## Project 3: Inferences for Two Populations

**Name:**

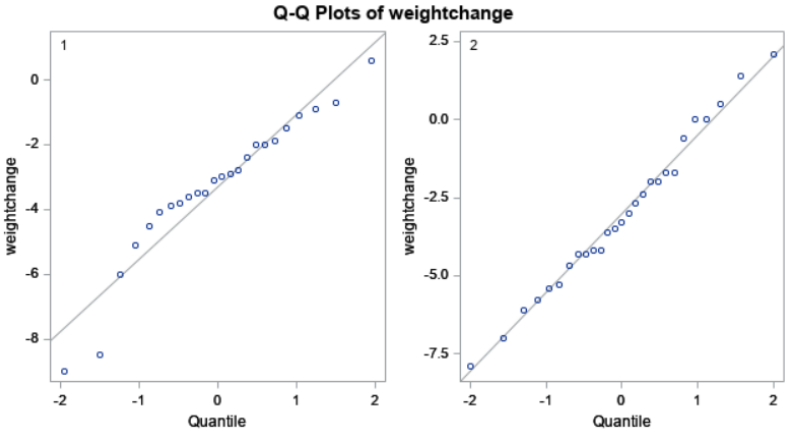
Fifty-one participants were placed on one of two different diets for 6 weeks. Determine whether the weight lost between the two diets was different. Use $\alpha$=0.05. Note: you can copy and paste the data (not the headings) into SAS.

| Diet | weightchange |
|---|---|
| 1 | -3.8 |
| 1 | -6 |
| 1 | -0.7 |
| 1 | -2.9 |
| 1 | -2.8 |
| 1 | -2 |
| 1 | -2 |
| 1 | -8.5 |
| 1 | -1.9 |
| 1 | -3.1 |
| 1 | -1.5 |
| 1 | -3 |
| 1 | -3.6 |
| 1 | -0.9 |
| 1 | 0.6 |
| 1 | -1.1 |
| 1 | -4.5 |
| 1 | -4.1 |
| 1 | -9 |
| 1 | -2.4 |
| 1 | -3.9 |
| 1 | -3.5 |
| 1 | -5.1 |
| 1 | -3.5 |
| 2 | 0 |
| 2 | 0 |
| 2 | 2.1 |
| 2 | -2 |
| 2 | -1.7 |
| 2 | -4.3 |

| | |
|---|---|
| 2 | -7 |
| 2 | -0.6 |
| 2 | -2.7 |
| 2 | -3.6 |
| 2 | -3 |
| 2 | -2 |
| 2 | -4.2 |
| 2 | -4.7 |
| 2 | -3.3 |
| 2 | 0.5 |
| 2 | -4.2 |
| 2 | -2.4 |
| 2 | -5.8 |
| 2 | -3.5 |
| 2 | -5.3 |
| 2 | -1.7 |
| 2 | -5.4 |
| 2 | -6.1 |
| 2 | -7.9 |
| 2 | 1.4 |
| 2 | -4.3 |

Customize the code in the Practice and Participation document to create a data set.

- Is this independent or paired data? Why?
  - This data is independent because it cannot be linked from one group to the other
- Which *t*-test would be appropriate to compare the two groups?
  - Independent ttest
- Check the assumptions for the chosen test. Paste your output and discuss your conclusions here.
  - Data appears to be linear, therefore we can conclude normality (diet 2 at least is extremely linear)

- 

- 
- Formally test (using a *t*-test or a nonparametric test as determined in the last part) to determine whether the mean weight losses are different. Use $\alpha$=0.05. Include your hypotheses, SAS output, SAS Code, and conclusions.
  - H0: $\mu_1 - \mu_2 = 0$  H1: $\mu_1 - \mu_2 \neq 0$
  - We will reject H0 if p < 0.05
  - P-val = 0.6851
  - Since our p-val is 0.6851 and 0.6851 > 0.05, we fail to reject the H0. So there is not enough evidence to suggest that the weight loss between the two diets are different.

| Method | Variances | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|
| Pooled | Equal | 49 | -0.41 | 0.6851 |
| Satterthwaite | Unequal | 49 | -0.41 | 0.6829 |

| Equality of Variances | | | | |
|---|---|---|---|---|
| Method | Num DF | Den DF | F Value | Pr > F |
| Folded F | 26 | 23 | 1.27 | 0.5675 |

