Statistics 147: Practice Exam II Solution

1 Using R

1. (i) Read in and print out the data, making sure the columns are accessible individually.

FROM R:

```
> # Read in and print the data
> age_data <-
read.table("c:/linda/summer2019/su19147/datafiles/pr2_agegroup_f18.dat",
header = TRUE, skip = 1)
> # Print the data as a check
> age_data
   G1 G2 G3 G4
1  29 20 37 28
2  33 21 25 29
. . .
10 22 25 33 32
> # Use the attach() function to make the columns accessible individually
> attach(age_data)
> # Use the names() function to obtain the column names
> names(age_data)
[1] "G1" "G2" "G3" "G4"
```

(ii)

(a) Normality Tests

FROM R:

```
> # Question 1, Part (ii) , subpart (a) Test for Normality using Anderson-Daling
> # Use ad.test
> # Install nortest package
> local({pkg <- select.list(sort(.packages(all.available = TRUE)),graphics=TRUE)
+ if(nchar(pkg)) library(pkg, character.only=TRUE)})
> # For Age Group 1
> ad.test(G1)
        Anderson-Darling normality test
data:  G1
A = 0.17998, p-value = 0.8878 ****
> # For Age Group 2
> ad.test(G2)
        Anderson-Darling normality test
data:  G2
A = 0.53232, p-value = 0.1283 ****
> # For Age Group 3
> ad.test(G3)
        Anderson-Darling normality test
data:  G3
A = 0.20139, p-value = 0.8325 ****
```

```
> # For Age Group 4
> ad.test(G4)
        Anderson-Darling normality test
data:  G4
A = 0.22518, p-value = 0.754 ****
```

- H_0: Age Group 1 is normally distributed
- H_a: Age Group 1 is not normally distributed
- TS: p-value = 0.7540
- RR: Reject H_0 if p-value $< \alpha = 0.05$
- Conclusion: Since p-value $= 0.8878 \not< \alpha = 0.05$ $\Rightarrow$ ok to assume normality for Age Group 1.

- H_0: Age Group 2 is normally distributed
- H_a: Age Group 2 is not normallydistributed
- TS: p-value = 0.1283
- RR: Reject H_0 if p-value $< \alpha = 0.05$
- Conclusion: Since p-value $= 0.1283 \not< \alpha = 0.05$ $\Rightarrow$ ok to assume normality for Age Group 2.

- H_0: Age Group 3 is normally distributed
- H_a: Age Group 3 is not normally distributed
- TS: p-value = 0.8325
- RR: Reject H_0 if p-value $< \alpha = 0.05$
- Conclusion: Since p-value $= 0.98325 \not< \alpha = 0.05$ $\Rightarrow$ ok to assume normality for Age Group 3

- H_0: Age Group 4 is normally distributed
- H_a: Age Group 4 is not normally distributed
- TS: p-value = 0.7540
- RR: Reject H_0 if p-value $< \alpha = 0.05$
- Conclusion: Since p-value $= 0.7540 \not< \alpha = 0.05$ $\Rightarrow$ ok to assume normality for Age Group 4.

(b) Test for equality (homogeneity) of variances.

FROM R:

```
> # Question 1, Part (ii), subpart (b) Test for Equality of Variances
> # using bartlett.test
> # First stack the data using the stack( ) function
> stack_ages <- stack (age_data)
> # Use attach( ) to make the columns accessible individually
> attach(stack_ages)
> # Use names( 0 function to obtain the column names
> names(stack_ages)
[1] "values" "ind"
> # Now perform equality of variances test
> # value = data values ind = classes
> # bartlett.test(values,ind)
> bartlett.test(values,ind)
        Bartlett test of homogeneity of variances
data:  values and ind
Bartlett's K-squared = 2.2615, df = 3, p-value = 0.5199****
```

- $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$
- H_a : at least one of the σ_i^2 is different
- p-value = 0.5199 (Bartlett's test)
- RR: Reject H_0 if p-value $< \alpha = 0.05$

- Conclusion: Since p-value $= 0.5199 \not< \alpha = 0.05 \Rightarrow$ do not reject $H_0 \Rightarrow$ ok to assume equality (homogeneity) of variances.

