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:

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

```
## 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
```

```
> ## 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
  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.)

```
## Subpart (ii)
# add in code to make the columns accessible individually and obtain the headers
attach(agility)
names(agility)
Cody
Dusty
Shadow
```

```
> ## Subpart (ii)
> # add in code to make the columns accessible individually and obtain the headers
> attach(agility)
> names(agility)
[1] "Cody"   "Dusty"  "Shadow"
> 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) **[NEW]** Using **R** to generate the appropriate output, test whether the true mean finishing time for **Cody** is more than 75 minutes. Use $\alpha = 0.05$. (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.

```
## 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 (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
[1] 5.986592
>
> # generate hypothesis conclusion from t.test
> t.test(Cody,alternative="greater",mu=75,sd=sd_a,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) **[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.

```
# Subpart (iv)

# generate confidence interval
t.test(Cody,alternative="two.sided",conf.level=0.96)
```

```
> # Subpart (iv)
>
> # generate confidence interval
> 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
```

SAS

1. **(10 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;

    /* 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;
```

Statistics 147 Assignment #3
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.)

See answer for part (i)

(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 = agility1;
     by dog;
run;

proc print noobs data = agility1;
     title6 'Print to check sorted';
run;
```

Statistics 147 Assignment #3
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 #3
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 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;
```

Statistics 147 Assignment #3
Summer 2020
Wesley Chang
Question 1
Part (v)
Only Codys scores

score
75
85
70
79
73
81
84
70

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

```
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 #3
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) **[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.

```
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 #3
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

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;
```