  - **Code: data three;**
  - **input diet weightchange @@;**
  - **cards;**
  - **1 -3.8**
  - **1 -6**
  - **1 -0.7**
  - **1 -2.9**
  - **1 -2.8**
  - **1 -2**
  - **1 -2**
  - **1 -8.5**
  - **1 -1.9**

- **1 -3.1**
- **1 -1.5**
- **1 -3**
- **1 -3.6**
- **1 -0.9**
- **1 0.6**
- **1 -1.1**
- **1 -4.5**
- **1 -4.1**
- **1 -9**
- **1 -2.4**
- **1 -3.9**
- **1 -3.5**
- **1 -5.1**
- **1 -3.5**
- **2 0**
- **2 0**
- **2 2.1**
- **2 -2**
- **2 -1.7**
- **2 -4.3**
- **2 -7**
- **2 -0.6**
- **2 -2.7**
- **2 -3.6**
- **2 -3**
- **2 -2**
- **2 -4.2**
- **2 -4.7**
- **2 -3.3**
- **2 0.5**
- **2 -4.2**
- **2 -2.4**
- **2 -5.8**
- **2 -3.5**
- **2 -5.3**
- **2 -1.7**
- **2 -5.4**
- **2 -6.1**
- **2 -7.9**
- **2 1.4**
- **2 -4.3**
- **;**
- **run;**
- **proc print data=three;**
- **run;**

    - **proc ttest data=three h0=0;**
    - **class diet;**
    - **var weightchange;**
    - **run;**
- If you conducted a *t*-test, construct a 95% confidence interval for the difference between the means.
    - -1.6241 < μ < 1.0759

---

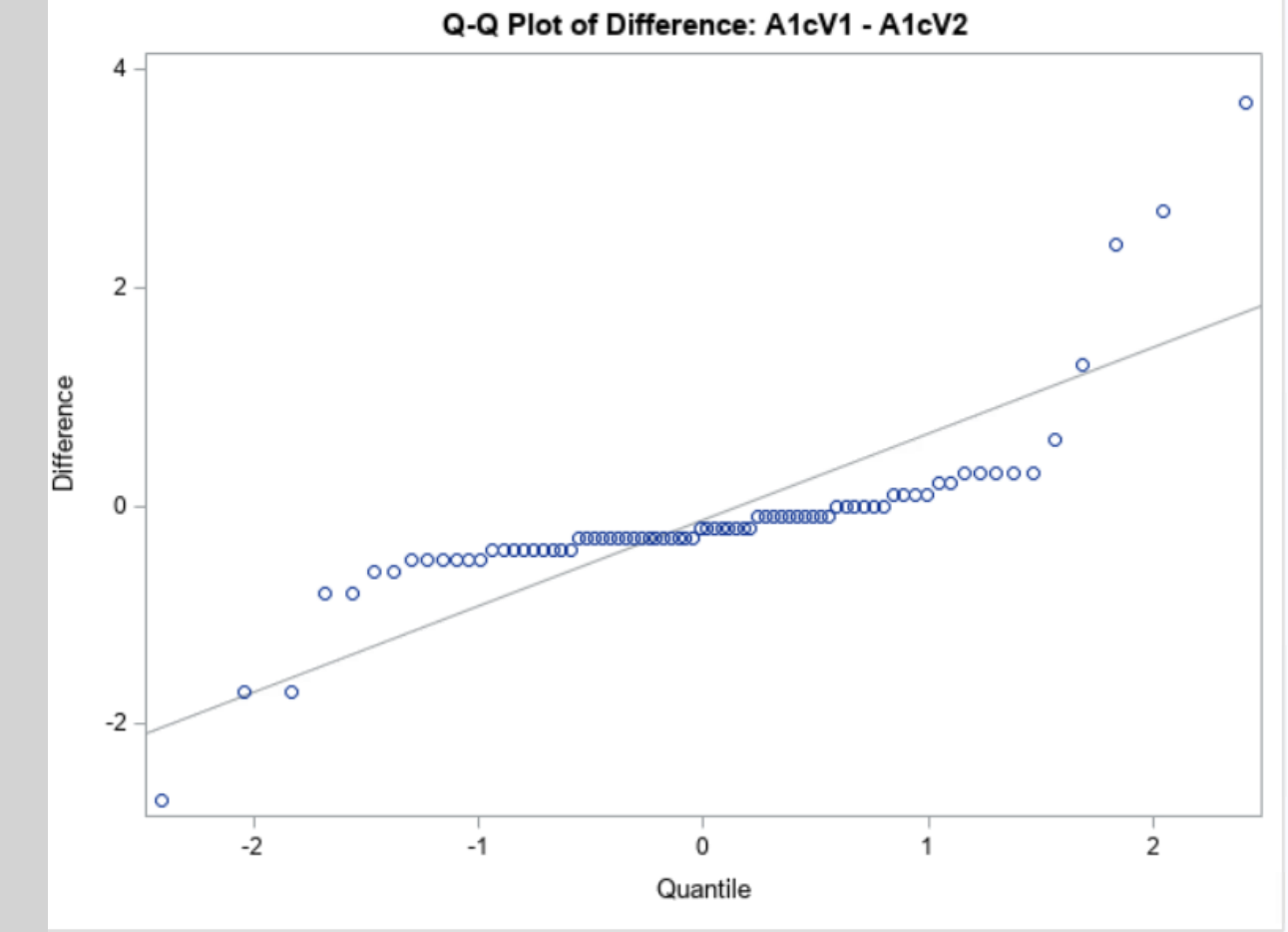
The data below is from a sample of participants in the Jackson Heart Study and includes A1C measures from their first and second visits. Test whether there is a difference in A1C measures between visits one and two. Use $\alpha$=0.05.

