Project 5: Two-Way ANOVA

Name:

The following SAS code may be helpful for this assignment:

Two-Way ANOVA (with interaction):

```
proc anova data=datasetname;
class varname1 varname2;
model varname3=varname1 varname2 varname1*varname2;
run;
```

Two-Way ANOVA (without interaction):

```
proc anova data=datasetname;
class varname1 varname2;
model varname3=varname1 varname2;
run;
```

Post-Hoc Tests:

```
proc anova data=datasetname;
class varname1 varname2;
model varname3=varname1 varname2;
means varname / tukey cldiff;
run;

proc glm data=datasetname;
class varname;
model varname2=varname;
means varname / bon;
run;
```

Part 1: Performing Hypothesis Tests from Data

An advertising firm specializes in TV commercials for children's products. The firm wants to determine if there are large differences in the mean length of time that the commercial is able to hold attention based on age and product type. Three age groups are used: 5-6 years (Age 1), 7-8 years (Age 2), and 9-10 years (Age 3). Two types of products are used: breakfast cereals (Product 1) and video games (Product 2). Use α=0.05 for all tests. Consider the following data:

Product	Age 1	Age 2	Age 3
Product 1	19	19	37
	36	25	6
	40	22	28
	30	28	4
	4	1	32
	10	27	16
	30	27	8
	5	16	41

	34	3	29
	21	18	18
Product 2	39	30	51
	18	47	52
	32	6	43
	22	27	48
	16	44	39
	2	26	33
	36	33	56
	43	48	43
	7	23	40
	16	21	51

1. Create a SAS data set with three variables: Age (for grouping), Product (for grouping), and Time.
2. Test for an interaction between age and product. Include your hypotheses, SAS output, SAS code, and conclusions.

   H_0: There is not an interaction between the main effect of age and product
   H_1: There is an interaction between the main effect of age and product
   We Reject H_0 if p-val < 0.05
   p-val: 0.1490
   Since our p-val is 0.1490 and 0.1490 is greater than 0.05, then we fail to reject the H_0. This means there is not sufficient evidence to suggest an interaction between age and product.

   Code:
```
data two;
input age product time @@;
cards;
1 1 19 1 1 36 1 1 40 1 1 30 1 1 4 1 1 10 1 1 30 1 1 5 1 1 34 1 1 21
2 1 19 2 1 25 2 1 22 2 1 28 2 1 1 2 1 27 2 1 27 2 1 16 2 1 3 2 1 18
1 2 39 1 2 18 1 2 32 1 2 22 1 2 16 1 2 2 1 2 36 1 2 43 1 2 7 1 2 16
2 2 30 2 2 47 2 2 6 2 2 27 2 2 44 2 2 26 2 2 33 2 2 48 2 2 23 2 2 21
;
run;
proc anova data=two;
class age product;
model time=age product age*product;
run;
quit;
```

Source	DF	Anova SS	Mean Square	F Value	Pr > F
age	1	24.0250000	24.0250000	0.15	0.6983
product	1	366.0250000	366.0250000	2.33	0.1360
age*product	1	342.2250000	342.2250000	2.17	0.1490

3. If appropriate, test the main effects of age and product. Include your hypotheses, SAS output, SAS code, and conclusions.

We will now evaluate the main effects.
H_0: There is no main effect of age
H_1: There is a main effect of age

H_0: There is no main effect of product
H_1: There is a main effect of product
We Reject H_0 if p-val < 0.05
Age p-val: 0.7027
Product p-val: 0.1418
Since both p-vals are greater than 0.05, then we fail to reject the H_0. This means there is not sufficient evidence to suggest a main effect of age or product on time watching commercials.

Code: proc anova data=two;
class age product;
model time=age product;
run;
quit;

Source	DF	Anova SS	Mean Square	F Value	Pr > F
age	1	24.0250000	24.0250000	0.15	0.7027
product	1	366.0250000	366.0250000	2.25	0.1418

4. If appropriate, perform Tukey's and Bonferroni's Post-Hoc tests. Include your hypotheses, SAS output, SAS Code, and conclusions.