(c) Test means:

FROM R:

```
> # Question 1, Part (ii), subpart (c) Use aov function to generate
> # ANOVA information
> # format: aov(response~factor, data = dataname)
> results2 = aov(values~ind,data=stack_ages)
> # Use summary( ) function to show the results
> summary(results2)
            Df Sum Sq Mean Sq F value  Pr(>F)
ind          3  318.6  106.20   4.643 0.00762 **
Residuals   36  823.4   22.87
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ (cannot conclude a significant effect due to age on the increase in heart rate, on the average)
- H_a: at least one μ_i is different (can conclude that at least one of the ages yields a significantly different increase in heart rate, on the average)
- Test Statistic: p-value $= 0.00762$
- R.R.: Reject H_0 if the p-value $<$ specified value of α (For our work, $\alpha = 0.05$.)
- Conclusion: Since p-value $= 0.00762 < \alpha = 0.05 \Rightarrow$ reject $H_0 \Rightarrow$ can conclude that the mean increase in heart rate is significantly different for at least one of the age groups

(d) Use Tukey's test:

FROM R:

```
> # Use TukeyHSD test for multiple comparisons
>  TukeyHSD(results2,conf.level=0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level
Fit: aov(formula = values ~ ind, data = stack_ages)
$ind
      diff        lwr       upr     p adj
G2-G1 -7.5 -13.260261 -1.739739 0.0064875
G3-G1 -1.4  -7.160261  4.360261 0.9132507
G4-G1 -2.7  -8.460261  3.060261 0.5922047
G3-G2  6.1   0.339739 11.860261 0.0344766
G4-G2  4.8  -0.960261 10.560261 0.1307469
G4-G3 -1.3  -7.060261  4.460261 0.9289331
```

Using p-value approach:

RECALL: If p-value $< \alpha = 0.045 \rightarrow$ can conclude a significant difference.

Pair Comparison	p-value	p-value $< \alpha = 0.05$ (Yes or No)	Can conclude a significant difference? (Yes or No)
Group2 - Group1	**0.0064875**	**YES**	**YES**
Group3 - Group1	**0.9132507**	**NO**	**NO**
Group4 - Group1	**0.5922047**	**NO**	**NO**
Group3 - Group2	**0.0344766**	**YES**	**YES**
Group4 - Group2	**0.1307469**	**NO**	**NO**
Group4 - Group3	**0.9289331**	**NO**	**NO**

OR

Using the confidence interval approach:

RECALL: If 0 is in the interval, cannot conlude a significant difference. If 0 is not in the interval, can conclude a significant difference.

Age 1 vs Age 2: $(-13.26, -1.74) \Rightarrow$ Since 0 is not in the interval, can conclude there is a significant difference in mean heart rate increase between Age Group 1 and Age Group 2

Age 1 vs Age 3: $(-7.16, 4.36,) \Rightarrow$ Since 0 is in the interval, cannot conclude there is a significant difference in mean heart rate increase between Age Group 1 and Age Group 3

Age 1 vs Age 4: $(-8.46, 3.06) \Rightarrow$ Since 0 is in the interval, cannot conclude there is a significant difference in mean heart rate increase between Age Group 1 and Age Group 4

Age 2 vs Age 3: $(0.34, 11.86) \Rightarrow$ Since 0 is not in the interval, can conclude there is a significant difference in mean heart rate increase between Age Group 2 and Age Group 3

Age 2 vs Age 4: $(-0.96, 10.56) \Rightarrow$ Since 0 is in the interval, cannot conclude there is a significant difference in mean heart rate increase between Age Group 2 and Age Group 4

Age 3 vs Age 4: $(-7.06, 4.46) \Rightarrow$ Since 0 is in the interval, cannot conclude there is a significant difference in mean heart rate increase between Age Group 3 and Age Group 4

OR IN CHART FORMAT

Comparison	Lower	Upper	0 in interval?	Can conclude sig. diff
1 vs 2	**−13.260261**	**−1.739729**	**NO**	**YES**
1 vs 3	**−7.160261**	**4.360261**	**YES**	**NO**
1 vs 4	**−8.460261**	**3.060261**	**YES**	**NO**
2 vs 3	**0.339739**	**11.860261**	**NO**	**YES**
2 vs 4	**−0.960261**	**10.560261**	**YES**	**NO**
3 vs 4	**−7.060261**	**4.460261**	**YES**	**NO**

(iii) Make sure you have the Teaching Demos packages installed and loaded.

(a) Test for equality of 2 variances.

FROM R:

```
> # Question 1, Part (iii), subpart (i) Test for equality of variances between G3 & G4
> # Use var.test function with alternative = "two.sided" and  ratio = 1
> var.test(G3,G4,alternative = "two.sided", ratio = 1,conf.level = 0.05)