| A1cV1 | A1cV2 |
|---|---|
| 5.7 | 6 |
| 9.7 | 7.3 |
| 5.6 | 5.7 |
| 6.1 | 5.5 |
| 5.5 | 7.2 |
| 6.2 | 5.9 |
| 5.5 | 6 |
| 5.6 | 6 |
| 6 | 6 |
| 6.9 | 5.6 |
| 6.2 | 5.9 |
| 9.1 | 11.8 |
| 5.5 | 6 |
| 5.3 | 5.3 |
| 5.2 | 5.8 |
| 5.7 | 6 |
| 6 | 5.9 |
| 4.6 | 4.8 |
| 5.2 | 5.3 |
| 5.1 | 5.2 |
| 5.3 | 5.4 |
| 5.1 | 5.6 |
| 5.7 | 6.1 |
| 5.5 | 6 |
| 5.4 | 6 |
| 5.5 | 5.7 |

| | |
|---|---|
| 11.1 | 8.4 |
| 5.1 | 5.4 |
| 5.6 | 5.6 |
| 5.8 | 6.2 |
| 5 | 5.2 |
| 6.2 | 6.4 |
| 6.5 | 7.3 |
| 5.4 | 5.8 |
| 5.3 | 5.7 |
| 5.6 | 5.6 |
| 5.3 | 5.6 |
| 5.5 | 5.7 |
| 5.9 | 6.2 |
| 5.6 | 6 |
| 5.1 | 5.5 |
| 4.4 | 4.5 |
| 5.7 | 5.6 |
| 6.3 | 6.8 |
| 6 | 6.1 |
| 10.7 | 7 |
| 5.3 | 5.2 |
| 5.1 | 5.4 |
| 5.2 | 5.3 |
| 5.8 | 6 |
| 6.2 | 6.7 |
| 5.1 | 5.4 |
| 5.5 | 5.5 |
| 5.5 | 5.8 |
| 5.6 | 5.6 |
| 5.4 | 5.6 |
| 5.3 | 5.6 |
| 5.3 | 5 |
| 5.4 | 5.7 |
| 6.5 | 6.8 |
| 6.4 | 6.8 |
| 5.2 | 5.3 |
| 7.4 | 9.1 |
| 5.6 | 6 |
| 6.1 | 5.9 |

| | |
|---|---|
| 6.1 | 6.4 |
| 5.2 | 5.5 |
| 6 | 6.3 |
| 5.7 | 5.8 |
| 5.7 | 6.5 |
| 5.5 | 5.8 |
| 5.4 | 5.7 |
| 5.9 | 5.7 |
| 6.2 | 5.9 |
| 5 | 5.2 |
| 5 | 4.9 |
| 4.7 | 4.8 |
| 6.2 | 5.9 |
| | |

- Is this independent or paired data? Why?
  - This data is paired because the same person is getting tested on both visits
- Which *t*-test would be appropriate to compare the two groups?
  - Paired ttest
- Check the assumptions for the chosen test. Paste your output and discuss your conclusions here.
  - The data is not normal since it doesn't fit a linear trend. We will conduct a Wilcoxon Signed Rank test




