Practice and Participation 3: Inferences for Two Populations

Name: Key

The following SAS code may be helpful for this assignment:

Independent t-test:

```
proc ttest data=datasetname h0=0;
class varname;
var varname2;
run;
```

Wilcoxon Rank Sum test:

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

Dependent t-test:

```
proc ttest data=datasetname;
paired var1*var2;
run;
```

Wilcoxon Signed Rank test:

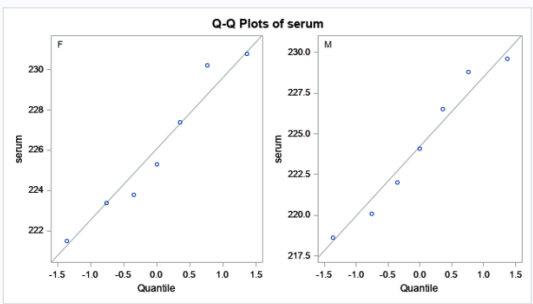
```
data datasetname;
set datasetname;
diff=reported-measured;
run;
proc univariate data=datasetname;
var diff;
run;
```

Part 1: Performing Hypothesis Tests from Data

Using the following data, test whether male and female turtles have different mean serum cholesterol concentrations.

Male Turtles	220.1	218.6	229.6	228.8	222.0	224.1	226.5
Female Turtles	223.4	221.5	230.2	225.3	223.8	230.8	227.4

1. Is this independent or paired data? Why? This is independent as there is no implied relationship between the participants.
2. Which *t*-test would be appropriate to compare the two groups? Independent t-test.
3. Check the assumptions for the chosen test. Paste your output and discuss your conclusions here.



Normality: Data points follow the line closely. The normality assumption is met.

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	6	6	1.45	0.6638

Equality of Variances:

I have not asked for the hypotheses and conclusions for this test, but if you include it that's ok. Either way, you need to discuss that the assumption for equality of variances is met. If you do the formal test:

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Since p=0.6638 is not less than 0.05, we fail to reject the null and conclude that we do not have evidence of a differences between the variances. The assumption is met.

We can do a t-test based on the assumptions.

4. Formally test (using a *t*-test or a nonparametric test as determined in the last part) to determine whether the mean serum concentrations are different. Use α=0.05. Include your hypotheses, SAS output, SAS Code, and conclusions.

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 - \mu_2 \neq 0$$

```
proc ttest data=eleven h0=0;
class sex;
var serum;
     run;
```

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	12	0.87	0.4025
Satterthwaite	Unequal	11.61	0.87	0.4031

The p-value (pooled) is 0.4025 which is not less that 0.05 so we fail to reject H0. We conclude that there is no evidence of a differences in the mean serum concentrations.

5. If you conducted a *t*-test, construct a 95% confidence interval for the difference between the means.
(-2.7408, 6.3694)

First, create a temporary dataset named “eleven” using the following code:

```
data eleven;
input sex serum @@;
cards;
1 220.1 1 218.6 1 229.6 1 228.8 1 222.0 1 224.1 1 226.5
2 223.4 2 221.5 2 230.2 2 225.3 2 223.8 2 230.8 2 227.4
;
run;
```

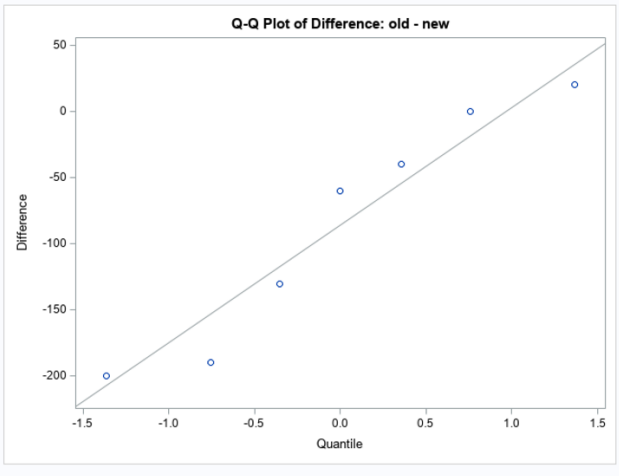
Next, customize the code at the beginning of this document to answer questions 3 – 5 above. Paste your code below.

```
proc ttest data=eleven h0=0;
class sex;
var serum;
     run;
```

