Project 4: Inferences for More Than Two Populations

Name:

The following SAS code may be helpful for this assignment:

ANOVA:
```
proc glm data=datasetname;
class varname;
model varname2=varname;
run;
```

Normality Assumption:
```
proc mixed data=datasetname;
class varname;
model varname2=varname / residual;
run;
```

Variances Assumption:
```
proc glm data=datasetname;
class varname;
model varname2=varname;
means varname / hovtest=bf;
run;
```

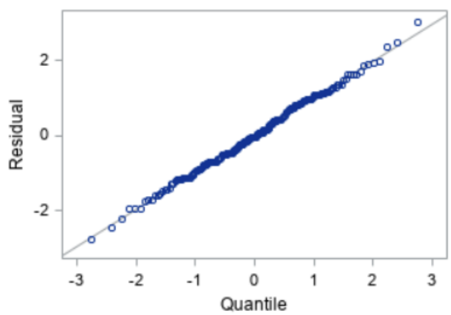
Kruskal-Wallis Test:
```
proc npar1way data=datasetname wilcoxon;
class varname;
var varname2;
run;
```

Post-Hoc Tests:
```
proc glm data=datasetname;
class varname;
model varname2=varname;
means varname / tukey;
run;

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

proc npar1way data=datasetname Wilcoxon dscf;
class varname;
var varname2;
run;
```

Run the following SAS code to create a work data set named “seven” from your p099.sas7bdat file.

```
libname in 'G:\My Drive\STA5990Data';

data seven;
set in.p099;
run;
```

1. We are interested in determining if there is a difference in diastolic blood pressure (dbp) based on income level (Income).
2. Test the assumptions of normality of the residuals using SAS. Include a screenshot of your output and discuss your findings. Include your SAS code.
   From the graph below, I say that the data is normally distributed. The points are very linear

   Code:
   proc mixed data=seven;
   class Income;
   model dbp=Income / residual;
   run;
3. Test the assumption of equality of variances at α=0.05. Include your hypotheses, SAS output, SAS code, and conclusions.
   H_0:σ^2_1= σ^2_2= σ^2_3= σ^2_4
   H_1: at least one is different
   We reject H_0 if p-val < 0.05
   p-val= 0.2548
   Since our p-val is 0.2548 and 0.2548 > 0.05, we fail to reject the H_0. This means there is not enough evidence to suggest that at least one mean is different.

Brown and Forsythe's Test for Homogeneity of dbp Variance ANOVA of Absolute Deviations from Group Medians					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Income	3	80.8236	26.9412	1.36	0.2548
Error	201	3968.9	19.7458		

   Code:
   proc glm data=seven;
   class Income;
   model dbp=Income;

```
means Income / hovtest=bf;
run;
```

4. Formally test (using ANOVA or the Kruskal-Wallis test as determined in the last part) to determine whether there at difference in the mean (or median) diastolic blood pressure in at least one of the income groups. Use α=0.05. Include your hypotheses, SAS output, SAS Code, and conclusions.
   H_0: $M_1=M_2=M_3=M_4$
   H_1: at least one median is different
   Reject H_0 if p-val < 0.05
   p-val = 0.7172
   Since the p-val is 0.7172 and 0.7172 > 0.05, we fail to reject the H_0. This means there is not enough evidence to suggest that at least one of the medians are different.

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
1.3504	3	0.7172

```
proc npar1way data=seven wilcoxon;
class Income;
var dbp;
run;
```

5. If you found a difference between at least one of the income groups, perform the appropriate post-hoc testing to determine which groups differ. For ANOVA, report Tukey and Bonferroni, for Kruskal-Wallis, report the DSCF. Use α=0.05. Include your hypotheses, SAS output, SAS Code, and conclusions.
   I didn’t find a difference between the income groups, so I will not perform a post-hoc test.
6. Were the findings what you expected? Why or why not.
   I hadn’t thought much about the link between income level and blood pressure, but the findings make sense. There isn’t a difference between blood pressure levels given an income level. Money won’t help or hurt your blood pressure.

Graduate Students: What are your assumptions about BMI among the income groups? Test your hypothesis.

I would say that there is no difference between the means of the BMI among income groups.

*** Data is normal, but variance aren’t equal, so KW test

H_0: $M_1=M_2=M_3=M_4$
H_1: at least one median is different
Reject H_0 if p-val < 0.05
p-val = 0.1965

Since the p-val is 0.1965 and 0.1965 > 0.05, we fail to reject the H_0. This means there is not enough evidence to suggest that at least one of the medians are different.