- Formally test (using a *t*-test or a nonparametric test as determined in the last part) to determine whether the A1C levels are different. Use $\alpha$=0.05. Include your hypotheses, SAS output, SAS Code, and conclusions.
  - H0: Distribution of A1c levels are the same between the first and second visit.
  - H1: Distribution of A1c levels are different between the first and second visit.
  - We will reject H0 if $p < 0.05$
  - Our p-val is less than 0.0001
  - Since our p-val is less than 0.0001 and this is less than 0.05, then we reject the H0. This means there is enough evidence to suggest that the distribution of A1c levels are different between the first and second visit.

| Tests for Location: Mu0=0 | | | | |
|---|---|---|---|---|
| Test | Statistic | | p Value | |
| Student's t | t | -1.42015 | Pr > \|t\| | 0.1596 |
| Sign | M | -20 | Pr >= \|M\| | <.0001 |
| Signed Rank | S | -744.5 | Pr >= \|S\| | <.0001 |

- **Code:** **data four;**
- **input A1cV1 A1cV2 @@;**
- **cards;**
- **5.7 6**
- **9.7 7.3**
- **5.6 5.7**
- **6.1 5.5**
- **5.5 7.2**
- **6.2 5.9**
- **5.5 6**
- **5.6 6**
- **6 6**
- **6.9 5.6**
- **6.2 5.9**
- **9.1 11.8**
- **5.5 6**
- **5.3 5.3**
- **5.2 5.8**
- **5.7 6**
- **6 5.9**
- **4.6 4.8**
- **5.2 5.3**
- **5.1 5.2**
- **5.3 5.4**
- **5.1 5.6**
- **5.7 6.1**
- **5.5 6**
- **5.4 6**
- **5.5 5.7**
- **11.1 8.4**
- **5.1 5.4**
- **5.6 5.6**
- **5.8 6.2**
- **5 5.2**
- **6.2 6.4**
- **6.5 7.3**
- **5.4 5.8**

- **5.3 5.7**
- **5.6 5.6**
- **5.3 5.6**
- **5.5 5.7**
- **5.9 6.2**
- **5.6 6**
- **5.1 5.5**
- **4.4 4.5**
- **5.7 5.6**
- **6.3 6.8**
- **6 6.1**
- **10.7 7**
- **5.3 5.2**
- **5.1 5.4**
- **5.2 5.3**
- **5.8 6**
- **6.2 6.7**
- **5.1 5.4**
- **5.5 5.5**
- **5.5 5.8**
- **5.6 5.6**
- **5.4 5.6**
- **5.3 5.6**
- **5.3 5**
- **5.4 5.7**
- **6.5 6.8**
- **6.4 6.8**
- **5.2 5.3**
- **7.4 9.1**
- **5.6 6**
- **6.1 5.9**
- **6.1 6.4**
- **5.2 5.5**
- **6 6.3**
- **5.7 5.8**
- **5.7 6.5**
- **5.5 5.8**
- **5.4 5.7**
- **5.9 5.7**
- **6.2 5.9**
- **5 5.2**
- **5 4.9**
- **4.7 4.8**
- **6.2 5.9**
- **;**
- **run;**

- **proc print data=four;**
- **run;**
- **proc ttest data=four sides=2;**
- **paired A1cV1*A1cV2;**
- **run;**
- **data four;**
- **set four;**
- **diff=A1cV1-A1cV2;**
- **run;**
- **proc univariate data=four;**
- **var diff;**
- **run;**

- If you conducted a *t*-test, construct a 95% confidence interval for the difference between the means.
  - Did not conduct a ttest, I conducted a Wilcoxon Signed Rank test