A farmer began using a new fertilizer in his fields and recorded yield from the old and new fertilizer in 7 plots of land. Determine whether the new fertilizer increased the output of the fields. Use α=0.01.

Plot	1	2	3	4	5	6	7
Old Fertilizer	1920	2020	2060	1960	1960	2140	1980
New Fertilizer	2050	2210	2060	2000	2160	2120	2040

1. Is this independent or paired data? Why? This is paired (dependent) data. Single plots with two different treatments.
2. Which *t*-test would be appropriate to compare the two groups? Dependent t-test
3. Check the assumptions for the chosen test. Paste your output and discuss your conclusions here.

Normality: 

Data are close to the line. The normality assumption is met. We can use the t-test.

4. Formally test (using a *t*-test or a nonparametric test as determined in the last part) to determine whether the fertilizer increased the yield. Use α=0.01. Include your hypotheses, SAS output, SAS Code, and conclusions.

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 - \mu_2 \neq 0$$

```
data fert;
input old new @@;
cards;
1920 2050 2020 2210 2060 2060 1960 2000 1960 2160 2140 2120 1980 2040
;
proc ttest data=fert h0=0;
paired old*new;
run;
```

DF	t Value	Pr > |t|
6	-2.56	0.0430

The p-value we calculated is 0.0430 which is not less than 0.01. We fail to reject H0 and conclude that there is no evidence of a difference between the fertilizers.

5. If you conducted a *t*-test, construct a 99% confidence interval for the difference between the means.
(-210.0, 38.5332)

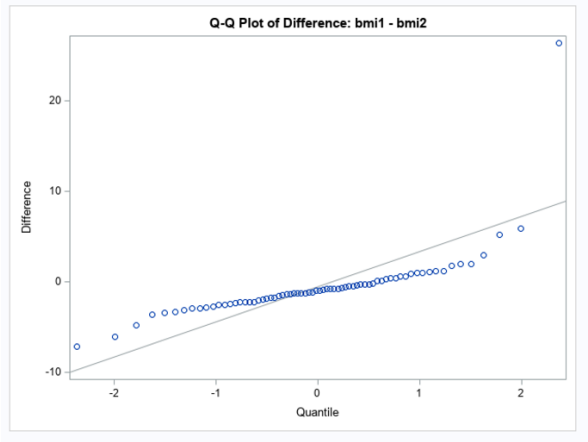
The following data is a sample of BMI measurements from 70 participants in the Jackson Heart Study from Visit 1 and Visit 2. Determine whether there is a difference in BMI between visit 1 and visit 2. Use α=0.05. (Note: you can copy and paste data after the cards statement in SAS.)

bmiV1	bmiV2
26.01	26.75
35.13	38.75
29.76	31.02
18.66	18.4
37.2	39.44
32.03	30.81
31.91	34.61
30.39	32.67
32.38	29.48
40.07	42.34
30.45	30.05
38.08	38.56

33.79	39.82
31.81	29.86
22.53	23.86
18.53	18.4
31.78	30.92
21.82	20.69
23.11	25.65
38.54	45.66
26.27	28.58
46.4	20.05
35.89	38.7
24.6	26.54
26.4	28.7
24.92	24.83
43.24	44.38
36.65	38.42
28.11	29.9
36.42	38.86
29.86	28.13
27.12	27.27
48.26	43.09
46.38	47.17
27.48	30.02
24.49	25.39
28.17	28.69
23.93	24.61
43.94	47.09
27.51	27.83
36.81	38.28
25.28	28.23
24.88	25.68
31.21	32.48
29.44	30.68
26.58	25.47
20.72	21.72
31.63	35.05
21.58	22.59
37.8	31.91
41.18	40.16

