**Statistics 147 Assignment #5**

**Summer 2020**

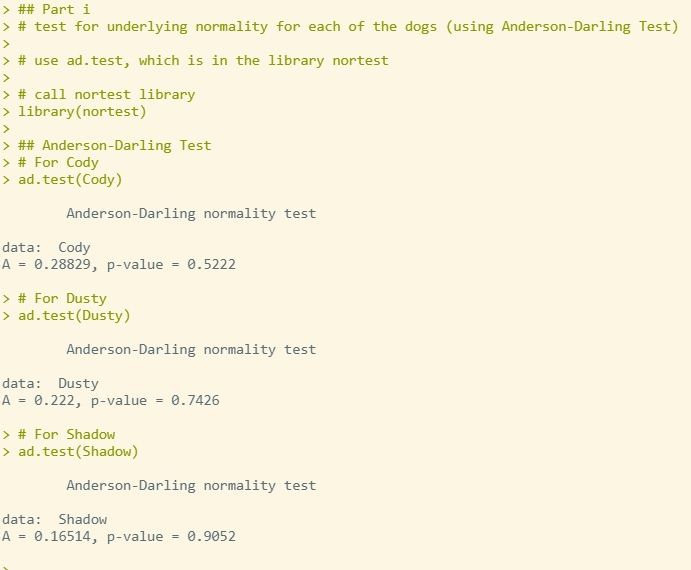
**Wesley Chang**

**0996**

**R Questions:**

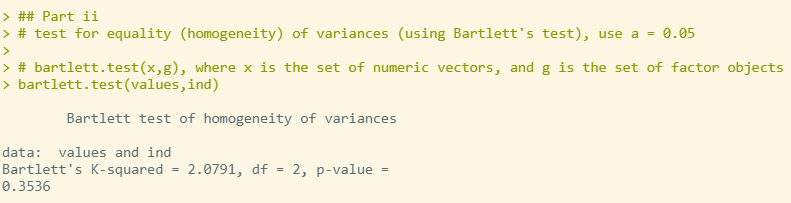
1. **Using R, complete the following:**
   1. Test for underlying normality for each of the dogs. (Use Anderson-Darling Test.) (4 pts)

**From the results of ad.test(), we can see that the results of the tests for Cody, Dusty, and Shadow return p-values that are larger than 0.05, which is the alpha at the 95% level. In all three cases, we do not reject the null hypothesis that the data is normally distributed. Therefore, we can assume that all three samples are normal.**



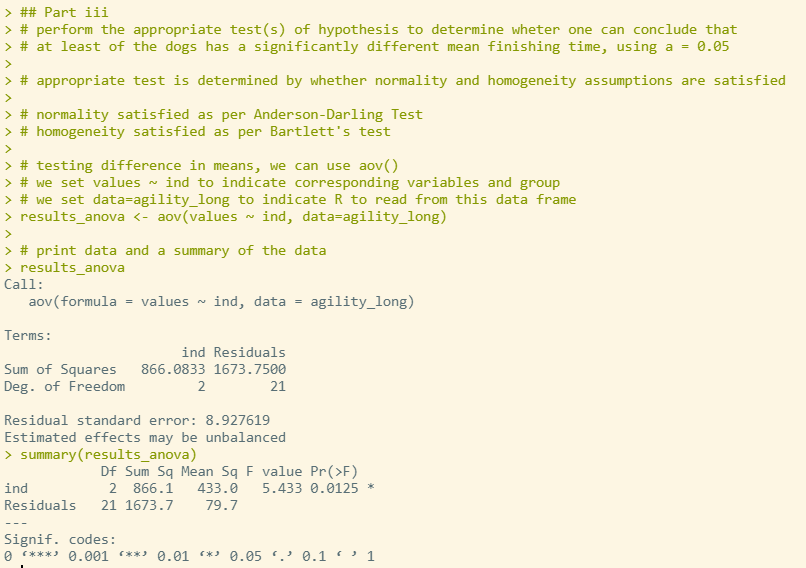
* 1. Test for equality (homogeneity) of variances. (Use Bartlett’s test.) Use *α* = 0*.*05. (2 pts)

**The results of Bartlett’s test give us a p-value of 0.3536. If we use alpha = 0.05 level of signficance, we see that the p-value is greater than the alpha, so we do not reject the null hypothesis that the variances are equal. Therefore, we can assume homogeneity of variances.**



