**Statistics 147 Assignment #3**

**Summer 2020**

**Wesley Chang**

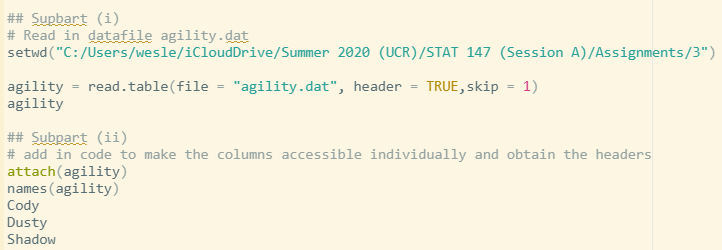
**0996**

**The Questions**

**R**

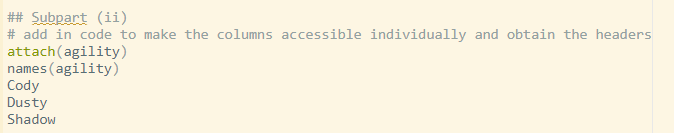
1. **(10 pts total)** (Use **R** for this problem!) Linda has three dogs (Cody, Dusty and Shadow) that she is training for a national agility championship. Linda records their times to finish the course (in seconds) for 30 runs. Lauren does not believe there is a significant difference in mean finishing times between the three dogs. To test this claim, Lauren takes three independent random samples of 8 times for each of the three dogs, yielding the following data:

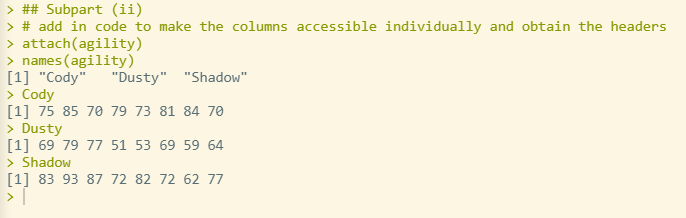
1. Write the R code to read in and print out the data. (Done as part of Assignment #2.)





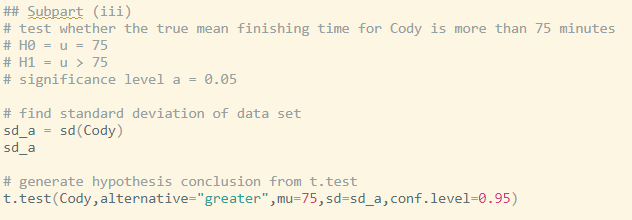
1. Add the appropriate lines of code to make the columns accessible individually and obtain the column headers. (Done as part of Assignment #2.)

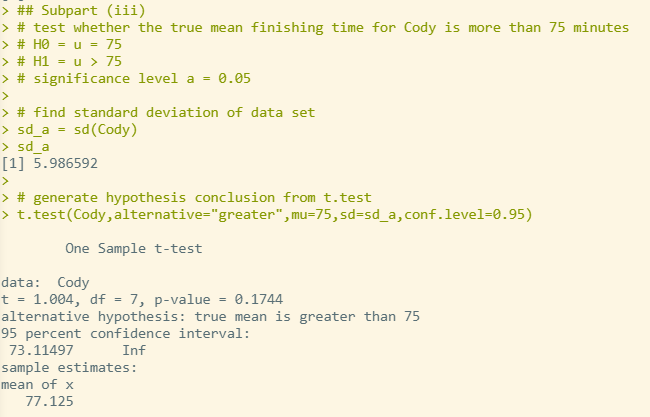




1. **[NEW]** Using **R** to generate the appropriate output, test whether the true mean finishing time for **Cody** is more than 75 minutes. Use *α* = 0*.*05. (5 pts)

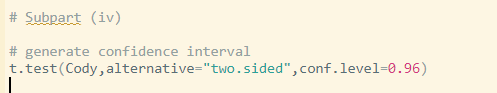
**The p-value 0.1744 is less than the alpha = 0.05, reject the null hypothesis in favor of H1. The true mean finishing time for Cody is statistically more than 75.**

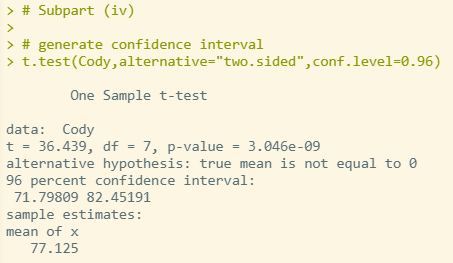




1. **[NEW]** Using **R** to generate the output, find and interpret a 96% confidence interval for the true mean finishing time of **Cody**. (5 pts)

