```
/* Question 1*/

data infantexposure;
length group$ 9;
infile "/home/u60672671/InfantExposure.txt" dlm = "09"x;
input group$ level;
if group = exposed then level2 = (level - 25);
if group NE exposed then level2 = level;
run;


*
                                    1B:
A) Stat the null and alternative hypothesis:

    Null hypothesis: The average level is 25 ng per mg
    Alternative hypothesis: The average level is not equal to 25 nh per mg

B) Indicate whether you should use pooled or unpooled variance:

    Pr > F is greater than 0.03. Therefore, the difference in variances is significant
    so we use pooled variance.

C) State the value of the test statistic:
T   he test statistic is equal to 2.28.

D) State your p=value and conclusion at alpha = 0.03:

    The p-value is equal to 0.0405.  Since this is greater than 0.03,
    there is not evidence to indicate that the average level of the population is equal to 25 ng per mg.


E) State the one-sided 97% confidence interval and interpret it

    The one-sided 97% confidence interval is (20.7820,148.6). This means
    we are 97% sure that the mean difference in ng per mg between the exposed
    and unexposed babies falls between 20.7820 and 148.6

F) Analyze whether any of the assumptions of the t-test have been violated
    While the exposed group is symmetrical and the data falls in line with what we would expect
    from a normal distribution. the unexposed group, however, has data which is less symmetrical,
    and does not line up with the values we would expect from a normal distribution.
    Therefore, the assumption that the data is normally distributed may be violated.

                                    2C:
a) Write the null and alternative hypothesis for the Wilcoxon Rank Sum test:

    Null Hypothesis: the distribution of the data is the same between the groups
    Alternative Hypothesis: The distribution of the data is different between the groups

b) State the value of the test statistic
    The test statistic is equal to 33.0

c) State your p-value (Are you using z approximation or t approximation?) and conclusion
at alpha = 0.03.

    We are using a t-approximation since the sample size is small (less than 10). the p-value is 0.0208.
    As this is less than 0.03, there is sufficient evidence to suggest that the distribution of data is not the same
    betweeen both groups.

                                    2D:

    It is better to use the Wilcoxon Rank Sum Test, because the sample size is small, and it does not seem
    that a normal distribution can be assumed. In this problem, it is importnt to choose the correct test
    because the assumptions are violated for the t-test, which would mean that
    you cannot rely on the conclusions drawn from that test.

;

Proc Ttest data = infantexposure h0 = 25 alpha = 0.03;
class group;
var level;
run;

Proc npar1way wilcoxon data = infantexposure;
class group;
var level2;
run;


/*Question #2 */

/* a) Write the appropriate one tailed null and alternative hypothesis
to test if the arsenic levels exceed 10 ppb

    Null Hypothesis: The arsenic levels are greater than or equal to 10 ppb.
    Alterntive Hypothesis: The arsenic levels are less than 10 ppb.

    b) See code below

    c) Provide the value of the test statistic and the p-value. State your conclusion
        at the 1% significance level

        The p-value is 0.001/2 or 0.0005. The t-value is equal to 103.04.
        With these results, there is enough evidence to suggest that the alternative hypothesis is true.
        (0.0005 < 0.01)
        In other words, we can say that the arsenic levels are less than 10 ppb.

    d) In the context of your statistical conclusion, should the citizens continue drinking the water or refrain from drinking
    the water? Explain.

        The citizens can continue drinking the water. We have seen that the level of arsenic is incredibly likely
        to  be below 10 ppb, therefore, it is safe to drink.

    e) Provide the 99% confidence interval and interpret it.

        The 99% confidence interval is (9.5478, 10.1413). This means that we are 99% certain that the arsenic levels
        in the water fall between 9.5478 ppb and 10.1413 ppb.
*/

Data arseniclevels;
input arseniclvl;
Datalines;
9.722
10.162
9.976
9.787
9.474
10.113
10.157
9.556
9.667
9.809
```