        F test to compare two variances
```

```
data:  G3 and G4
F = 0.70994, num df = 9, denom df = 9, p-value = 0.618 ****
alternative hypothesis: true ratio of variances is not equal to 1
5 percent confidence interval:
 0.6800750 0.7411222
sample estimates:
ratio of variances
         0.7099428
```

- $H_0 : \sigma_3^2 = \sigma_4^2$
- $H_a : \sigma_3^2 \neq \sigma_4^2$
- p-value = 0.618
- RR: Reject H_0 if p-value $< \alpha = 0.05$
- Conclusion: Since p-value $= 0.618 \not< \alpha = 0.05 \Rightarrow$ do not reject $H_0 \Rightarrow$ ok to assume equality (homogeneity) of variances.

(b) Now test means, assuming equality of variances

FROM R:

```
> # Use t.test function to test for equality of means
> # Use alternative  = "two.sided" and var.equal = TRUE
> t.test(G3,G4,alternative = "two.sided",var.equal= TRUE, conf.level = 0.95)

        Two Sample t-test

data:  G3 and G4
t = 0.56403, df = 18, p-value = 0.5797 ****
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.542259  6.142259
sample estimates:
mean of x mean of y
     29.5      28.2
```

- $H_0 : \mu_3 = \mu_4$ (cannot conclude a significant difference in mean increase in heart rate between age group 3 (40-59) and 4 (60-69))
- H_a: $\mu_3 \neq \mu_4$ (can conclude a significant difference in mean increase in heart rate between age group 3 (40-59) and 4 (60-69))
- p-value = 0.5797
- R.R.: Reject H_0 if the p-value < specified value of α (For our work, $\alpha = 0.05$.)
- Conclusion: Since p-value $= 0.5797 \not< \alpha = 0.05 \Rightarrow$ do not reject $H_0 \Rightarrow$ cannot conclude that the mean increase in heart rate is significantly different between Age Group 3 and Age Group 4.

(iii) Test $\mu_3 = 25$ vs $H_a : \mu_3 > 25$

FROM R:

```
> # Question 1, Part (iv), subpart (ii)
> # Use t.test with mu = 25, alternative = "greater"
> t.test(G3,mu=25,alternative = "greater", conf.level = 0.05)
```

```
        One Sample t-test
data:  G3
t = 3.0301, df = 9, p-value = 0.007122 *****
alternative hypothesis: true mean is greater than 25
5 percent confidence interval:
 32.22238      Inf
sample estimates:
mean of x
     29.5
```

- $H_0 : \mu_3 = 25$
- $H_a : \mu_3 > 25$
- p-value = 0.007122
- R.R.: Reject H_0 if the p-value < specified value of α (For our work, $\alpha = 0.05$.)
- Conclusion: Since p-value = $0.007122 < \alpha = 0.05 \Rightarrow$ reject $H_0 \Rightarrow$ can conclude that the mean increase in heart rate is greater than 25 for Age Group 3.

(iv) 98% CI for Group 4

FROM R:

```
> # Question 1, Part (v) 98% confidence interval for G4
> # Use t.test with alternative = "Two.sided", conf.level = 0.98
> t.test(G4, alternative = "two.sided", conf.level = 0.98)

        One Sample t-test

data:  G4
t = 15.999, df = 9, p-value = 6.435e-08
alternative hypothesis: true mean is not equal to 0
98 percent confidence interval:
 23.22701 33.17299
sample estimates:
mean of x
     28.2
```

LIMITS: (23.22701, 33.17299)

INTERPRETATION: One can be 98% confident that the true mean increase in heart rate for Age Group 4 (60 - 69) is between 23.22701 and 33.17299.

2. Let $\mu_D = \mu_{Male} - \mu_{Female}$

(i) Read in and print out the data

FROM R:

```
> # Question 2
> # Part (i) Read in and print the data
> litter_data <-
read.table("c:/linda/summer2019/su19147/datafiles/litter2_f18.dat", header = TRUE,skip =1)
> litter_data
   Litter Male Female
```

```
1       1  475    425
2       2  775    750
. . .
10     10  750    725
> # Use attach( ) function to make the columns accessible individually
> attach(litter_data)
> # Use names( ) function to obtain the column names
> names(litter_data)
[1] "Litter" "Male"   "Female"
```

(ii)

FROM R:

```
> # Question 2, Part (ii) Test means using paired-difference t-test
> # Use t.test with paired = TRUE and alternative = "greater"
> t.test(Male,Female,paired = TRUE, alternative = "greater", conf.level = 0.95)
        Paired t-test
data:  Male and Female
t = 3.5654, df = 9, p-value = 0.003034 ****
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 17.24797      Inf
sample estimates:
mean of the differences
                   35.5
```

- H_0 : $\underline{\quad \mu_D = 0 \quad}$ (cannot conclude the average selling price of male puppies is significantly higher than the average selling price of the female puppies)
- H_a : $\underline{\quad \mu_D \neq 0 \quad}$ (can conclude the average selling price of male puppies is significantly higher than the average selling price of the female puppies)
- p-value = **0.003034**
- Rejection Region: Reject H_0 if p-value $< \alpha$.
- Conclusion: Since p-value = 0.003034$< \alpha = 0.05 \Rightarrow$ reject $H_0 \Rightarrow$ can conclude the average selling price of male puppies is significantly higher than the average selling price of the female puppies.

2 Complete using SAS

3. Now using SAS

(i) ANOVA **(a) Normality tests**

```
STATISTICS 147  PRACTICE EXAM II
Summer 2018
LINDA M. PENAS
QUESTION 3
Part (i)
***********************************************************
level1=Sixty to 69
                  Tests for Normality
Test                 --Statistic---    -----p Value------
Shapiro-Wilk         W      0.94848    Pr < W      0.6506**
```

```
Kolmogorov-Smirnov    D     0.152307    Pr > D     >0.1500
Cramer-von Mises      W-Sq  0.032349    Pr > W-Sq  >0.2500
Anderson-Darling      A-Sq  0.225183    Pr > A-Sq  >0.2500
***********************************************************
level1=Ten to 19
                   Tests for Normality
Test                  --Statistic---    -----p Value------
Shapiro-Wilk          W     0.977531    Pr < W      0.9506**
Kolmogorov-Smirnov    D     0.156978    Pr > D     >0.1500
Cramer-von Mises      W-Sq  0.031083    Pr > W-Sq  >0.2500
Anderson-Darling      A-Sq  0.179985    Pr > A-Sq  >0.2500
***********************************************************
level1=Thirty to 59
                   Tests for Normality
Test                  --Statistic---    -----p Value------
Shapiro-Wilk          W     0.971085    Pr < W      0.9007**
Kolmogorov-Smirnov    D     0.171943    Pr > D     >0.1500
Cramer-von Mises      W-Sq  0.033449    Pr > W-Sq  >0.2500
Anderson-Darling      A-Sq  0.201393    Pr > A-Sq  >0.2500
***********************************************************
level1=Twenty to 39
                   Tests for Normality
Test                  --Statistic---    -----p Value------
Shapiro-Wilk          W     0.876139    Pr < W      0.1178**
Kolmogorov-Smirnov    D     0.247198    Pr > D      0.0824
Cramer-von Mises      W-Sq   0.08995    Pr > W-Sq   0.1367
Anderson-Darling      A-Sq  0.532319    Pr > A-Sq   0.1334
***********************************************************
```

- H_0: Age Group 1 is normally distributed
- H_a: Age Group 1 is not normally distributed
- TS: p-value = 0.9506
- RR: Reject H_0 if p-value $< \alpha = 0.05$
- Conclusion: Since p-value $= 0.9506 \not< \alpha = 0.05$ $\Rightarrow$ ok to assume normality for Age Group 1.

- H_0: Age Group 2 is normally distributed
- H_a: Age Group 2 is not normally distributed
- TS: p-value = 0.1178
- RR: Reject H_0 if p-value $< \alpha = 0.05$
- Conclusion: Since p-value $= 0.1178 \not< \alpha = 0.05$ $\Rightarrow$ ok to assume normality for Age Group 2.

- H_0: Age Group 3 is normally distributed
- H_a: Age Group 3 is not normally distributed
- TS: p-value = 0.9007
- RR: Reject H_0 if p-value $< \alpha = 0.05$
- Conclusion: Since p-value $= 0.9007 \not< \alpha = 0.05$ $\Rightarrow$ ok to assume normality for Age Group 3

- H_0: Age Group 4 is normally distributed
- H_a: Age Group 4 is not normally distributed
- TS: p-value = 0.6506
- RR: Reject H_0 if p-value $< \alpha = 0.05$
- Conclusion: Since p-value $= 0.6506 \not< \alpha = 0.05$ $\Rightarrow$ ok to assume normality for Age Group 4.

(b) Homogeneity (equality) of variances

```
QUESTION 3
Part (i)
The GLM Procedure
Bartlett's Test for Homogeneity of hartrate Variance
Source        DF    Chi-Square    Pr > ChiSq
level1         3        2.2615        0.5199
```

- $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$
- H_a : at least one of the σ_i^2 is different

- p-value $= 0.5199$ (Bartlett's test)
- RR: Reject H_0 if p-value $< \alpha = 0.05$
- Conclusion: Since p-value $= 0.5199 \not< \alpha = 0.05 \Rightarrow$ do not reject $H_0 \Rightarrow$ ok to assume equality (homogeneity) of variances.

(c) Test for means