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





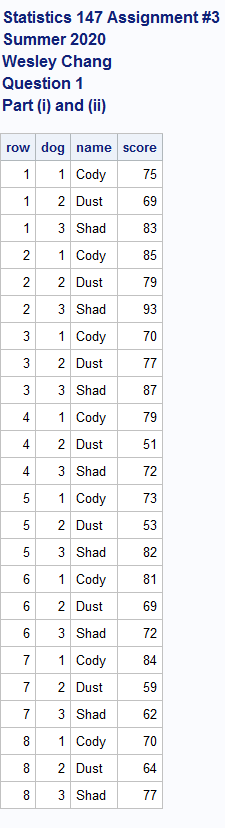
**SAS**

1. **(10 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.)

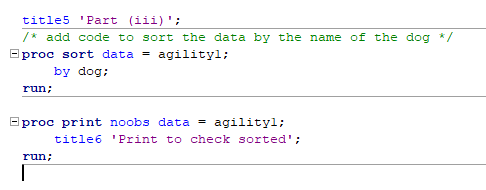


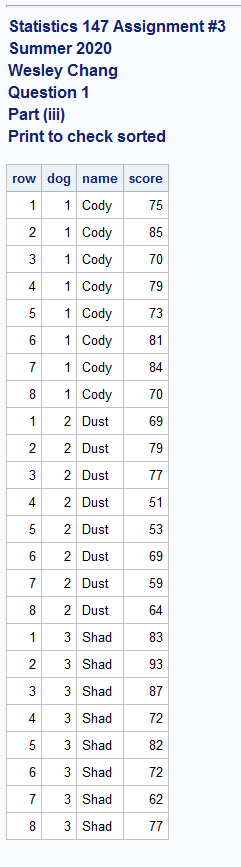


1. Using if-then-else structures, name the dogs as follows: (Done as part of Assignment #2.)

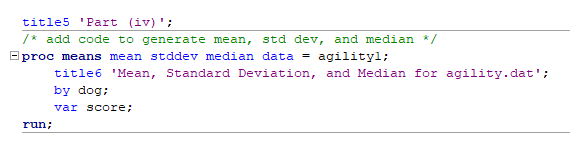
**See answer for part (i)**

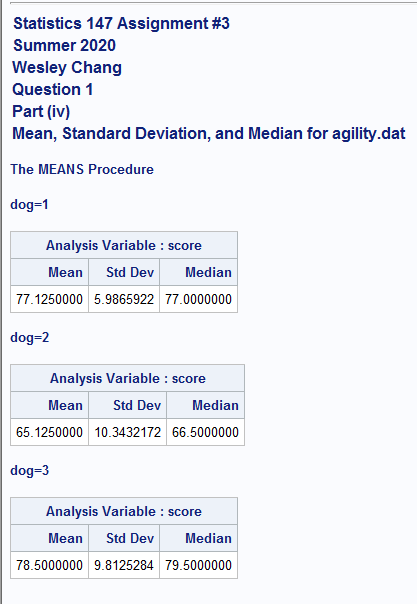
1. Add the appropriate lines of code to sort the data by the **name** of the dog. (Done as part of Assignment #2.)





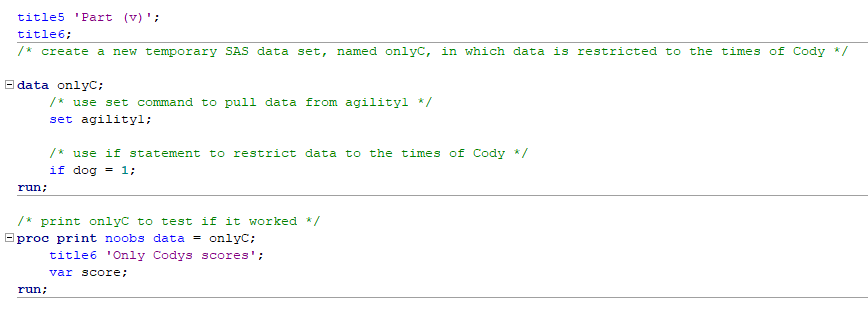
1. Add the appropriate lines of code to your program to generate the mean, standard deviation and median for each of the dogs. (Done as part of Assignment #2.)