* 1. Perform the appropriate test(s) of hypothesis to determine whether one can conclude that at least one of the dogs has a significantly different mean finishing time. Use *α* = 0*.*05. (4 pts)

**Since the above results show that there is normality and a homogeneity of variances, we can test the difference of means. The function aov() gives us a p-value of 0.0125, so we reject the null hypothesis that there is an equality in means at the 0.05 level of significance.**



* 1. If appropriate, use **Tukey’s test** *and* the **p-value method** to determine which mean(s) is(are) significantly different. (Be sure to justify your answer!) (3 pts)

**We can use Tukey’s test here since the data is independent, there is assumed normality, and assumed homogeneity in variances.**

**Confidence Interval approach:**

**Dusty-Cody: 0 does not fall into the confidence interval, so we can conclude that there is a difference in means between Dusty and Cody.**

**Shadow-Cody: 0 falls into the confidence interval, so we cannot conclude that is a difference in means between Shadow and Cody.**

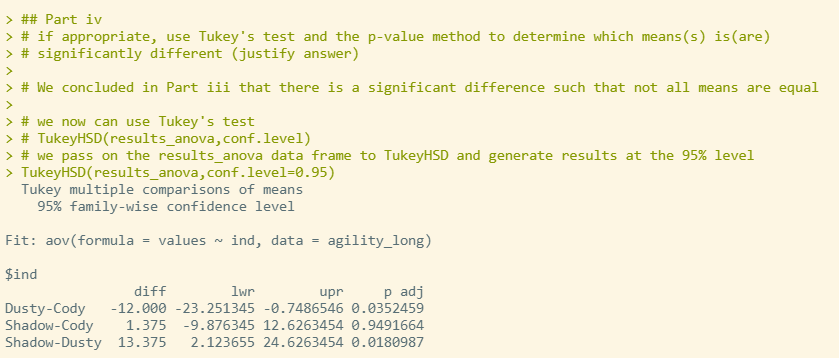
**Shadow-Dusty: 0 does not fall into the confidence interval, so we can conclude that there is a difference in means between Shadow and Dusty**

**P-value method:**

**Dusty-Cody: The p-value is lower than the 0.05 significance level, so we reject the null hypothesis that there is no difference, so we can conclude that there is a difference in means between Dusty and Cody.**

**Shadow-Cody: The p-value is higher than the 0.05 significance level, so we do not reject the null hypothesis that there is no difference, and cannot conclude that there is a difference in means between Shadow and Cody**

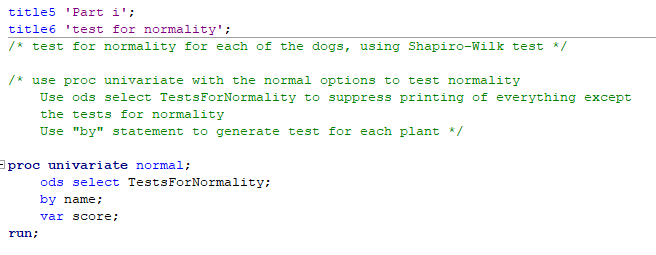
**Shadow-Dusty: The p-value is lower than the 0.05 significance level, so we reject the null hypothesis that there is no difference, so we can conclude that there is a difference in means between Shadow and Dusty.**

