Statistics 147 Assignment #4

Summer 2020; 50 pts

DUE DATE: Wednesday, July 15, 2020, by 6:00 pm

1 GENERAL INSTRUCTIONS:

- Templates for the solution (in Word and LaTeX) to this assignment has been uploaded to Blackboard. They were designed to minimize your typing of the solution. You may use them if you like. They are located under **Course Materials → Handouts → Templates**. (Be sure if you do not use one of these templates that you include all the steps in the tests of hypothesis!)
- **No late papers will be accepted for credit!**
- Your write-up should be neat, well-organized and concise but complete. Your write-up should include a cover page which includes your name, the last 4 digits of your student ID. Your write-up must be typed. Use a word processor such as Word or LaTeX, etc. No hand-written papers will be accepted for credit!
- Be sure to include copies of any/all charts, graphs and output.
- You will need the following data file, that is available for download on **ilearn** under **Data Files**

```
agility.dat     dogdive.dat
```

- For your SAS session, name your program file *hwk4_su20.sas*.

  (i) **(2 pts)** Include the following titles:

```
title1 'Statistics 147 Assignment #4';
title2 'Summer 2020';
title3 'Your Name';
title4 'SAS Question XX'    where XX = question number
title5 'SubPart (xx)';     /* where xx is something like ii */
```

  (ii) **(4 pts)** Be sure to include documentation in your SAS program.

  (iii) **(2 pts)** Be sure to turn in your SAS program and output!

- For your R session,

  ♠ **(2 pts)** be sure include the following titles (as comments)

```
# Statistics 147 Assignment #4
# Summer 2020
# Your Name
# R Question ZZ (where ZZ = question number)
# Subpart XX (where XX = subpart number, something like ii)
```

  ♠ **(2 pts)** Be sure to include a copy of your R code. (You may copy and paste it from the R Console into Word, LaTeX, etc.)

  ♠ **(3 pts)** Be sure to include documentation in your R code.

2 The Questions

2.1 R

1. **(6 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:

```
File name: agility.dat
Cody Dusty Shadow
75      69      83
85      79      93
70      77      87
79      51      72
73      53      82
81      69      72
84      59      62
70      64      77
```

NOTE:

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

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

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

(iii) Using **R** to generate the appropriate output, test whether the true mean finishing time for **Cody** is more than 75 minutes. Use $\alpha = 0.05$. (Done as part of Assignment #3.)

(iv) 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.)

(v) **[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 $\alpha = 0.05$. (6 pts)

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

2. **(9 pts total)** *The Great Outdoor Games*, sponsored by *ESPN*, feature the *Big Air Dogs* event. In this event, also known as *dock diving*, the dogs race down a dock and fly off the end into the water, with long distance being the goal. During the competition, each dog is allowed 2 or 3 jumps. Linda and Lauren, believing that the dogs dive distance is **significantly different** on the second dive than on the first dive, attended a recent competition and recorded the dive distance for the first two dives for the top 11 dogs, and records them in a data file named **dogdive.dat**. (NOTE: The data represents the distance of each dive, measured in feet, for each dog.)

```
Filename: dogdive.dat
Dog     Dive1  Dive2
Cody     28.2   29.0
Dusty    25.9   26.4
Shadow   25.1   25.7
Korra    25.2   24.9
Kali     23.8   23.9
Dexter   22.5   22.8
Lakota   29.1   29.4
Abby     23.3   23.2
MaxG     25.2   26.1
MaxS     24.1   24.3
Mercedes 23.2   23.3
```

NOTE: The headings are included in the data file. The actual data begins on line 3.

(i) 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)

(ii) 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 $\alpha = 0.05$. (5 pts)

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

2.2 SAS

1. **(9 pts total)** Refer to **R** Question 1.

(i) 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.)

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

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

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

(iv) 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.)

(v) 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.)

(a) Test whether the true mean time of the **Cody** team is 75 minutes. Use $\alpha = 0.05$. (Done as part of Assignment #3.)

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

(vi) 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.)

(a) **[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 $\alpha = 0.05$. (6 pts)

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

(c) **[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)

2. **[NEW] (6 pts total)** Refer to R Question 2.

(i) Read in the data using an infile statement. (Do not use DO loops! Be sure to print your data as a check!) (4 pts)

(ii) Perform the test of hypothesis to determine whether the mean dive distance for **Dive1** is **less** than the mean dive distance for **Dive2**. Use $\alpha = 0.05$. (5 pts)

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