---

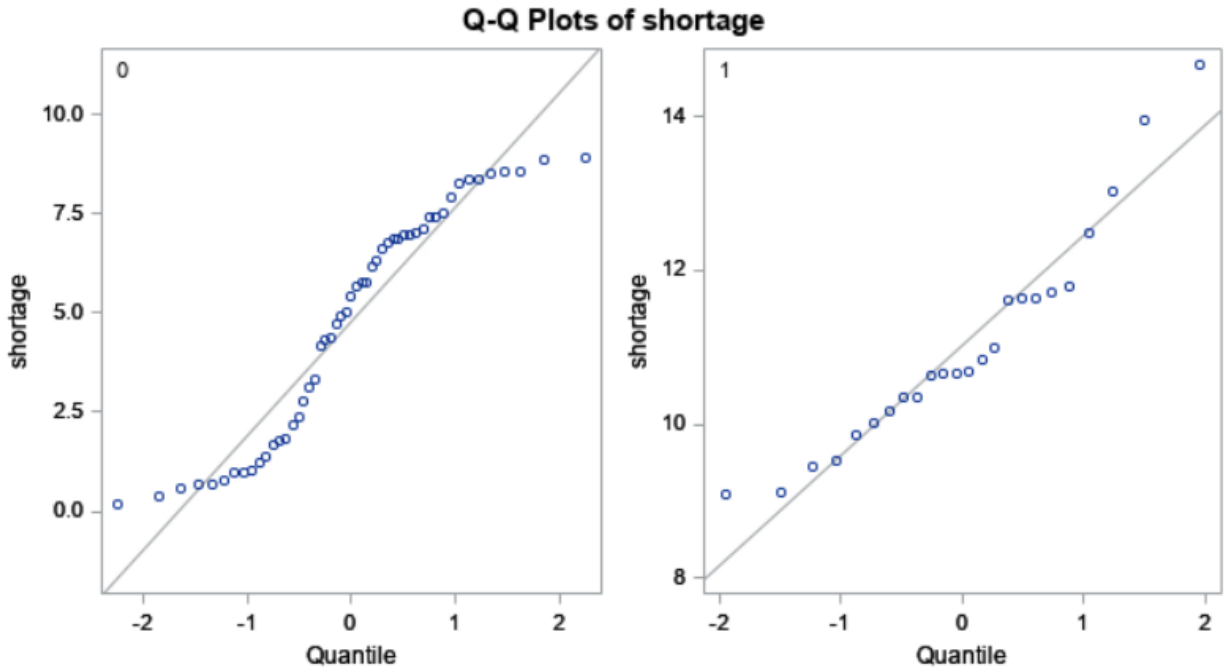
Below is a sample of the average nursing shortage of hospitals in each county of two states on a particular day. Test whether the two states are having different levels of staffing shortages. Note: The number of nurses each county is short is the sum of the shortage divided by the number of hospitals. Use $\alpha$=0.05.

| Shortage | State |
|---|---|
| 6.84 | 0 |
| 7 | 0 |
| 6.97 | 0 |
| 7.11 | 0 |
| 7.53 | 0 |
| 10.36 | 1 |
| 10.65 | 1 |
| 10.65 | 1 |
| 10.69 | 1 |
| 10.62 | 1 |
| 10.84 | 1 |
| 11.78 | 1 |
| 11.71 | 1 |
| 11.65 | 1 |
| 11.6 | 1 |
| 11.64 | 1 |
| 0.69 | 0 |
| 0.78 | 0 |

| | |
|---|---|
| 0.99 | 0 |
| 1.05 | 0 |
| 1.21 | 0 |
| 1.37 | 0 |
| 1.84 | 0 |
| 4.19 | 0 |
| 4.71 | 0 |
| 4.92 | 0 |
| 5.74 | 0 |
| 6.17 | 0 |
| 6.75 | 0 |
| 7.91 | 0 |
| 8.26 | 0 |
| 8.53 | 0 |
| 8.87 | 0 |
| 8.89 | 0 |
| 9.52 | 1 |
| 10.02 | 1 |
| 0.19 | 0 |
| 0.39 | 0 |
| 0.58 | 0 |
| 0.68 | 0 |
| 0.97 | 0 |
| 1.66 | 0 |
| 3.13 | 0 |
| 4.37 | 0 |
| 5.01 | 0 |
| 5.74 | 0 |
| 6.3 | 0 |
| 6.86 | 0 |
| 6.95 | 0 |
| 7.42 | 0 |
| 7.41 | 0 |
| 8.34 | 0 |
| 8.52 | 0 |
| 9.44 | 1 |
| 10.35 | 1 |
| 5.43 | 0 |
| 8.56 | 0 |

| | |
|---|---|
| 9.09 | 1 |
| 10.18 | 1 |
| 1.8 | 0 |
| 2.19 | 0 |
| 2.38 | 0 |
| 2.76 | 0 |
| 3.34 | 0 |
| 4.31 | 0 |
| 5.66 | 0 |
| 6.61 | 0 |
| 8.34 | 0 |
| 9.1 | 1 |
| 9.85 | 1 |
| 10.99 | 1 |
| 12.5 | 1 |
| 13.04 | 1 |
| 13.96 | 1 |
| 14.68 | 1 |
| | |