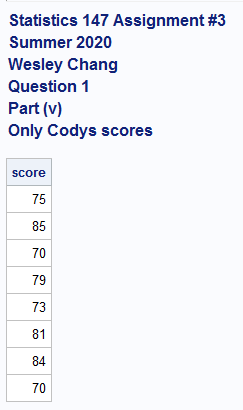




1. Create a new temporary SAS data set, named **onlyC**, in which the data is restricted to the times of

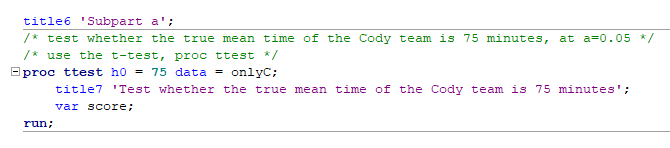
*Cody*. (Be sure to print the data as a check.) (Done as part of Assignment #2.)

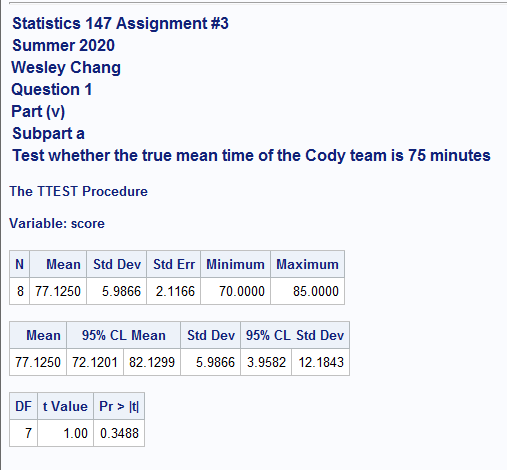




1. **[NEW]** Test whether the true mean time of the **Cody** team is 75 minutes. Use *α* = 0*.*05. (5 pts)

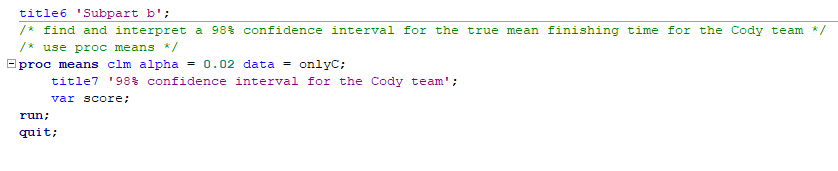
**Based of this data, we get the p-value 0.3488, which is less that the significance level a = 0.05. This means that we reject the null hypothesis and can conclude that the true mean time of the Cody team is statistically not 75 minutes.**

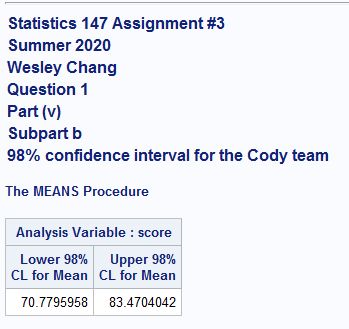




1. **[NEW]** Find and interpret a 98% confidence interval for the true mean finishing time for the *Cody* team. (5 pts)

**The confidence interval is (70.7795958, 83.4704042) at 98% confidence. We can statistically conclude with 98% confidence that the true mean finishing time for the Cody team falls within this range.**





**R code:**

# Statistics 147 Assignment #3

# Summer 2020

# Wesley Chang

# Question 1

## Supbart (i)

# Read in datafile agility.dat

setwd("C:/Users/wesle/iCloudDrive/Summer 2020 (UCR)/STAT 147 (Session A)/Assignments/3")

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

agility

## Subpart (ii)

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

attach(agility)

names(agility)

Cody

Dusty

Shadow

## Subpart (iii)

# test whether the true mean finishing time for Cody is more than 75 minutes

# H0 = u = 75

# H1 = u > 75

# significance level a = 0.05

# find standard deviation of data set

sd\_a = sd(Cody)

sd\_a

# generate hypothesis conclusion from t.test

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

# Subpart (iv)

# generate confidence interval

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

**SAS code:**

title1 'Statistics 147 Assignment #3';

title2 'Summer 2020';

title3 'Wesley Chang';

/\* set up options \*/

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;

/\* read from file \*/

infile 'C:\Users\wesle\iCloudDrive\Summer 2020 (UCR)\STAT 147 (Session A)\Assignments\3\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**;