23.87	25.09
38.03	37.04
42.46	47.23
39.52	41.34
29.88	29.48
28.15	28.44
23.54	25.09
24.58	24.02
22.06	24.96
46.93	44.93
29.48	31.52
34.35	33.78
25.89	27.3
38.48	41.81
32.43	32.85
43.35	44.1
27.58	28.86
37.68	38.31
28.87	29.13

1. Is this independent or paired data? Why? This is paired data. Individual participants measured twice.
2. Which *t*-test would be appropriate to compare the two groups? Paired (dependent) t-test.
3. Check the assumptions for the chosen test. Paste your output and discuss your conclusions here.
   Normality:



Data trend off the line. This is not normal data. We need to change to the Wilcoxon Signed Rank Test.

4. Formally test (using a *t*-test or a nonparametric test as determined in the last part) to determine whether there is a difference in BMI. Use α=0.05. Include your hypotheses, SAS output, SAS Code, and conclusions.

$$H_0: M_1 - M_2 = 0$$
$$H_1: M_1 - M_2 \neq 0$$

```
data nine;
set jhs;
diff=bmi1-bmi2;
run;
proc univariate data=nine;
      var diff; run;
```

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	-1.21125	Pr > |t|	0.2299
Sign	M	-15	Pr >= |M|	0.0004
Signed Rank	S	-643	Pr >= |S|	<.0001

The p-value for the signed rank is <0.0001 which is less than 0.05. We reject H0 and conclude that there is evidence of a difference in BMI from visit 1 to visit 2.

5. If you conducted a *t*-test, construct a 95% confidence interval for the difference between the means. N/A

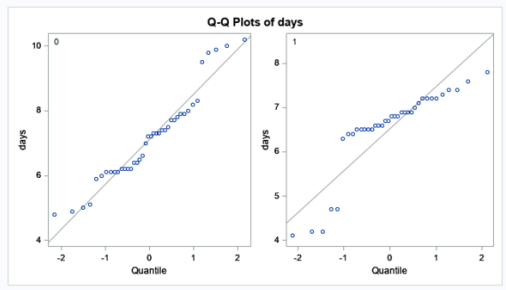
The following data is recorded a sample of 75 patients and their lengths of stay (in days) after a major surgical procedure. Data includes data from two different hospitals. Determine whether patients are being kept in the hospital for a different amount of time based on the hospital they choose for surgery. Use α=0.01.

Days	Hospital
6.6	0
6.4	0
6.5	0
6.4	0
6.2	0
6.1	0
6.2	0
6.1	0
6.2	0
6.2	0
6.1	0
6.1	0

6	0
5.9	0
5.1	0
5	0
4.9	0
4.8	0
4.7	1
4.7	1
4.2	1
4.2	1
4.1	1
9.5	0
8.3	0
8.2	0
7.8	1
7.6	1
7.4	1
7.3	1
7.2	1
7.2	1
6.9	1
6.9	1
6.8	1
6.8	1
6.7	1
6.6	1
6.5	1
6.5	1
6.5	1
6.5	1
6.5	1
6.4	1
6.4	1
6.3	1
8	0
7.9	0
7.9	0
7.8	0
7.7	0

7.4	1
7.2	1
7.2	1
7.1	1
7	1
6.9	1
6.9	1
6.8	1
6.7	1
6.6	1
6.6	1
10.2	0
10	0
9.9	0
9.8	0
7.4	0
7.2	0
7	0
7.3	0
7.3	0
7.3	0
7.2	0
7.4	0
7.5	0
7.7	0

1. Is this independent or paired data? Why? Independent. There is not an implied relationship between patients or hospitals.
2. Which *t*-test would be appropriate to compare the two groups? Independent t-test.
3. Check the assumptions for the chosen test. Paste your output and discuss your conclusions here.
   Normality:



Data in sample 1 is pretty far off normality. We should be cautious and change to the Wilcoxon Rank Sum. Since this

assumption is broken, we do not need to test variances. If we did, p=0.0246 which is less than 0.05 so the variances are not equal.