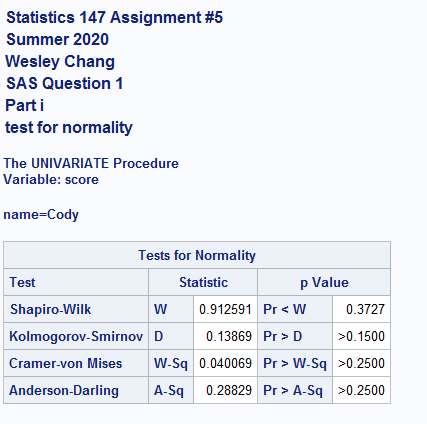


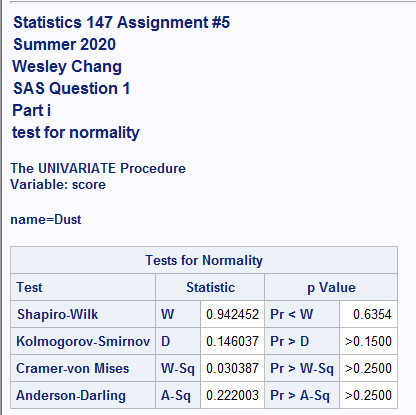
**SAS Questions:**

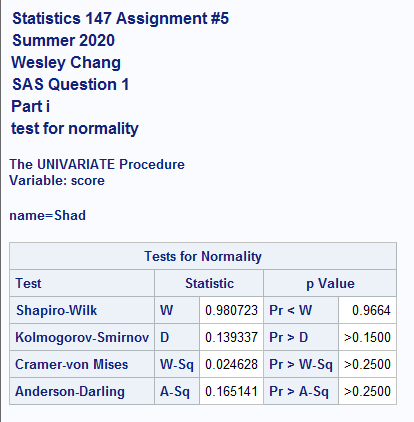
1. **Using SAS, complete the following.** Modify your existing **SAS** program file (from Assignments #3 and #4) to complete the following.
   1. Test for underlying normality for each of the dogs. (Use Shapiro-Wilk Test.) (4 pts)

**According to the results of the Shapiro-Wilk test, each dog’s p-value is larger than the level of significance 0.05. This means that we do not reject the null hypothesis that each dog’s results are normal, and can assume normality in all three samples.**



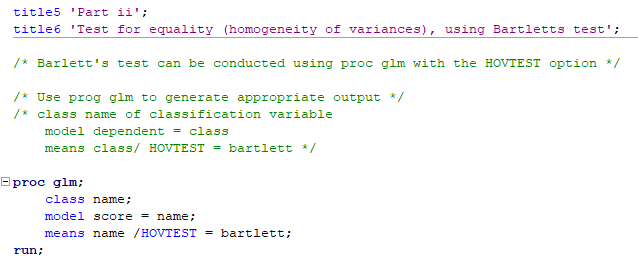


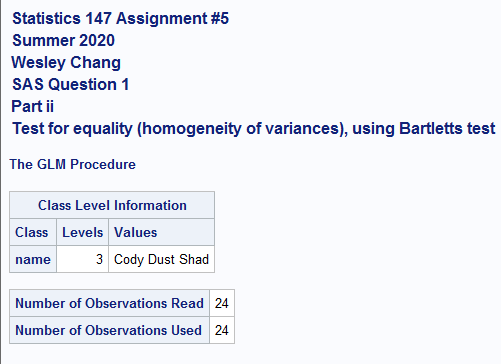


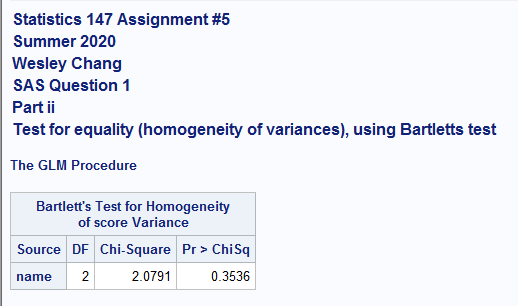


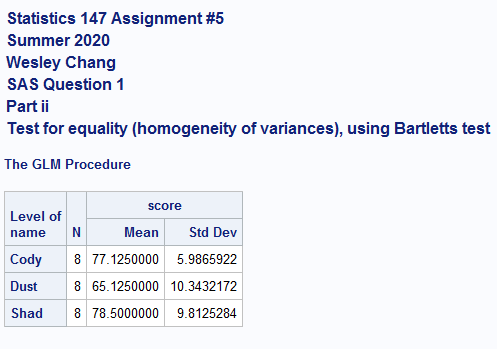
* 1. Test for equality (homogeneity) of variances. (Use Bartlett’s test.) Use *α* = 0*.*05. (2 pts)

**According to the results of Bartlett’s test, we get a p-value of 0.3536, which is larger than the level of significance 0.05. We do not reject the null hypothesis that there is equality of variances and can therefore assume homogeneity of variances.**