Suppose we would like to examine the effect of diet and maternal age on the birthweights of newborns (in oz.). Use α=0.05 for all tests. Consider the following data:

Age	Diet 1	Diet 2	Diet 3
20 – 29	157.78	139.72	129.35
	136.79	125.47	110.73
	138.84	117.14	118.38
30 – 39	137.07	117.46	97.43
	146.28	128.54	125.26
	130.27	99.16	115.42

1. Create a SAS data set with three variables: Age (for grouping), Diet (for grouping), and weight.
2. Test for an interaction between age and diet. Include your hypotheses, SAS output, SAS code, and conclusions.

   H_0: There is not an interaction between the main effect of age and diet
   H_1: There is an interaction between the main effect of age and diet
   We Reject H_0 if p-val < 0.05
   p-val: 0.8908
   Since our p-val is 0.8908 and 0.8908 is greater than 0.05, then we fail to reject the H_0. This means there is not sufficient evidence to suggest an interaction between age and diet.

   Code:
```
data three;
input age diet weight @@;
cards;
2 1 157.78 2 2 139.72 2 3 129.35
2 1 136.79 2 2 125.47 2 3 110.73
2 1 138.84 2 2 117.14 2 3 118.38
3 1 137.07 3 2 117.46 3 3 97.43
3 1 146.28 3 2 128.54 3 3 125.26
3 1 130.27 3 2 99.16 3 3 115.42
;
run;
proc anova data=three;
class age diet;
model weight=age diet age*diet;
run;
quit;
```

Source	DF	Anova SS	Mean Square	F Value	Pr > F
age	1	332.046450	332.046450	2.38	0.1486
diet	2	2104.670533	1052.335267	7.55	0.0075
age*diet	2	32.516133	16.258067	0.12	0.8908

3. If appropriate, test the main effects of age and diet. Include your hypotheses, SAS output, SAS code, and conclusions.

We will now evaluate the main effects.
H_0: There is no main effect of age
H_1: There is a main effect of age

H_0: There is no main effect of diet
H_1: There is a main effect of diet
We Reject H_0 if p-val < 0.05
Age p-val: 0.1208
Diet p-val: 0.0036
Since the age p-val is greater than 0.05, then we fail to reject that H_0. This means there is not sufficient evidence to suggest a main effect of age on weight.
Since the diet p-val is less than 0.05, we **reject** that H_0. This means there is enough evidence to suggest a main effect of diet on weight.

Code: proc anova data=three;
class age diet;
model weight=age diet;
run;
quit;

Source	DF	Anova SS	Mean Square	F Value	Pr > F
age	1	332.046450	332.046450	2.73	0.1208
diet	2	2104.670533	1052.335267	8.65	0.0036

4. If appropriate, perform Tukey's and Bonferroni's Post-Hoc tests. Include your hypotheses, SAS output, SAS Code, and conclusions.

According to Tukey's test, we can conclude:
$\mu_1 \neq \mu_2$
$\mu_1 \neq \mu_3$
$\mu_3 \neq \mu_2$
There is no evidence to suggest a difference in weight between the means of the diets

Code: proc anova data=three;
class age diet;
model weight= age diet;
means diet / tukey cldiff;
run;
quit;

Comparisons significant at the 0.05 level are indicated by ***.				
diet Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
1 - 2	19.923	3.253	36.594	***
1 - 3	25.077	8.406	41.747	***
2 - 1	-19.923	-36.594	-3.253	***
2 - 3	5.153	-11.517	21.824	
3 - 1	-25.077	-41.747	-8.406	***
3 - 2	-5.153	-21.824	11.517	

Part 2: Using a import file for data.

Download the JHS.csv data set from eLearning.

Use the data import wizard to create a data set “JHS1”

Using this data set, use a Two-Way ANOVA to determine whether there is an effect of sex and obesity on A1CV1. Use α=0.05.