The TTEST Procedure

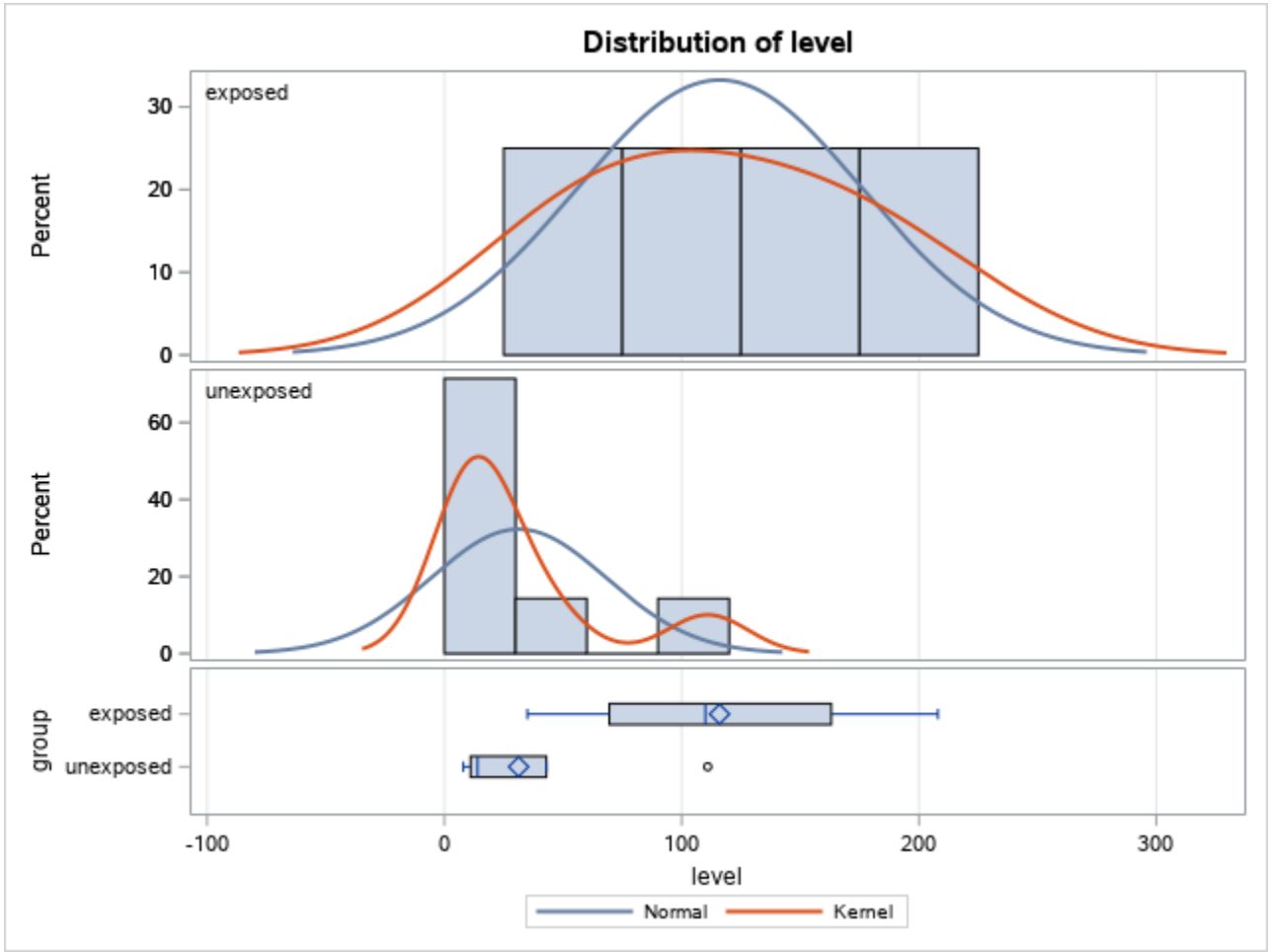
Variable: level

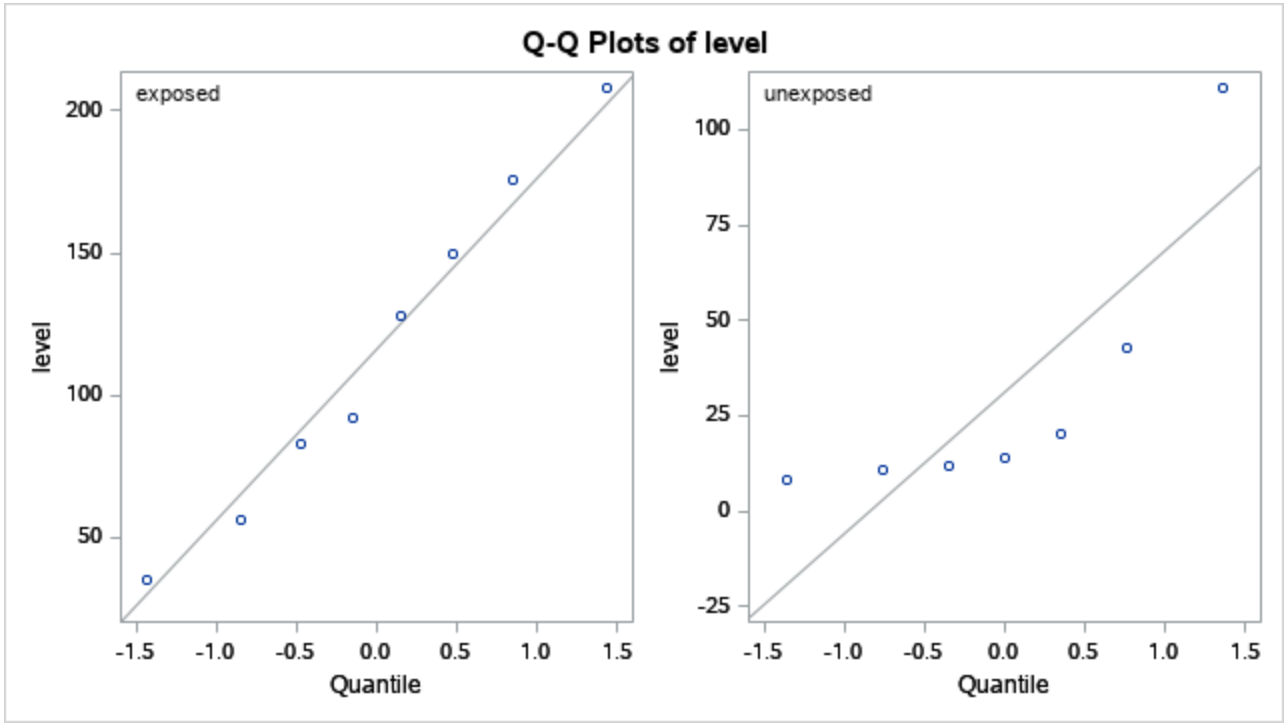
group	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
exposed		8	116.0	59.9881	21.2090	35.0000	208.0
unexposed		7	31.2857	37.0662	14.0097	8.0000	111.0
Diff (1-2)	Pooled		84.7143	50.7129	26.2464		
Diff (1-2)	Satterthwaite		84.7143		25.4184		

group	Method	Mean	97% CL Mean		Std Dev	97% CL Std Dev	
exposed		116.0	58.4266	173.6	59.9881	38.0504	133.3
unexposed		31.2857	-8.3468	70.9182	37.0662	22.8579	90.0773
Diff (1-2)	Pooled	84.7143	20.7820	148.6	50.7129	35.5845	86.4295
Diff (1-2)	Satterthwaite	84.7143	22.0385	147.4			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	13	2.28	0.0405
Satterthwaite	Unequal	11.817	2.35	0.0371