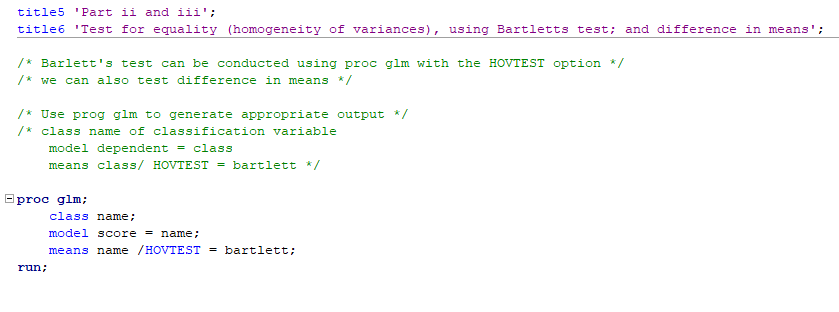


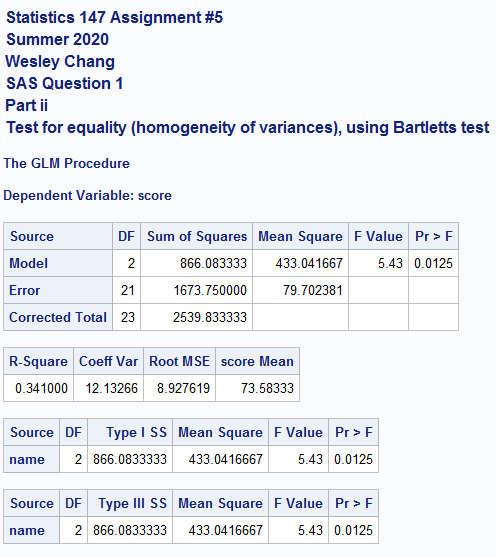




* 1. Perform the appropriate test(s) of hypothesis to determine whether one can conclude that at least one of the dogs has a significantly different mean finishing time. Use α = 0.05. (4 pts)

**From parts i and ii, we concluded that there is normality and homogeneity in variances. From this, we can continue on with testing the difference in means. The output was already generated in the proc glm function, so we refer to that section for the p-value. Since the p-value of that result is 0.0125, we reject the null hypothesis that the means are equal at the 0.05 level of significance. Therefore, we can conclude that there is a significant difference in means.**





* 1. If appropriate, use Tukey’s test and both the grouping and confidence interval methods to determine which mean(s) is(are) significantly different. (Be sure to justify your answer!) (3 pts)

**We can use Tukey’s test here since the data is independent, there is assumed normality, and assumed homogeneity in variances.**

**Confidence Interval approach:**

**Dusty-Cody: 0 does not fall into the confidence interval, so we can conclude that there is a difference in means between Dusty and Cody.**

**Shadow-Cody: 0 falls into the confidence interval, so we cannot conclude that is a difference in means between Shadow and Cody.**

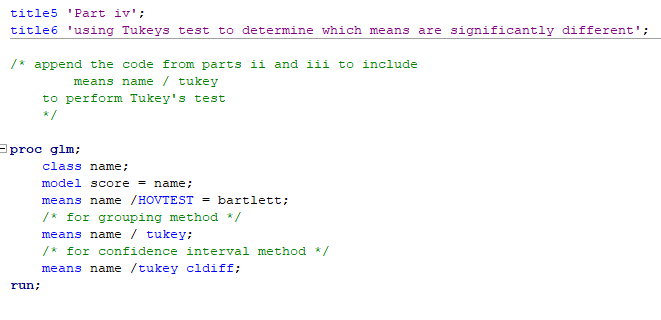
**Shadow-Dusty: 0 does not fall into the confidence interval, so we can conclude that there is a difference in means between Shadow and Dusty**