```
QUESTION 3
Part (i)
The GLM Procedure
                        Class Level Information
Class         Levels    Values
level1            4    Sixty to 69 Ten to 19 Thirty to 59 Twenty to 39
Number of observations    40
Dependent Variable: hartrate
                                     Sum of
Source                     DF        Squares    Mean Square   F Value   Pr > F
Model                       3     318.600000     106.200000      4.64   0.0076 ****
Error                      36     823.400000      22.872222
Corrected Total            39    1142.000000
R-Square     Coeff Var      Root MSE    hartrate Mean
0.278984      17.08033      4.782491         28.00000
Source                     DF      Type I SS    Mean Square   F Value   Pr > F
level1                      3    318.6000000    106.2000000      4.64   0.0076
Source                     DF    Type III SS    Mean Square   F Value   Pr > F
level1                      3    318.6000000    106.2000000      4.64   0.0076
```

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ (cannot conclude a significant effect due to age on the increase in heart rate, on the average)
- H_a: at least one μ_i is different (can conclude that at least one of the ages yields a significantly different increase in heart rate, on the average)
- Test Statistic: p-value $= 0.0076$
- R.R.: Reject H_0 if the p-value $<$ specified value of α (For our work, $\alpha = 0.05$.)
- Conclusion: Since p-value $= 0.0076 < \alpha = 0.05 \Rightarrow$ reject $H_0 \Rightarrow$ conclude that the mean increase in heart rate is significantly different for at least one of the age groups (on the average)

(d) Tukey's test

Using the Grouping Method:

```
QUESTION 3
Part (i)
The GLM Procedure
Tukey's Studentized Range (HSD) Test for hartrate
NOTE: This test controls the Type I experimentwise error rate, but it
generally has a higher Type II error rate than REGWQ.
Alpha                                  0.05
Error Degrees of Freedom                 36
Error Mean Square                  22.87222
Critical Value of Studentized Range  3.80880
Minimum Significant Difference       5.7603
Means with the same letter are not significantly different.
Tukey
Grouping
                Mean      N    level1
     A        30.900     10    Ten to 19  (1)
     A
     A        29.500     10    Thirty to 59 (3)
```

```
     A
B    A        28.200     10    Sixty to 69 (4)
B
B             23.400     10    Twenty to 39 (2)
```

Comparison			
Age Group	**Age Group**	**Same Letter**	**Can conclude a significant difference**
10 to 19 (1)	**20 to 39 (2)**	**No**	**Yes**
10 to 19 (1)	**30 to 59 (3)**	**Yes**	**No**
10 to 19 (1)	**60 to 69 (4)**	**Yes**	**No**
20 to 39 (2)	**30 to 59 (3)**	**No**	**Yes**
20 to 39 (2)	**60 to 69 (4)**	**Yes**	**No**
30 to 59 (3)	**60 to 69 (4)**	**Yes**	**No**

Using the Confidence Interval Method:

```
The GLM Procedure
Tukey's Studentized Range (HSD) Test for hartrate
NOTE: This test controls the Type I experimentwise error rate.
Alpha                                      0.05
Error Degrees of Freedom                     36
Error Mean Square                      22.87222
Critical Value of Studentized Range  3.80880
Minimum Significant Difference        5.7603
 Comparisons significant at the 0.05 level are indicated by ***.
                               Difference      Simultaneous
          level1                  Between     95% Confidence
        Comparison                  Means          Limits
Ten to 19    - Thirty to 59         1.400     -4.360   7.160
Ten to 19    - Sixty to 69          2.700     -3.060   8.460
Ten to 19    - Twenty to 39         7.500      1.740  13.260  ***
Thirty to 59 - Ten to 19           -1.400     -7.160   4.360
Thirty to 59 - Sixty to 69          1.300     -4.460   7.060
Thirty to 59 - Twenty to 39         6.100      0.340  11.860  ***
Sixty to 69  - Ten to 19           -2.700     -8.460   3.060
Sixty to 69  - Thirty to 59        -1.300     -7.060   4.460
Sixty to 69  - Twenty to 39         4.800     -0.960  10.560
Twenty to 39 - Ten to 19           -7.500    -13.260  -1.740  ***
Twenty to 39 - Thirty to 59        -6.100    -11.860  -0.340  ***
Twenty to 39 - Sixty to 69         -4.800    -10.560   0.960
```

Comparison	**Lower**	**Upper**	**0 in interval?**	**Can conclude sig. diff**
1 vs 2	**1.74**	**13.26**	**NO**	**YES**
1 vs 3	**−4.36**	**7.16**	**YES**	**NO**
1 vs 4	**−3.06**	**8.46**	**YES**	**NO**
2 vs 3	**−11.86**	**−0.34**	**NO**	**YES**
2 vs 4	**−10.56**	**0.96**	**YES**	**NO**
3 vs 4	**−4.46**	**7.06**	**YES**	**NO**

(ii)