1. Test for an interaction between sex and obesity. Include your hypotheses, SAS output, SAS code, and conclusions.
   H_0: There is not an interaction between the main effect of sex and obesity
   H_1: There is an interaction between the main effect of sex and obesity
   We Reject H_0 if p-val < 0.05
   p-val: 0.3502
   Since our p-val is 0.3502 and 0.3502 is greater than 0.05, then we fail to reject the H_0. This means there is not sufficient evidence to suggest an interaction between sex and obesity.

   Code: pROC IMPORT OUT= WORK.jhs1
   DATAFILE= "G:\My Drive\STA5990Data\JHS.csv"
   DBMS=CSV REPLACE;
   GETNAMES=YES;
   DATAROW=2;
   RUN;

   proc anova data=jhs1;
   class sex obesity;

model A1cV1 = obesity sex obesity*sex;
run;
quit;

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Obesity	2	145.2229480	72.6114740	61.99	<.0001
sex	1	0.7663054	0.7663054	0.65	0.4187
sex*Obesity	2	2.4593302	1.2296651	1.05	0.3502

2. If appropriate, test the main effects of sex and obesity. Include your hypotheses, SAS output, SAS code, and conclusions.

We will now evaluate the main effects.
H_0: There is no main effect of sex
H_1: There is a main effect of sex

H_0: There is no main effect of obesity
H_1: There is a main effect of obesity
We Reject H_0 if p-val < 0.05
Sex p-val: 0.4187
Obesity p-val <.0001

Since the sex p-val is greater than 0.05, then we fail to reject that H_0. This means there is not sufficient evidence to suggest a main effect of sex on A1CV1.
Since the obesity p-val is less than 0.05, we **reject** that H_0. This means there is enough evidence to suggest an effect of obesity on A1CV1.

Code: proc anova data=jhs1;
class sex obesity;
model A1cV1 = obesity sex;
run;
quit;

3. If appropriate, perform Tukey's and Bonferroni's Post-Hoc tests. Include your hypotheses, SAS output, SAS Code, and conclusions.

According to Tukey's test, we can conclude:
$\mu_1 \neq \mu_2$
$\mu_1 \neq \mu_3$
$\mu_3 \neq \mu_2$
There is no evidence to suggest a difference in A1CV1 between the means of the obesity

Code: proc anova data=jhs1;
class sex obesity;

```
model A1cV1= sex obesity;
means obesity / tukey cldiff;
run;
quit;
```

Comparisons significant at the 0.05 level are indicated by ***.

Obesity Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
3 - 2	0.34506	0.23409	0.45603	***
3 - 1	0.67055	0.51381	0.82730	***
2 - 3	-0.34506	-0.45603	-0.23409	***
2 - 1	0.32549	0.15889	0.49210	***
1 - 3	-0.67055	-0.82730	-0.51381	***
1 - 2	-0.32549	-0.49210	-0.15889	***

Graduate students only: Use the same data set and test for a difference in fgpV1 using sex and obesity for grouping variables.

H_0: There is not an interaction between the main effect of sex and obesity
H_1: There is an interaction between the main effect of sex and obesity
We Reject H_0 if p-val < 0.05
p-val: 0.0736
Since our p-val is 0.0736 and 0.0736 is greater than 0.05, then we fail to reject the H_0. This means there is not sufficient evidence to suggest an interaction between sex and obesity.

We will now check for main effects
H_0: There is no main effect of sex
H_1: There is a main effect of sex

H_0: There is no main effect of obesity
H_1: There is a main effect of obesity
We Reject H_0 if p-val < 0.05
Sex p-val: 0.3786
Obesity p-val <.0001

Since the sex p-val is greater than 0.05, then we fail to reject that H_0. This means there is not sufficient evidence to suggest a main effect of sex on FpgV1.
Since the obesity p-val is less than 0.05, we **reject** that H_0. This means there is enough evidence to suggest an effect of obesity on FpgV1.

According to Tukey's test, we can conclude:
$\mu_1 \neq \mu_2$
$\mu_1 \neq \mu_3$
$\mu_3 \neq \mu_2$

There is no evidence to suggest a difference in FpgV1 between the means of the obesity