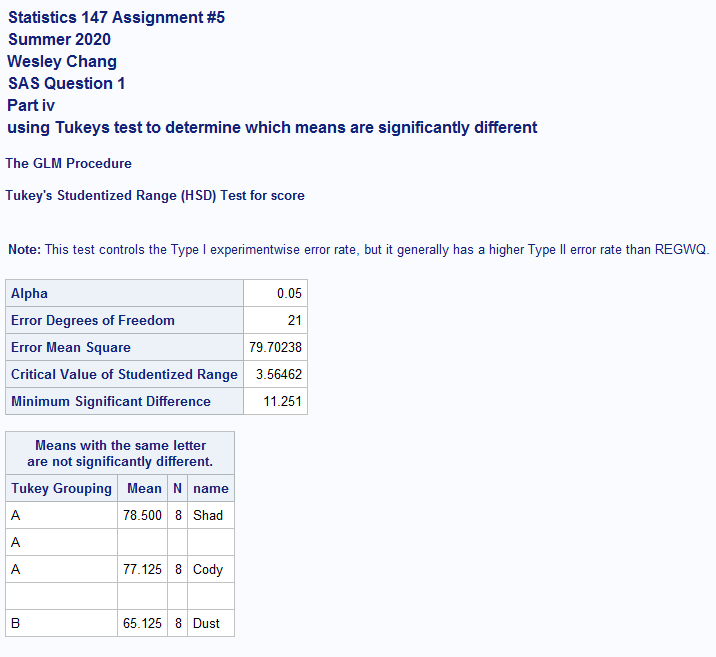
**Grouping method:**

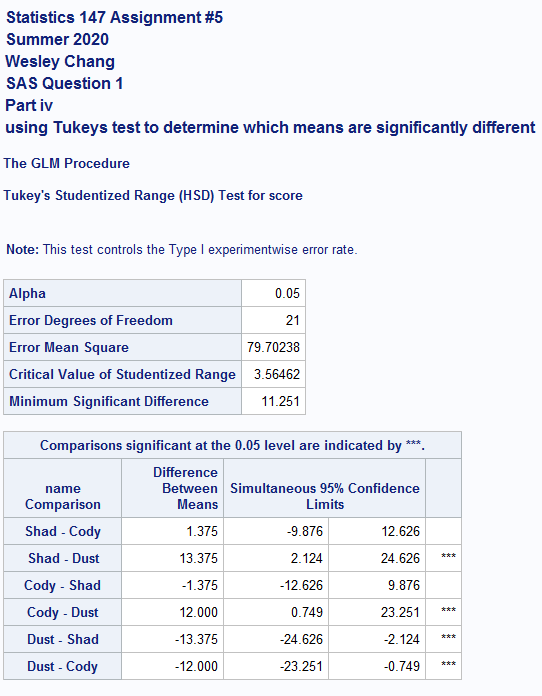
**Group A: Includes Shadow and Cody**

**Group B: Dusty**

**There is a significant difference in means between Group A, which contains Shadow and Cody and Group B, which only contains Dusty.**







**R Code:**

# Statistics 147 Assignment #5

# Summer 2020

# Wesley Chang

# open file agility.dat, read into data set agility

setwd("C:/Users/wesle/iCloudDrive/Summer 2020 (UCR)/STAT 147 (Session A)/Assignments/5")

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

agility

# attach column names, and verify

attach(agility)

names(agility)

Cody

Dusty

Shadow

# convert data frame from wide to long

agility\_long <- stack(agility)

agility\_long

# attach column names, and verify

attach(agility\_long)

names(agility\_long)

values

ind

### R Question 1

## Part i

# test for underlying normality for each of the dogs (using Anderson-Darling Test)

# use ad.test, which is in the library nortest

# call nortest library

library(nortest)

## Anderson-Darling Test

# For Cody

ad.test(Cody)

# For Dusty

ad.test(Dusty)

# For Shadow

ad.test(Shadow)

## Part ii

# test for equality (homogeneity) of variances (using Bartlett's test), use a = 0.05

# bartlett.test(x,g), where x is the set of numeric vectors, and g is the set of factor objects

bartlett.test(values,ind)

## Part iii

# perform the appropriate test(s) of hypothesis to determine wheter one can conclude that

# at least of the dogs has a significantly different mean finishing time, using a = 0.05

# appropriate test is determined by whether normality and homogeneity assumptions are satisfied

# normality satisfied as per Anderson-Darling Test

# homogeneity satisfied as per Bartlett's test

# we can proceed with testing for equality of means

# testing difference in means, we can use aov()

# we set values ~ ind to indicate corresponding variables and group

# we set data=agility\_long to indicate R to read from this data frame

results\_anova <- aov(values ~ ind, data=agility\_long)

# print data and a summary of the data

results\_anova

summary(results\_anova)

