mean time of the Cody team is 75 minutes, at a=0.05 \*/

/\* use the t-test, proc ttest \*/

**proc** **ttest** h0 = **75** data = onlyC;

title7 'Test whether the true mean time of the Cody team is 75 minutes';

var score;

**run**;

title6 'Subpart b';

/\* find and interpret a 98% confidence interval for the true mean finishing time for the Cody team \*/

/\* use proc means \*/

**proc** **means** clm alpha = **0.02** data = onlyC;

title7 '98% confidence interval for the Cody team';

var score;

**run**;

**quit**;

title5 'Part vi';

title6;

title7;

**data** bothDS;

set agility1;

if name = 'Dust' or name = 'Shad';

**run**;

**proc** **print** noobs data = bothDS;

**run**;

title6 'Subpart a';

/\* perform appropriate test of hypothesis to determine whether one can conclude that the mean finishing time of Dusty is significantly less than the mean finishing time of Shadow \*/

/\* use a = 0.05 \*/

/\* test for equality of variances \*/

**proc** **ttest** data = bothDS;

title7 'Checking for equality of variances';

class name;

var score;

**run**;

title6 'Subpart b';

/\* test if the mean finishing time for Dusty is significantly less that the mean finishing time of Shadow \*/

**proc** **ttest** sides = upper alpha = **0.05** data = bothDS;

title7 'Testing Dusty vs Shadow';

class name;

var score;

**run**;

title6 'Subpart c';

/\* find 98% CI for the difference in the mean finishing time of Dusty and the mean finishing time of Shadow \*/

**proc** **ttest** alpha = **0.02** data = bothDS;

title7 '98% CI';

class name;

var score;

**run**;

title4 'Question 2';

title5 'Part i';

title6;

title7;

/\* read in the data using an infile statment \*/

**data** question2;

infile 'C:\Users\wesle\iCloudDrive\Summer 2020 (UCR)\STAT 147 (Session A)\Assignments\4\dogdive.dat' firstobs = **3**;

input dog $ Dive1 @@ Dive2 @@;

output;

**run**;

**proc** **print** noobs data = question2;

**run**;

title5 'Part ii';

/\* test whether the mean dive distancec for Dive1 is less than the mean dive distance for Dive2, use a = 0.05 \*/

**proc** **ttest** alpha = **0.05** sides = lower data = question2;

var dog;

class Dive1\*Dive2;

**run**;