1. Is this independent or paired data? Why?
    - This data is independent because the data isn't linked
2. Which *t*-test would be appropriate to compare the two groups?
    - Independent ttest
3. Check the assumptions for the chosen test. Paste your output and discuss your conclusions here.
    - It does not appear to be normal. So we will use the Wilcoxon Ranked Sum Test




4. Formally test (using a *t*-test or a nonparametric test as determined in the last part) to determine whether the staffing shortages are different. Use $\alpha$=0.05. Include your hypotheses, SAS output, SAS Code, and conclusions.
    - H0: The staffing shortages between the states are the same
    - H1: The staffing shortages between the states are different
    - We reject H0 if p < 0.05
    - Our p-val is less than 0.0001
    - Since our p-val is less than 0.0001, the it is less than 0.05, so we will reject the H0. So, there is enough evidence to suggest that the staffing shortages are different between the states
    - **Code:**
    - **data five;**
    - **input shortage state @@;**
    - **cards;**
    - **6.84 0**
    - **7 0**
    - **6.97 0**
    - **7.11 0**
    - **7.53 0**
    - **10.36 1**
    - **10.65 1**
    - **10.65 1**
    - **10.69 1**
    - **10.62 1**
    - **10.84 1**
    - **11.78 1**
    - **11.71 1**
    - **11.65 1**

- **11.6 1**
- **11.64 1**
- **0.69 0**
- **0.78 0**
- **0.99 0**
- **1.05 0**
- **1.21 0**
- **1.37 0**
- **1.84 0**
- **4.19 0**
- **4.71 0**
- **4.92 0**
- **5.74 0**
- **6.17 0**
- **6.75 0**
- **7.91 0**
- **8.26 0**
- **8.53 0**
- **8.87 0**
- **8.89 0**
- **9.52 1**
- **10.02 1**
- **0.19 0**
- **0.39 0**
- **0.58 0**
- **0.68 0**
- **0.97 0**
- **1.66 0**
- **3.13 0**
- **4.37 0**
- **5.01 0**
- **5.74 0**
- **6.3 0**
- **6.86 0**
- **6.95 0**
- **7.42 0**
- **7.41 0**
- **8.34 0**
- **8.52 0**
- **9.44 1**
- **10.35 1**
- **5.43 0**
- **8.56 0**
- **9.09 1**
- **10.18 1**
- **1.8 0**

- **2.19 0**
- **2.38 0**
- **2.76 0**
- **3.34 0**
- **4.31 0**
- **5.66 0**
- **6.61 0**
- **8.34 0**
- **9.1 1**
- **9.85 1**
- **10.99 1**
- **12.5 1**
- **13.04 1**
- **13.96 1**
- **14.68 1**
- **;**
- **run;**
- **proc print data=five;**
- **run;**
- **proc ttest data=five h0=0;**
- **class state;**
- **var shortage;**
- **run;**
- **proc npar1way data=five wilcoxon;**
- **class state;**
- **var shortage;**
- **run;**

| Wilcoxon Two-Sample Test | | | | | |
|---|---|---|---|---|---|
| | | | | t Approximation | |
| Statistic | Z | Pr > Z | Pr > \|Z\| | Pr > Z | Pr > \|Z\| |
| 1524.000 | 6.9454 | <.0001 | <.0001 | <.0001 | <.0001 |
| Z includes a continuity correction of 0.5. | | | | | |

5. If you conducted a *t*-test, construct a 95% confidence interval for the difference between the means.
   - Did not conduct a ttest – conducted Wilcoxon