```
QUESTION 3
Part (ii)
The TTEST Procedure
Variable:  hartrate
agegrp          N        Mean     Std Dev     Std Err     Minimum     Maximum

1              10     30.9000      5.1951      1.6428     22.0000     39.0000
2              10     23.4000      3.3731      1.0667     20.0000     30.0000
Diff (1-2)             7.5000      4.3799      1.9587
agegrp        Method              Mean       95% CL Mean        Std Dev
1                              30.9000     27.1837  34.6163      5.1951
2                              23.4000     20.9870  25.8130      3.3731
Diff (1-2)    Pooled            7.5000      4.1034  Infty        4.3799
Diff (1-2)    Satterthwaite     7.5000      4.0727  Infty
agegrp        Method            95% CL Std Dev
1                               3.5734   9.4842
2                               2.3201   6.1580
Diff (1-2)    Pooled            3.3095   6.4771
Diff (1-2)    Satterthwaite
Method           Variances        DF    t Value    Pr > t
Pooled           Equal            18       3.83    0.0006 ****
Satterthwaite    Unequal      15.443       3.83    0.0008
              Equality of Variances
Method      Num DF    Den DF    F Value    Pr > F
Folded F         9         9       2.37    0.2143 ****
```

(a) Test variances

- $H_0 : \sigma_1^2 = \sigma_2^2$
- $H_a : \sigma_1^2 \neq \sigma_2^2$
- p-value = 0.2143 (F-test)
- RR: Reject H_0 if p-value $< \alpha = 0.05$
- Conclusion: Since p-value $= 0.2143 \not< \alpha = 0.05 \Rightarrow$ do not reject $H_0 \Rightarrow$ ok to assume equality (homogeneity) of variances.

(b) Now test means assuming equal variances

- $H_0 : \mu_1 = \mu_2$ (cannot conclude that the mean increase in heart rate age group 1 (10-19) is significantly larger than the mean increase in heart rate for group 2 (20-39).)
- H_a: $\mu_1 > \mu_2$ (can conclude that the mean increase in heart rate age group 1 (10-19) is significantly larger than the mean increase in heart rate for group 2 (20-39).)
- Test Statistic: p-value = 0.0006
- R.R.: Reject H_0 if the p-value < specified value of α (For our work, $\alpha = 0.05$.)
- Conclusion: Since p-value $= 0.0006 < \alpha = 0.05 \Rightarrow$ reject $H_0 \Rightarrow$ conclude that the mean increase in heart rate for Age Group 1 is significantly larger than the mean increase in heart rate for Age Group 2.

(iii) Test $\mu_3 > 25$

```
QUESTION 3
Part (iii)
The TTEST Procedure
Variable:  hartrate

  N        Mean     Std Dev     Std Err     Minimum     Maximum
 10     29.5000      4.6963      1.4851     22.0000     37.0000
```

```
    Mean       95% CL Mean        Std Dev      95% CL Std Dev
 29.5000     26.7776  Infty        4.6963      3.2303   8.5737

    DF    t Value    Pr > t
     9       3.03    0.0071
*******************************************************************
The UNIVARIATE Procedure
Variable:  hartrate
           Tests for Location: Mu0=25
Test           -Statistic-    -----p Value------
Student's t    t  3.030076    Pr > |t|    0.0142 **** divide by 2
Sign           M       3.5    Pr >= |M|   0.0391
Signed Rank    S        19    Pr >= |S|   0.0234
```

- $H_0 : \mu_3 = 25$ (cannot conclude that the mean increase in heart rate is significantly greater than 25 for Age Group 3)
- $H_a : \mu_3 > 25$ (can conclude that the mean increase in heart rate is significantly greater than 25 for Age Group 3)
- Test Statistic: p-value = 0.0071
- R.R.: Reject H_0 if the p-value < specified value of α (For our work, $\alpha = 0.05$.)
- Conclusion: Since p-value = $0.010071 < \alpha = 0.05 \Rightarrow$ reject $H_0 \Rightarrow$ can conclude that the mean increase in heart rate is significantly greater than 25 for Age Group 3.

(iv)