4. Formally test (using a *t*-test or a nonparametric test as determined in the last part) to determine whether the mean hospital stay times are different. Use α=0.01. Include your hypotheses, SAS output, SAS Code, and conclusions.

$$H_0: M_1 - M_2 = 0$$
$$H_1: M_1 - M_2 \neq 0$$

```
proc npar1way data=hosp wilcoxon;
class hospital;
var days;
     run;
```

Wilcoxon Two-Sample Test					
				t Approximation	
Statistic	Z	Pr < Z	Pr > |Z|	Pr < Z	Pr > |Z|
1253.500	-1.3746	0.0846	0.1692	0.0867	0.1733
Z includes a continuity correction of 0.5.					

The p-value is 0.1733 which is not less than 0.01. We fail to reject H0. There is no evidence of a difference in the length of stays between the hospitals.

5. If you conducted a *t*-test, construct a 99% confidence interval for the difference between the means. N/A

Part 2: Using a .sas7bdat file for data.

Run the following SAS code to create a work data set named “twelve” from your p099.sas7bdat file. Note: you will be using a similar file for this week’s project. This will be included in the project instructions.

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

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

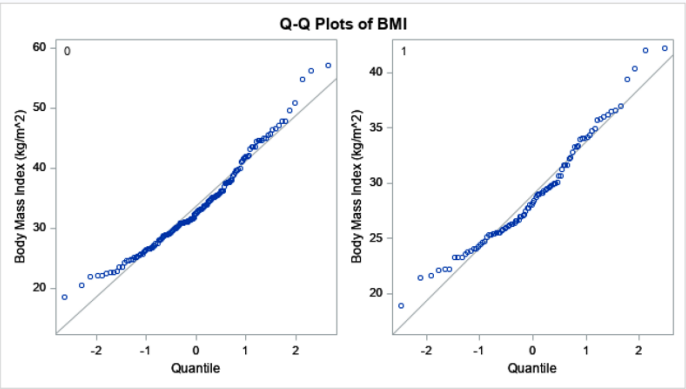
Paste your output here:

Use this data set to determine whether there is a difference in BMI between males and females. Use α=0.01.

1. Is this independent or paired data? Why? Independent. There is no implied relationship between participants.

2. Which *t*-test would be appropriate to compare the two groups? Independent t-test.
3. Check the assumptions for the chosen test. Paste your output and discuss your conclusions here.

Normality:



The trend in both of these are slightly curved. I would suggest being cautious and using the nonparametric test. Especially if the variances are unequal.

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	154	94	2.47	<.0001

For the variances assumption, p<0.0001. We would conclude that the variances are unequal. The assumption is not met.

4. Formally test (using a *t*-test or a nonparametric test as determined in the last part) to determine whether the mean BMIs are different. Use α=0.01. Include your hypotheses, SAS output, SAS Code, and conclusions.

$$H_0: M_1 - M_2 = 0$$
$$H_1: M_1 - M_2 \neq 0$$

Wilcoxon Two-Sample Test					
				t Approximation	
Statistic	Z	Pr < Z	Pr > |Z|	Pr < Z	Pr > |Z|
9112.500	-5.0624	<.0001	<.0001	<.0001	<.0001
Z includes a continuity correction of 0.5.					

The p-value<0.0001 which is less than 0.01. We reject H0 and conclude that there is evidence of a difference in BMI between males and females.

5. If you conducted a *t*-test, construct a 99% confidence interval for the difference between the means. N/A

Paste your SAS code here:

```
proc npar1way data=twelve wilcoxon;
class male;
var bmi;
run;
```