Statistics 147 Assignment #4

Summer 2020

Wesley Chang

0996

2 The Questions

2.1 R

Question 1:

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

```
> ### 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
  Cody Dusty Shadow
1   75    69     83
2   85    79     93
3   70    77     87
4   79    51     72
5   73    53     82
6   81    69     72
7   84    59     62
8   70    64     77
> |
```

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

```
> ## Part ii
> # add code to make the columns accessible individually and obtain their column headers
> attach(question1)
> names(question1)
[1] "Cody"   "Dusty"  "Shadow"
>
> # print column data to make sure it came out right
> Cody
[1] 75 85 70 79 73 81 84 70
> Dusty
[1] 69 79 77 51 53 69 59 64
> Shadow
[1] 83 93 87 72 82 72 62 77
> |
```

(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.)

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.

```
> # 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)

        One Sample t-test

data:  Cody
t = 1.004, df = 7, p-value = 0.1744
alternative hypothesis: true mean is greater than 75
95 percent confidence interval:
 73.11497      Inf
sample estimates:
mean of x
   77.125
```

(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.)

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

```
> ## 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)

        One Sample t-test

data:  Cody
t = 36.439, df = 7, p-value = 3.046e-09
alternative hypothesis: true mean is not equal to 0
96 percent confidence interval:
 71.79809 82.45191
sample estimates:
mean of x
   77.125
```

(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)

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.

```
> ## 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)

        F test to compare two variances

data:  Cody and Dusty
F = 0.335, num df = 7, denom df = 7, p-value = 0.1723
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.06706879 1.67330709
sample estimates:
ratio of variances
         0.3350025

>
> # 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)

        Two Sample t-test

data:  Cody and Dusty
t = 2.8401, df = 14, p-value = 0.006552
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 4.558037      Inf
sample estimates:
mean of x mean of y
   77.125    65.125
```

(vi) 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:

(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)

```
> ### Question 2
>
> ## Part i
>
> question2 = read.table(file = "dogdive.dat", header = TRUE, skip = 1)
> question2
       Dog Dive1 Dive2
1      Cody  28.2  29.0
2     Dusty  25.9  26.4
3    Shadow  25.1  25.7
4     Korra  25.2  24.9
5      Kali  23.8  23.9
6    Dexter  22.5  22.8
7    Lakota  29.1  29.4
8      Abby  23.3  23.2
9      MaxG  25.2  26.1
10     MaxS  24.1  24.3
11 Mercedes  23.2  23.3
>
> attach(question2)
> names(question2)
[1] "Dog"   "Dive1" "Dive2"
> Dog
 [1] "Cody"     "Dusty"    "Shadow"   "Korra"    "Kali"     "Dexter"   "Lakota"   "Abby"     "MaxG"
[10] "MaxS"     "Mercedes"
> Dive1
 [1] 28.2 25.9 25.1 25.2 23.8 22.5 29.1 23.3 25.2 24.1 23.2
> Dive2
 [1] 29.0 26.4 25.7 24.9 23.9 22.8 29.4 23.2 26.1 24.3 23.3
>
```

(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)

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.

```
> ## 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)

        F test to compare two variances

data:  Dive1 and Dive2
F = 0.84679, num df = 10, denom df = 10, p-value = 0.7977
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2278278 3.1473366
sample estimates:
ratio of variances
         0.8467885

>
> # 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)

        Two Sample t-test

data:  Dive1 and Dive2
t = -0.33653, df = 20, p-value = 0.74
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.224967  1.606785
sample estimates:
mean of x mean of y
 25.05455  25.36364
```

(iii) 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.