```
QUESTION 3
Part (iv)
The MEANS Procedure
Analysis Variable : hartrate
  Lower 98%       Upper 98%
 CL for Mean     CL for Mean
------------------------------
  23.2270072      33.1729928
------------------------------
```

LIMITS: (23.2270072, 33.1729928)

INTERPRETATION: One can be 98% confident that the true mean increase in heart rate for Age Group 4 (60 - 69) is between 23.2270072 and 33.1729928.

4. Let $\mu_D = \mu_{Male} - \mu_{Female}$

(i) See attached SAS Program.

(ii)

- H_0 : $\underline{\quad \mu_D = 0 \quad}$ (cannot conclude there is a significant difference, on the average, in the selling price of male and female puppies)
- H_a : $\underline{\quad \mu_D \neq 0 \quad}$ (can conclude there is a significant difference, on the average, in the selling price of male and female puppies)
- p-value = **0.0061**
- Rejection Region: Reject H_0 if p-value < α.
- Conclusion: Since p-value = $0.0061 < \alpha = 0.05 \Rightarrow$ reject $H_0 \Rightarrow$ can conclude there is a significant difference, on the average, in the selling price of male and female puppies.

From SAS:

```
The TTEST Procedure
Difference:  male - female
  N        Mean     Std Dev     Std Err     Minimum     Maximum
 10     35.5000     31.4863      9.9569    -25.0000       100.0

    Mean       95% CL Mean        Std Dev      95% CL Std Dev
 35.5000     12.9760  58.0240     31.4863     21.6574  57.4818

    DF    t Value    Pr > |t|
     9       3.57      0.0061 ****
```

STATISTICS 147 PRACTICE EXAM II SAS PROGRAM

```
options ls=78 nocenter nodate nonumber ps=50;
/* Use DM to clear all windows except the editor window */
DM log "odsresults; clear; out; clear; log; clear;";
ods graphics off;
/* Create temporary SAS dataset */
/* Use infile statement to open data file
       firstobs   line number on which data begins */
/* Read in data from external data file named
   'pr2_agegroup_f18.dat' */
/* QUESTION 2, PART (i) */
data alldata;
     infile 'c:\linda\summer2019\su19147\datafiles\pr2_agegroup_f18.dat' firstobs = 3;
   title1 'STATISTICS 147  PRACTICE EXAM II';
   title2 'FALL 2018';
   title3 'LINDA M. PENAS ';
   title4 'QUESTION 3';
   title5 'Part (i) ';
/* Create do loops for rows and columns */
    do rows = 1 to 10;
       do agegrp = 1 to 4;
/* Name the age groups */
            if agegrp = 1 then level1 = 'Ten to 19   ';
       else if agegrp = 2 then level1 = 'Twenty to 39';
       else if agegrp = 3 then level1 = 'Thirty to 59';
       else                    level1 = 'Sixty to 69 ';
/* Input and output data */
          input hartrate @@;
          output;
/* Close the do loops */
       end;
    end;
/* Print the data as a check */
proc print;
/* Sort the data by level in order to perform the
   check for normality      */
proc sort;
   by level1;
```

```
/* Check for normality of each area using proc univariate
   with normal option
        by    to group according to level
       ods select   to select what output to print
        var   variable list     */
proc univariate normal;
   by level1;
   ods select TestsForNormality;
   var hartrate;
/* Perform test of homogeneity of variances, test for
equality of means and multiple comparisons using proc glm
   class      classification variable
   model      dependent(response) = class variable(s)
   HOVTEST    homogeneity of variances
   tukey      tukey's multiple comparison test
   lsd        Fisher's lsd multiple comparison test   */

proc glm;
   class level1;
   model hartrate = level1;
   means level1 / HOVTEST = bartlett;
   means level1 / tukey;
   means level1/cldiff tukey;
/* Create temporary SAS dataset with Age Groups  1 and 2 only */
/* Use set command to bring in existing SAS data set */
data only12;
   set alldata;
   title5 'Part (ii)';
   /* Use if structure to restrict data */
        if agegrp = 1 or agegrp = 2;
/* Use proc ttest to generate appropriate information
      class   classification variable
      var     response variable */
proc ttest sides = U;
     class agegrp;
     var hartrate;
/* Create temporary SAS data set with Age Group 3 only
   Use set command to bring in existing SAS data set */
data only3;
   set alldata;
   title5 'Part (iii)';
/* Use  if structure to restrict data to Age Group 3 */
      if agegrp = 3;
/* Can  use proc univariate with mu0 = 25 */
proc univariate mu0 = 25;
     ods select TestsForLocation;
     var hartrate;
/* Use proc ttest to generate test information
   use sides = U to test mu > 25
   h0 = value_to_be_tested */
proc ttest h0 = 25 sides = U alpha = 0.05;
```