---

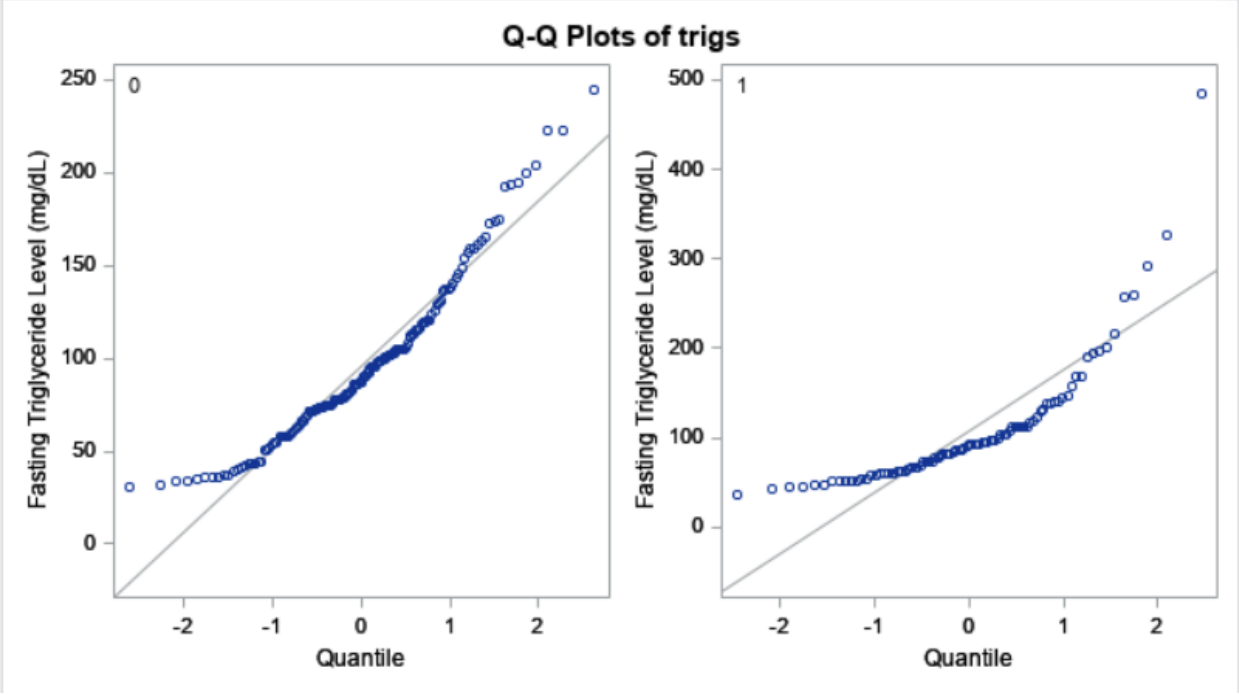
Part 2: Using a .sas7bdat file for data.

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

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

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

Use this data set to determine whether there is a difference in triglycerides (trigs) between males and females. Use $\alpha$=0.01.

6. Is this independent or paired data? Why?
   - This data is independent because it is not linked.
7. Which *t*-test would be appropriate to compare the two groups?
   - Independent ttest
8. Check the assumptions for the chosen test. Paste your output and discuss your conclusions here.
   - It is not a normal distribution since it isn't a linear trend. So we will run a Wilcoxon Rank Sum test
   - 

9. Formally test (using a *t*-test or a nonparametric test as determined in the last part) to determine whether the triglycerides are different between men and women. Use $\alpha$=0.01. Include your hypotheses, SAS output, SAS Code, and conclusions.
   - H0: the triglycerides in men and women are the same
   - H1: the triglycerides in men and women are different
   - We reject the H0 if the p-val is less than 0.01
   - Our p-val is 0.5982

- Since our p-val is 0.5982 and 0.5982 is greater than 0.01, we fail to reject the H0. There is not enough evidence to suggest that the triglycerides are different in men and women.

| Wilcoxon Two-Sample Test | | | | | |
|---|---|---|---|---|---|
| | | | | t Approximation | |
| Statistic | Z | Pr > Z | Pr > \|Z\| | Pr > Z | Pr > \|Z\| |
| 11006.00 | 0.5270 | 0.2991 | 0.5982 | 0.2994 | 0.5987 |
| Z includes a continuity correction of 0.5. | | | | | |

10. If you conducted a *t*-test, construct a 99% confidence interval for the difference between the means.
    - Did not conduct ttest – conducted Wilcoxon test

Paste your SAS code here:

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

data ten;
set in.p099;
run;

proc print data=ten;
run;

proc ttest data=ten h0=0;
class male;
var trigs;
run;

proc npar1way data=ten wilcoxon;
class male;
var trigs;
run;
```

---

Graduate students only:

Use a proc print statement to explore the variables in your p099 data set. Determine a hypothesis that you would like to answer that will require use of an independent t-test (or Wilcoxon test). Write your hypotheses and perform all evaluations of assumptions and conclusions. Include your SAS code and output.