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	7	6	2.62	0.2613



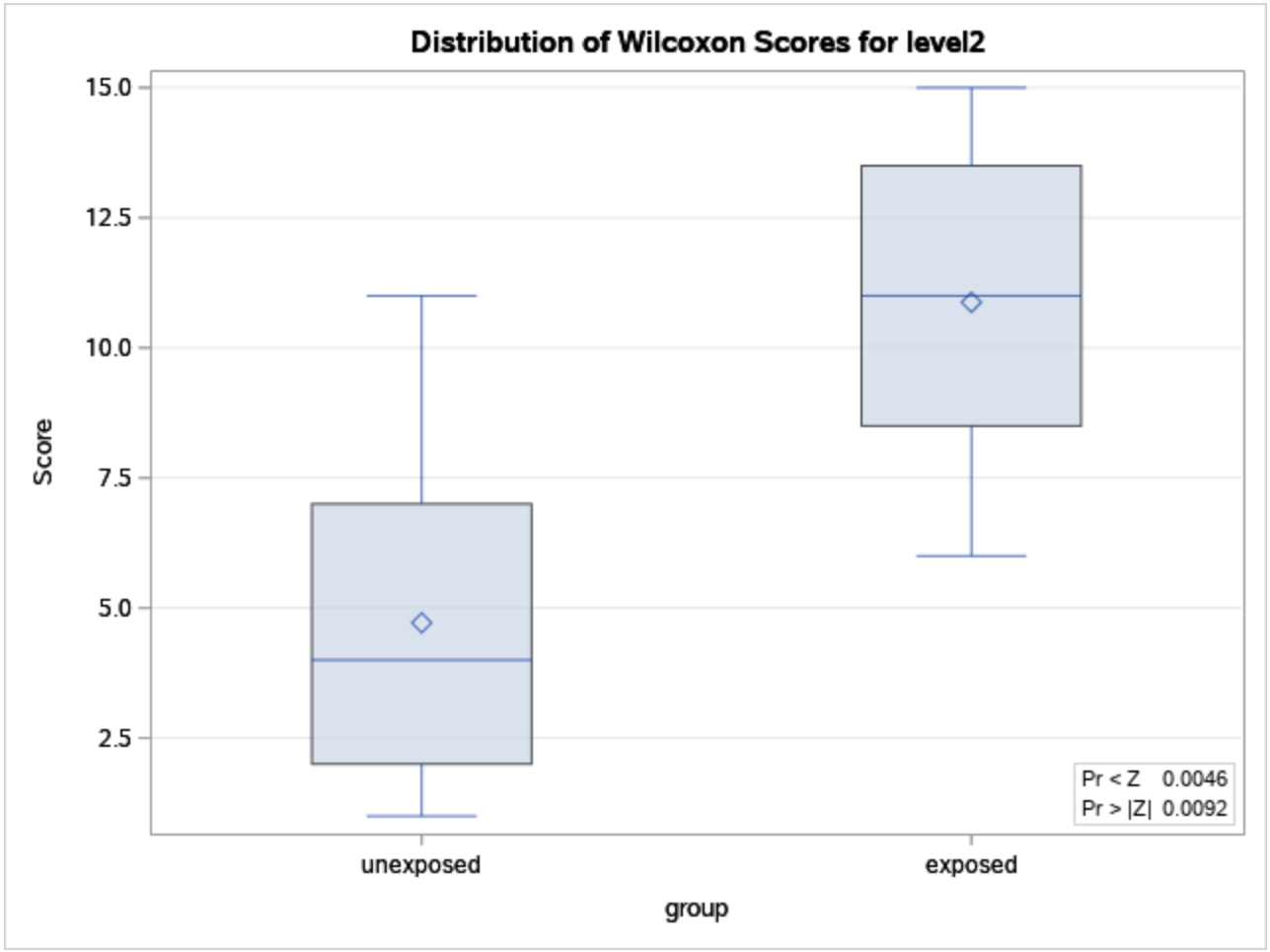


The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable level2 Classified by Variable group					
group	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
unexposed	7	33.0	56.0	8.640988	4.714286
exposed	8	87.0	64.0	8.640988	10.875000

Wilcoxon Two-Sample Test					
				t Approximation	
Statistic	Z	Pr < Z	Pr > |Z|	Pr < Z	Pr > |Z|
33.0000	-2.6039	0.0046	0.0092	0.0104	0.0208
Z includes a continuity correction of 0.5.					