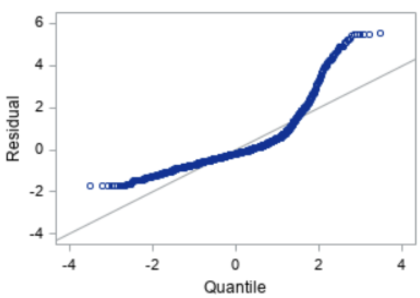
The Jackson Heart Study measured BMI, A1C, and age for participants for the first two visits of the study. Download the “JHS.csv” file from the module on eLearning and save it in your STA5990Data folder on Google Drive.

1. Use the Data Import Wizard to create a data set in SAS called “jhs2” and include your resulting SAS code here.
   I imported the data, but I don’t think I have any code to post
2. Use proc print data=jhs2 (obs=5); run; to explore the variables in the data set. Include your output here.

   The SAS System

Obs	subjid	sex	ageV1	bmiV1	A1cV1	fpgV1	ageV2	bmiV2	A1cV2	fpgV2	Obesity
1	2054	Male	63.4	26.01	5.7	94	67.5	26.75	6	96	2
2	2013	Female	56	35.13	9.7	173	60.8	38.75	7.3	125	3
3	455	Female	56.5	29.76	5.6	86	61.1	31.02	5.7	107	2
4	5	Female	40	18.66	6.1	89	44.3	18.4	5.5	90	1
5	2408	Female	47.7	37.2	5.5	98	51.8	39.44	7.2	167	3

3. We are interested to see if there is a difference A1C at the first visit (A1cV1) based on Obesity. Use α=0.05 for the test. Note: Participants are grouped into “normal weight = 1, overweight = 2, and obese = 3” in the “Obesity” variable.
4. Test the assumptions of normality of the residuals using SAS. Include a screenshot of your output and discuss your findings. Include your SAS code.
   I would say that the data is not normally distributed. It doesn’t follow a linear path

   Code:
   proc mixed data=jhs2;
   class Obesity;
   model A1cV1=Obesity / residual;
   run;
5. Test the assumption of equality of variances. Include your hypotheses, SAS output, SAS code, and conclusions.
   H_0:σ^2_1= σ^2_2= σ^2_3
   H_1: At least one if different
   We reject H_0 if p-val < 0.05
   p-val <.0001
   Since p-val <.0001 , we reject the H_0. This means that there is enough evidence to suggest that at least one of the means is different.

Brown and Forsythe's Test for Homogeneity of A1cV1 Variance ANOVA of Absolute Deviations from Group Medians					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Obesity	2	33.3123	16.6562	21.21	<.0001
Error	2562	2012.3	0.7855		

Code:

```
proc glm data=jhs2;
class Obesity;
model A1cV1=Obesity;
means Obesity / hovtest=bf;
run;
```

6. Formally test (using ANOVA or the Kruskal-Wallis test as determined in the last part) to determine whether there at difference in the mean (or median) A1C at Visit 1 among the obesity groups. Use α=0.05. Include your hypotheses, SAS output, SAS Code, and conclusions.

   ***I'm using KW because the data isn't Normal

   H_0: $M_1=M_2=M_3$

   H_1: at least one median is different

   Reject H_0 if p-val < 0.05

   p-val < 0.0001

   Since the p-val is <0.0001 and 0.0001 < 0.05, we reject the H_0. This means there is enough evidence to suggest that at least one of the medians are different.

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
175.2426	2	<.0001

Code:

```
proc npar1way data=jhs2 wilcoxon;
class Obesity;
var A1cV1;
run;
```

7. If you found a difference between A1C in at least one of the groups, perform the appropriate post-hoc testing to determine which groups differ. For ANOVA, report Tukey and Bonferroni, for Kruskal-Wallis, report the DSCF. Use α=0.05. Include your hypotheses, SAS output, SAS Code, and conclusions.

   H_0: $M_1 = M_2 = M_3$

   H_1: $M_1 \neq M_2$, $M_1 \neq M_3$, $M_3 \neq M_2$

   We reject H_0 if p-val < 0.05

   M_1 vs M_2 p-val: < 0.0001 and M_1 vs M_3 p-val: < 0.0001 and M_3 vs M_2 p-val: < 0.0001

   Since all p-vals were < 0.0001 and that is less than 0.05, we can safely reject the H_0. This means that none of the medians are equal to each other.

The SAS System

The NPAR1WAY Procedure

Pairwise Two-Sided Multiple Comparison Analysis			
Dwass, Steel, Critchlow-Fligner Method			
Variable: A1cV1			
Obesity	Wilcoxon Z	DSCF Value	Pr > DSCF
2 vs. 3	-8.4638	11.9696	<.0001
2 vs. 1	6.1721	8.7287	<.0001
3 vs. 1	11.8872	16.8110	<.0001