   (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.)

```
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;
```

Statistics 147 Assignment #4
Summer 2020
Wesley Chang
Question 1
Part i and ii

row	dog	name	score
1	1	Cody	75
1	2	Dust	69
1	3	Shad	83
2	1	Cody	85
2	2	Dust	79
2	3	Shad	93
3	1	Cody	70
3	2	Dust	77
3	3	Shad	87
4	1	Cody	79
4	2	Dust	51
4	3	Shad	72
5	1	Cody	73
5	2	Dust	53
5	3	Shad	82
6	1	Cody	81
6	2	Dust	69
6	3	Shad	72
7	1	Cody	84
7	2	Dust	59
7	3	Shad	62
8	1	Cody	70
8	2	Dust	64
8	3	Shad	77

(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**

See above

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

```
title5 'Part (iii)';
/* add code to sort the data by the name of the dog */
proc sort data = agilityl;
     by dog;
run;

proc print noobs data = agilityl;
     title6 'Print to check sorted';
run;
```

Statistics 147 Assignment #4
Summer 2020
Wesley Chang
Question 1
Part (iii)
Print to check sorted

row	dog	name	score
1	1	Cody	75
2	1	Cody	85
3	1	Cody	70
4	1	Cody	79
5	1	Cody	73
6	1	Cody	81
7	1	Cody	84
8	1	Cody	70
1	2	Dust	69
2	2	Dust	79
3	2	Dust	77
4	2	Dust	51
5	2	Dust	53
6	2	Dust	69
7	2	Dust	59
8	2	Dust	64
1	3	Shad	83
2	3	Shad	93
3	3	Shad	87
4	3	Shad	72
5	3	Shad	82
6	3	Shad	72
7	3	Shad	62
8	3	Shad	77

(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.)

```
title5 'Part iv';
/* add code to generate mean, std dev, and median */
proc means mean stddev median data = agilityl;
     title6 'Mean, Standard Deviation, and Median for agility.dat';
     by dog;
     var score;
run;
```

Statistics 147 Assignment #4
Summer 2020
Wesley Chang
Question 1
Part iv
Mean, Standard Deviation, and Median for agility.dat

The MEANS Procedure

dog=1

Analysis Variable : score		
Mean	Std Dev	Median
77.1250000	5.9865922	77.0000000

dog=2

Analysis Variable : score		
Mean	Std Dev	Median
65.1250000	10.3432172	66.5000000

dog=3

Analysis Variable : score		
Mean	Std Dev	Median
78.5000000	9.8125284	79.5000000

(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.)

```
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 agilityl */
    set agilityl;

    /* 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;
```

Statistics 147 Assignment #4
Summer 2020
Wesley Chang
Question 1
Part v
Only Codys scores

score
75
85
70
79
73
81
84
70

(a) Test whether the true mean time of the **Cody** team is 75 minutes. Use $\alpha = 0.05$. (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.

```
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;
```

Statistics 147 Assignment #4
Summer 2020
Wesley Chang
Question 1
Part v
Subpart a
Test whether the true mean time of the Cody team is 75 minutes

The TTEST Procedure

Variable: score

N	Mean	Std Dev	Std Err	Minimum	Maximum
8	77.1250	5.9866	2.1166	70.0000	85.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
77.1250	72.1201	82.1299	5.9866	3.9582	12.1843

DF	t Value	Pr > |t|
7	1.00	0.3488

(b) 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.

```
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;
```

Statistics 147 Assignment #4
Summer 2020
Wesley Chang
Question 1
Part v
Subpart b
98% confidence interval for the Cody team

The MEANS Procedure

Analysis Variable : score	
Lower 98% CL for Mean	Upper 98% CL for Mean
70.7795958	83.4704042

(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.)

```
title5 'Part vi';
title6;
title7;
data bothDS;
    set agility1;

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

proc print noobs data = bothDS;
run;
```

Statistics 147 Assignment #4
Summer 2020
Wesley Chang
Question 1
Part vi

row	dog	name	score
1	2	Dust	69
2	2	Dust	79
3	2	Dust	77
4	2	Dust	51
5	2	Dust	53
6	2	Dust	69
7	2	Dust	59
8	2	Dust	64
1	3	Shad	83
2	3	Shad	93
3	3	Shad	87
4	3	Shad	72
5	3	Shad	82
6	3	Shad	72
7	3	Shad	62
8	3	Shad	77