```
     var hartrate;
/* Create temporary SAS data set with Age Group 4 only
   Use set command to bring in existing SAS data set */
data only4;
   set alldata;
   title5 'Part (iii)';
/* Use  if structure to restrict data to Age Group 4 */
      if agegrp = 4;
/* Use proc means to generate confidence interval  information
    clm     confidence interval */
proc means clm alpha = 0.02;
       title5 'Part (iv)';
       var hartrate;
/* QUESTION 4 */
/* Create temporary SAS data set */
data dog1;
   title5 'Question 4';
/* Use infile statement to open the data file.
   Use firstobs command to indicate the data begins on line 3.
   Use input statement to enter variables list. */
   infile 'c:\linda\summer2019\su19147\datafiles\litter2_f18.dat' firstobs=3;
   input litter male female;
/* Print the data as a check */
   proc print;
/* Use proc ttest with paired option
    proc ttest;
         paired variable1* variable2   */
 proc ttest;
    paired male*female;
/* END QUESTION 4 */
run;
quit;
```

Statistics 147 Practice Exam II R SCRIPT

```
# ###############################################
# Statistics 147 Practice Exam II R Script
# Question 1, Part (i) Read in and print the data
age_data <-
read.table("c:/linda/summer2019/su19147/datafiles/pr2_agegroup_f18.dat", header = TRUE, skip = 1)
# Print the data as a check
# Use the attach() function to make the column accessible individually
attach(age_data)
# Use the names() function to obtain the column names
names(age_data)
# ###############################################
# Question 1, Part (ii) , subpart (a) Test for Normality using Anderson-Daling
# Use ad.test
# Install nortest package
# For Age Group 1
ad.test(G1)
# For Age Group 2
ad.test(G2)
# For Age Group 3
ad.test(G3)
# For Age Group 4
ad.test(G4)
# ###############################################
# Question 1, Part (ii), subpart (b) Test for Equality of Variances
# using bartlett.test
# First stack the data using the stack( ) function
stack_ages <- stack (age_data)
# Use attach( ) to make the olumn accessible individually
attach(stack_ages)
# Use names( 0 function to obtain the column names
names(stack_ages)
# Now perform equality of variances test
# value = data values ind = classes
bartlett.test(values,ind)
# ###############################################
# Question 1, Part (ii), subpart (c) Use aov function to generate
# ANOVA information
# format: aov(response~factor, data = dataname)
results2 = aov(values~ind,data=stack_ages)
# Use summary( ) function to show the results
summary(results2)
# Use TukeyHSD test for multiple comparisons
TukeyHSD(results2,conf.level=0.95)
# ###############################################
# Question 1, Part (iii), subpart (i) Test for equality of variances between G3 & G4
# Use var.test function with alternative = "two.sided" and  ratio = 1
var.test(G3,G4,alternative = "two.sided", ratio = 1,conf.level = 0.05)
# Use t.test function to test for equality of means
# Use alternative  = "two.sided" and var.equal = TRYE
```

```
t.test(G3,G4,alternative = "two.sided",var.equal= TRUE, conf.level = 0.95)
# ###############################################
#  Question 1, Part (ivi) Test5 mu(G3) > 25
# Use t.test with mu = 25, alternative = "greater"
t.test(G3,mu=25,alternative = "greater", conf.level = 0.05)
# ###############################################
# Question 1, Part (v) 98% confidence interval for G4
# Use t.test with alternative = "Two.sided", conf.level = 0.98
t.test(G4, alternative = "two.sided", conf.level = 0.98)
# ###############################################
# ###############################################
# Question 2
# Part (i) Read in and print the data
litter_data <-
read.table("c:/linda/summer2019/su19147/datafiles/litter2_f18.dat", header = TRUE,skip =1)
# Use attach( ) function to make the columns accessible individually
attach(litter_data)
# Use names( ) function to obtain the column names
names(litter_data)
# ###############################################
# Question 2, Part (ii) Test means using paired-difference t-test
# Use t.test with paired = TRUE and alternative = "greater"
t.test(Male,Female,paired = TRUE, alternative = "greater", conf.level = 0.95)
# ###############################################
```