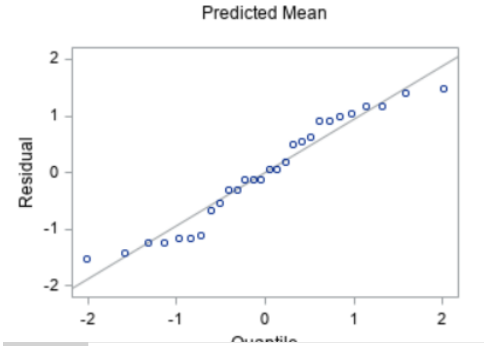
Code:

```
proc npar1way data=jhs2 wilcoxon dscf;
class Obesity;
var A1cV1;
run;
```

A research team grouped 28 participants into 4 groups. Group 1 was asked to rate their level of stress in a normal setting. Group 2 was asked to perform a mental arithmetic calculation and rate their level of stress. Group 3 was asked to perform a social role play and rate their level of stress, and Group 4 was asked to complete a mock job interview and rate their level of stress.

1	2	3	4
10	17	15	16
10	14	11	20
12	14	12	14
11	13	14	16
7	11	16	19
7	17	17	16
12	14	10	18

1. Create a SAS data set with two variables: group and stress.
2. Test the assumptions of normality of the residuals using SAS. Include a screenshot of your output and discuss your findings. Include your SAS code.
   I would say that the data is normal because the points follow a linear trend.



Code:
```
data anova2;
input group stress @@;
cards;
1 10 1 10 1 12 1 11 1 7 1 7 1 12
2 17 2 14 2 14 2 13 2 11 2 17 2 14
3 15 3 11 3 12 3 14 3 16 3 17 3 10
4 14 4 20 4 14 4 16 4 19 4 16 4 18
;
run;

proc mixed data=anova2;
class group;
model stress=group / residual;
run;
```

3. Test the assumption of equality of variances. Include your hypotheses, SAS output, SAS code, and conclusions.

H_0: $\sigma^2_1 = \sigma^2_2 = \sigma^2_3 = \sigma^2_4$

H_1: At least one if different

We reject H_0 if p-val < 0.05

p-val = 0.7841

Since p-val = 0.7841, we fail to reject the H_0. This means that there is not enough evidence to suggest that at least one of the means is different.

Brown and Forsythe's Test for Homogeneity of stress Variance ANOVA of Absolute Deviations from Group Medians					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
group	3	2.1071	0.7024	0.36	0.7841
Error	24	47.1429	1.9643		

Code:
```
proc glm data=anova2;
class group;
model stress=group;
means group / hovtest=bf;
run;
```

4. Formally test (using ANOVA or the Kruskal-Wallis test as determined in the last part) to determine whether there at difference in the mean (or median) stress level

between the groups. Use α=0.05. Include your hypotheses, SAS output, SAS Code, and conclusions.

***I'm using KW because the variance isn't equal

H_0: $M_1=M_2=M_3=M_4$

H_1: at least one median is different

Reject H_0 if p-val < 0.05

p-val = 0.0019

Since the p-val is 0.0019 and 0.0019< 0.05, we reject the H_0. This means there is enough evidence to suggest that at least one of the medians are different.

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
14.8658	3	0.0019

Code:

```
proc npar1way data=anova2 wilcoxon;
class group;
var stress;
run;
```

5. If you found a difference between at least one of the groups, perform the appropriate post-hoc testing to determine which groups differ. For ANOVA, report Tukey and Bonferroni, for Kruskal-Wallis, report the DSCF. Use α=0.05. Include your hypotheses, SAS output, SAS Code, and conclusions.

H_0: $M_1 = M_2 = M_3 = M_4$

H_1: $M_1 \neq M_2$, $M_1 \neq M_3$, $M_3 \neq M_2$, $M_1 \neq M_4$, $M_4 \neq M_3$, $M_4 \neq M_2$

We reject H_0 if p-val < 0.05

M_1 vs M_2 p-val: 0.0234 and M_1 vs M_3 p-val: 0.1242 and M_3 vs M_2 p-val: 0.9687 and M_1 vs M_4 p-val: 0.0089 and M_4 vs M_3 p-val: .2414 and M_4 vs M_2 p-val: 0.2887

Because of the p-vals that were less than 0.05, we can say that $M_1 \neq M_2$ and $M_1 \neq M_4$. We don't have enough evidence to say anything about the other median comparisons.

Pairwise Two-Sided Multiple Comparison Analysis			
Dwass, Steel, Critchlow-Fligner Method			
Variable: stress			
group	Wilcoxon Z	DSCF Value	Pr > DSCF
1 vs. 2	-2.8393	4.0153	0.0234
1 vs. 3	-2.1965	3.1063	0.1242
1 vs. 4	-3.1478	4.4517	0.0089
2 vs. 3	0.4548	0.6431	0.9687
2 vs. 4	-1.7682	2.5007	0.2887
3 vs. 4	-1.8692	2.6435	0.2414

Code:

```
proc npar1way data=anova2 wilcoxon dscf;
class group;
var stress;
```

run;