(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)

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.

```
title6 'Subpart a';
/* perform appropriate test of hypothesis to determine whether one can conclude that the mean
/* use a = 0.05 */

/* test for equality of variances */
proc ttest data = bothDS;
    title7 'Checking for equality of variances';
    class name;
    var score;
run;
```

Statistics 147 Assignment #4
Summer 2020
Wesley Chang
Question 1
Part vi
Subpart a
Checking for equality of variances

The TTEST Procedure

Variable: score

name	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Dust		8	65.1250	10.3432	3.6569	51.0000	79.0000
Shad		8	78.5000	9.8125	3.4693	62.0000	93.0000
Diff (1-2)	Pooled		-13.3750	10.0814	5.0407		
Diff (1-2)	Satterthwaite		-13.3750		5.0407		

name	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Dust		65.1250	56.4779	73.7721	10.3432	6.8387	21.0513
Shad		78.5000	70.2965	86.7035	9.8125	6.4878	19.9712
Diff (1-2)	Pooled	-13.3750	-24.1862	-2.5638	10.0814	7.3808	15.8993
Diff (1-2)	Satterthwaite	-13.3750	-24.1890	-2.5610			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	14	-2.65	0.0189
Satterthwaite	Unequal	13.961	-2.65	0.0189

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	7	7	1.11	0.8930

```
 /* 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;
```

Statistics 147 Assignment #4
Summer 2020
Wesley Chang
Question 1
Part vi
Subpart a
Testing Dusty vs Shadow

The TTEST Procedure

Variable: score

name	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Dust		8	65.1250	10.3432	3.6569	51.0000	79.0000
Shad		8	78.5000	9.8125	3.4693	62.0000	93.0000
Diff (1-2)	Pooled		-13.3750	10.0814	5.0407		
Diff (1-2)	Satterthwaite		-13.3750		5.0407		

name	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Dust		65.1250	56.4779	73.7721	10.3432	6.8387	21.0513
Shad		78.5000	70.2965	86.7035	9.8125	6.4878	19.9712
Diff (1-2)	Pooled	-13.3750	-22.2532	Infty	10.0814	7.3808	15.8993
Diff (1-2)	Satterthwaite	-13.3750	-22.2549	Infty			

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	14	-2.65	0.9905
Satterthwaite	Unequal	13.961	-2.65	0.9905

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	7	7	1.11	0.8930

(b) **[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.

(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)

```
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;
```

Statistics 147 Assignment #4
Summer 2020
Wesley Chang
Question 1
Part vi
Subpart c
98% CI

The TTEST Procedure

Variable: score

name	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Dust		8	65.1250	10.3432	3.6569	51.0000	79.0000
Shad		8	78.5000	9.8125	3.4693	62.0000	93.0000
Diff (1-2)	Pooled		-13.3750	10.0814	5.0407		
Diff (1-2)	Satterthwaite		-13.3750		5.0407		

name	Method	Mean	98% CL Mean		Std Dev	98% CL Std Dev	
Dust		65.1250	54.1619	76.0881	10.3432	6.3666	24.5845
Shad		78.5000	68.0993	88.9007	9.8125	6.0400	23.3231
Diff (1-2)	Pooled	-13.3750	-26.6042	-0.1458	10.0814	6.9876	17.4731
Diff (1-2)	Satterthwaite	-13.3750	-26.6089	-0.1411			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	14	-2.65	0.0189
Satterthwaite	Unequal	13.961	-2.65	0.0189

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	7	7	1.11	0.8930

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)

```
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;
```

Statistics 147 Assignment #4
Summer 2020
Wesley Chang
Question 2
Part i

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

(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)

```
title5 'Part ii';

/* test whether the mean dive distancec for Divel 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 Divel*Dive2;
run;
```

(iii) 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;
```