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
7.0848	1	0.0078



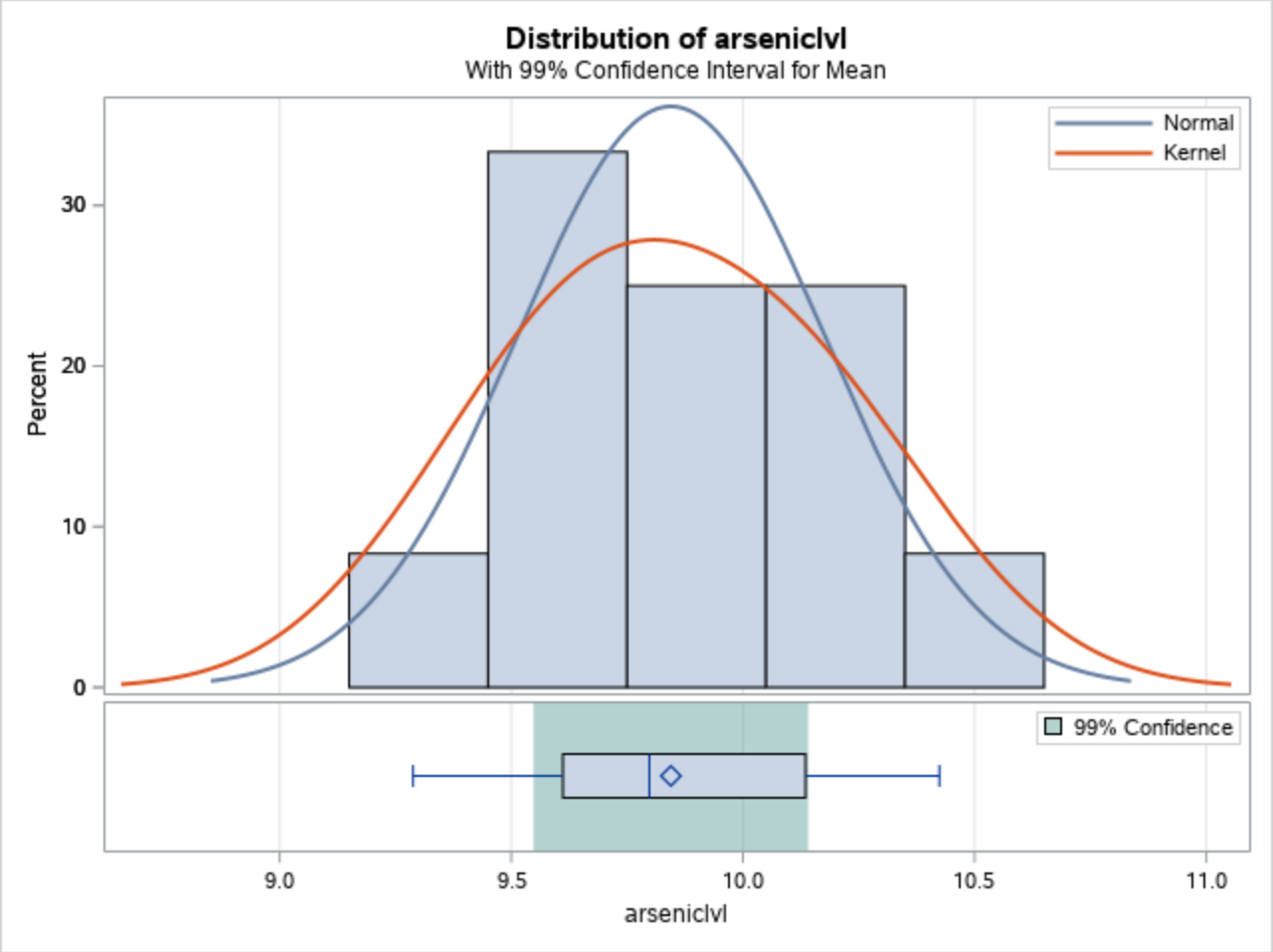
The TTEST Procedure

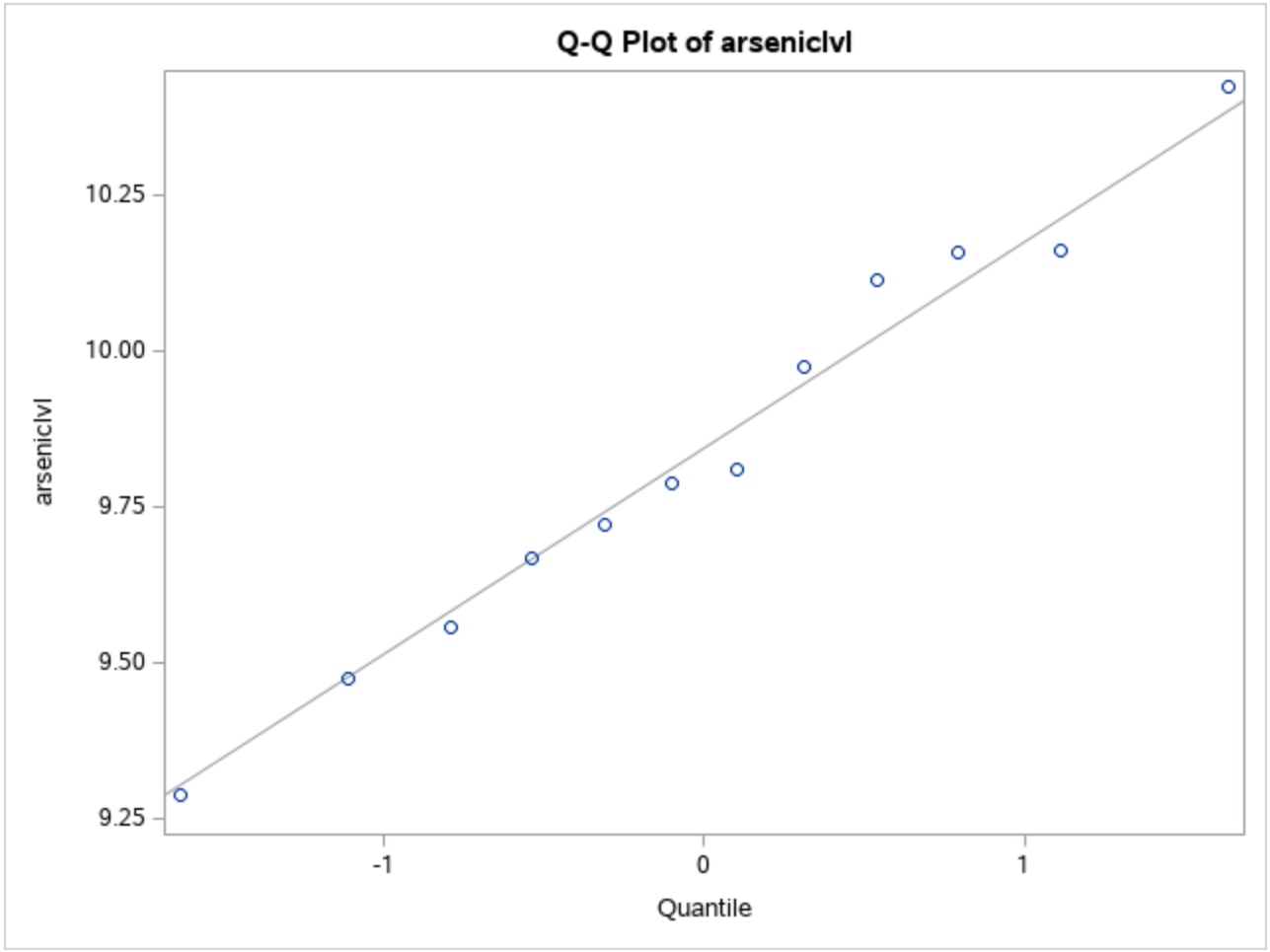
Variable: arseniclvl

N	Mean	Std Dev	Std Err	Minimum	Maximum
12	9.8446	0.3310	0.0955	9.2880	10.4240

Mean	99% CL Mean		Std Dev	99% CL Std Dev	
9.8446	9.5478	10.1413	0.3310	0.2122	0.6803

DF	t Value	Pr > |t|
11	103.04	<.0001


Distribution of arseniclvl
With 99% Confidence Interval for Mean
Normal
Kernel
Percent
30
20
10
0
99% Confidence
9.0
9.5
10.0
10.5
11.0
arseniclvl