## Part iv

# if appropriate, use Tukey's test and the p-value method to determine which means(s) is(are)

# significantly different (justify answer)

# We concluded in Part iii that there is a significant difference such that not all means are equal

# we now can use Tukey's test

# TukeyHSD(results\_anova,conf.level)

# we pass on the results\_anova data frame to TukeyHSD and generate results at the 95% level

TukeyHSD(results\_anova,conf.level=0.95)

# Statistics 147 Assignment #5

# Summer 2020

# Wesley Chang

# open file agility.dat, read into data set agility

setwd("C:/Users/wesle/iCloudDrive/Summer 2020 (UCR)/STAT 147 (Session A)/Assignments/5")

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

agility

# attach column names, and verify

attach(agility)

names(agility)

Cody

Dusty

Shadow

# convert data frame from wide to long

agility\_long <- stack(agility)

agility\_long

# attach column names, and verify

attach(agility\_long)

names(agility\_long)

values

ind

### R Question 1

## Part i

# test for underlying normality for each of the dogs (using Anderson-Darling Test)

# use ad.test, which is in the library nortest

# call nortest library

library(nortest)

## Anderson-Darling Test

# For Cody

ad.test(Cody)

# For Dusty

ad.test(Dusty)

# For Shadow

ad.test(Shadow)

## Part ii

# test for equality (homogeneity) of variances (using Bartlett's test), use a = 0.05

# bartlett.test(x,g), where x is the set of numeric vectors, and g is the set of factor objects

bartlett.test(values,ind)

## Part iii

# perform the appropriate test(s) of hypothesis to determine wheter one can conclude that

# at least of the dogs has a significantly different mean finishing time, using a = 0.05

# appropriate test is determined by whether normality and homogeneity assumptions are satisfied

# normality satisfied as per Anderson-Darling Test

# homogeneity satisfied as per Bartlett's test

# we can proceed with testing for equality of means

# testing difference in means, we can use aov()

# we set values ~ ind to indicate corresponding variables and group

# we set data=agility\_long to indicate R to read from this data frame

results\_anova <- aov(values ~ ind, data=agility\_long)

# print data and a summary of the data

results\_anova

summary(results\_anova)

## Part iv

# if appropriate, use Tukey's test and the p-value method to determine which means(s) is(are)

# significantly different (justify answer)

# We concluded in Part iii that there is a significant difference such that not all means are equal

# we now can use Tukey's test

# TukeyHSD(results\_anova,conf.level)

# we pass on the results\_anova data frame to TukeyHSD and generate results at the 95% level

TukeyHSD(results\_anova,conf.level=0.95)

**SAS Code:**

title1 'Statistics 147 Assignment #5';

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 'SAS Question 1';

**data** agility;

infile 'C:\Users\wesle\iCloudDrive\Summer 2020 (UCR)\STAT 147 (Session A)\Assignments\5\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 = agility;

**run**;

/\* add code to sort the data by the name of the dog \*/

**proc** **sort** data = agility;

by dog;

**run**;

**proc** **print** noobs data = agility;

title6 'Print to check sorted';

**run**;

title5 'Part i';

title6 'test for normality';

/\* test for normality for each of the dogs, using Shapiro-Wilk test \*/

/\* use proc univariate with the normal options to test normality

Use ods select TestsForNormality to suppress printing of everything except

the tests for normality

Use "by" statement to generate test for each plant \*/

**proc** **univariate** normal;

ods select TestsForNormality;

by name;

var score;

**run**;

title5 'Part ii and iii';

title6 'Test for equality (homogeneity of variances), using Bartletts test; and difference in means';

/\* Barlett's test can be conducted using proc glm with the HOVTEST option \*/

/\* we can also test difference in means \*/

/\* Use prog glm to generate appropriate output \*/

/\* class name of classification variable

model dependent = class

means class/ HOVTEST = bartlett \*/

**proc** **glm**;

class name;

model score = name;

means name /HOVTEST = bartlett;

**run**;

title5 'Part iv';

title6 'using Tukeys test to determine which means are significantly different';

/\* append the code from parts ii and iii to include

means name / tukey

to perform Tukey's test

\*/

**proc** **glm**;

class name;

model score = name;

means name /HOVTEST = bartlett;

/\* for grouping method \*/

means name / tukey;

/\* for confidence interval method \*/

means name /tukey cldiff;

**run**;