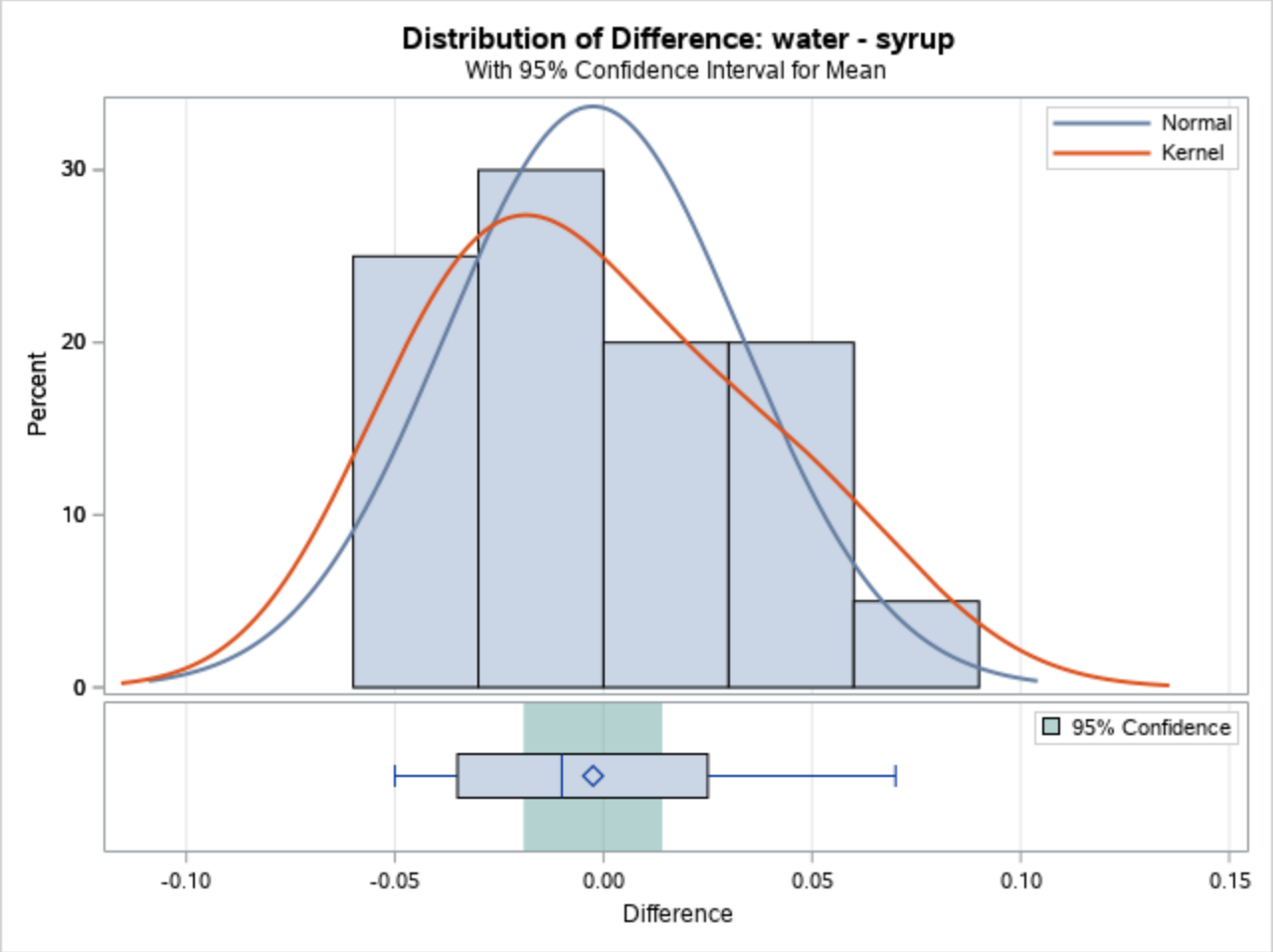
The TTEST Procedure

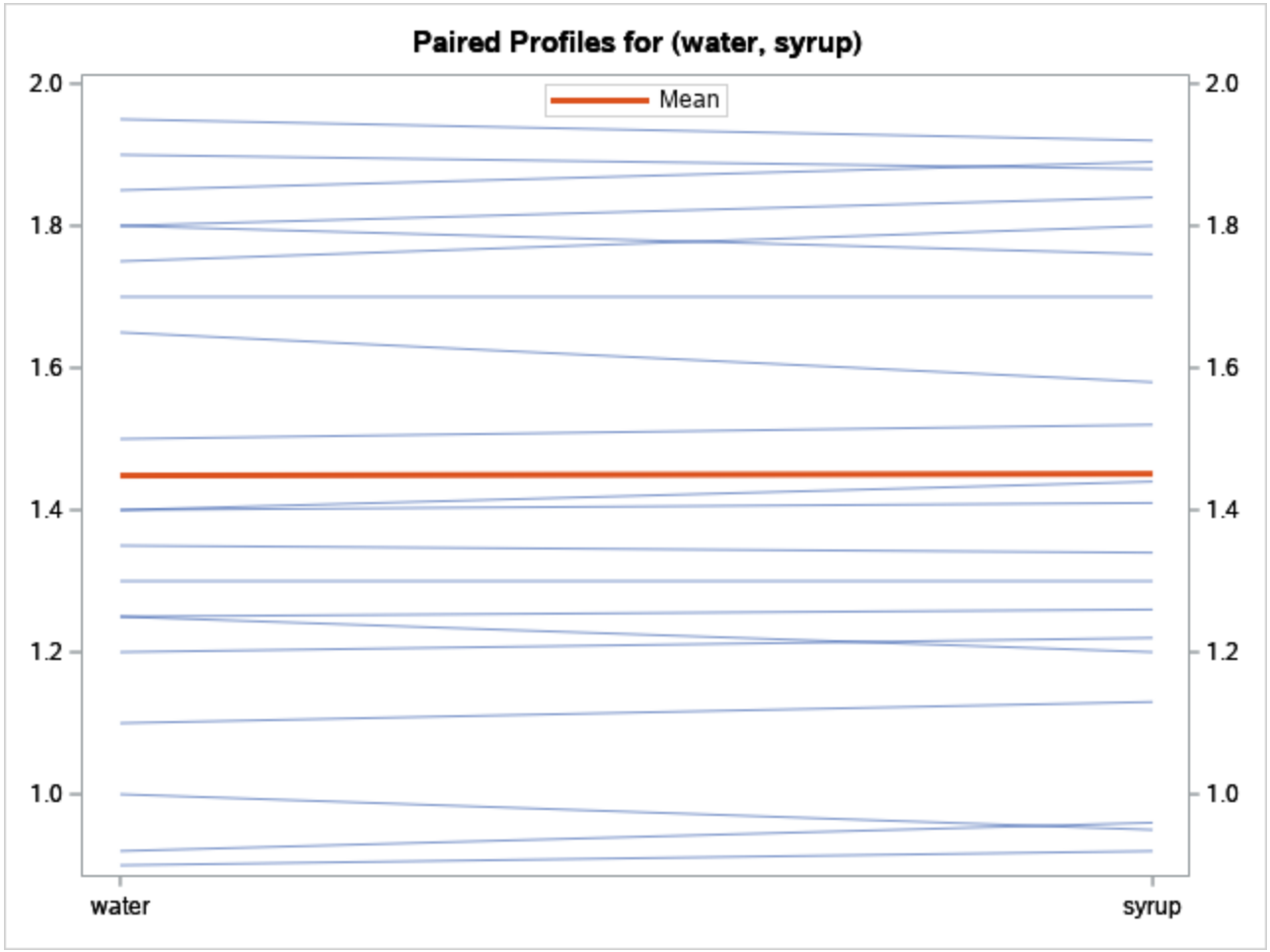
Difference: water - syrup

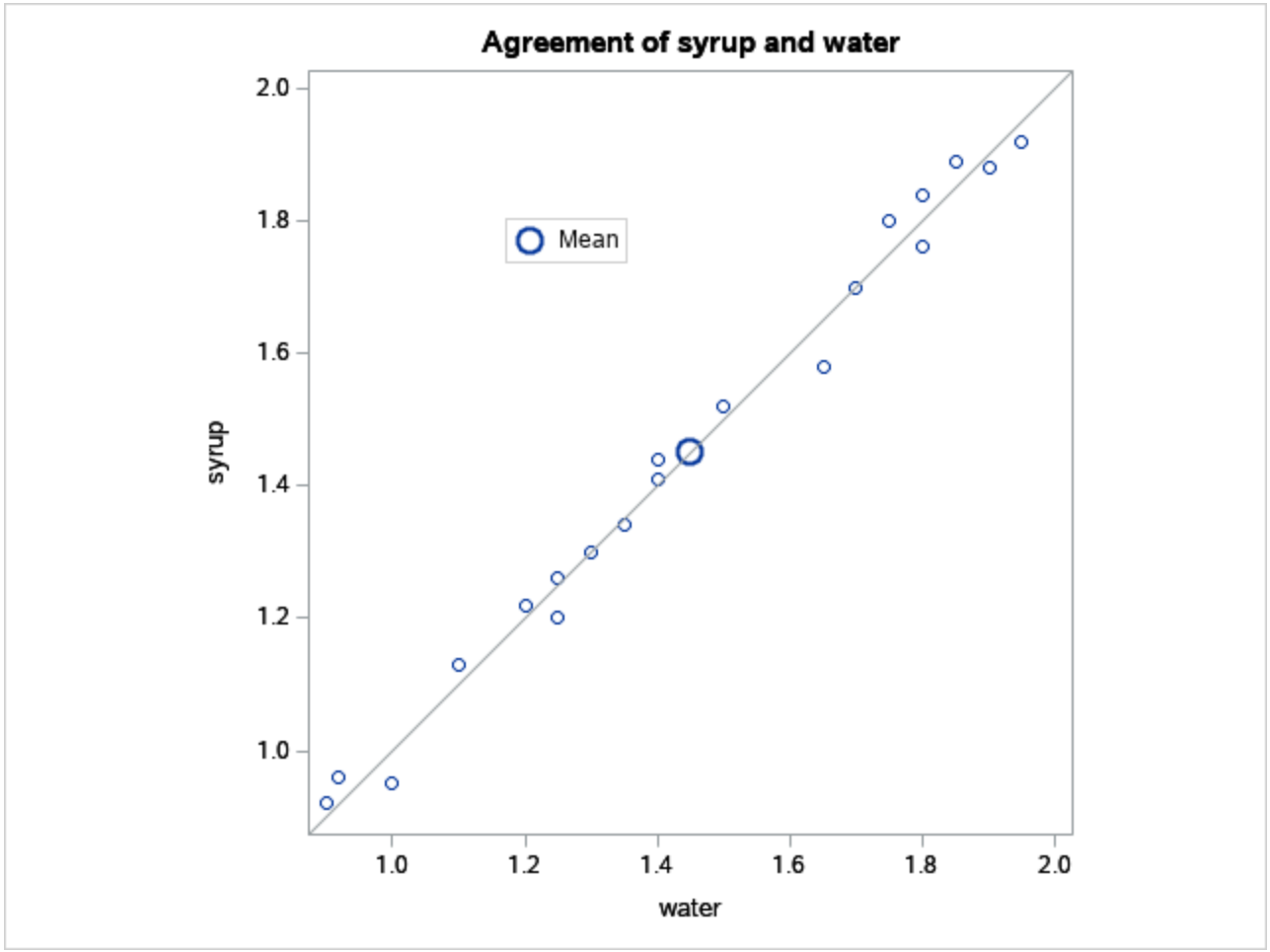
N	Mean	Std Dev	Std Err	Minimum	Maximum
20	-0.00250	0.0355	0.00794	-0.0500	0.0700

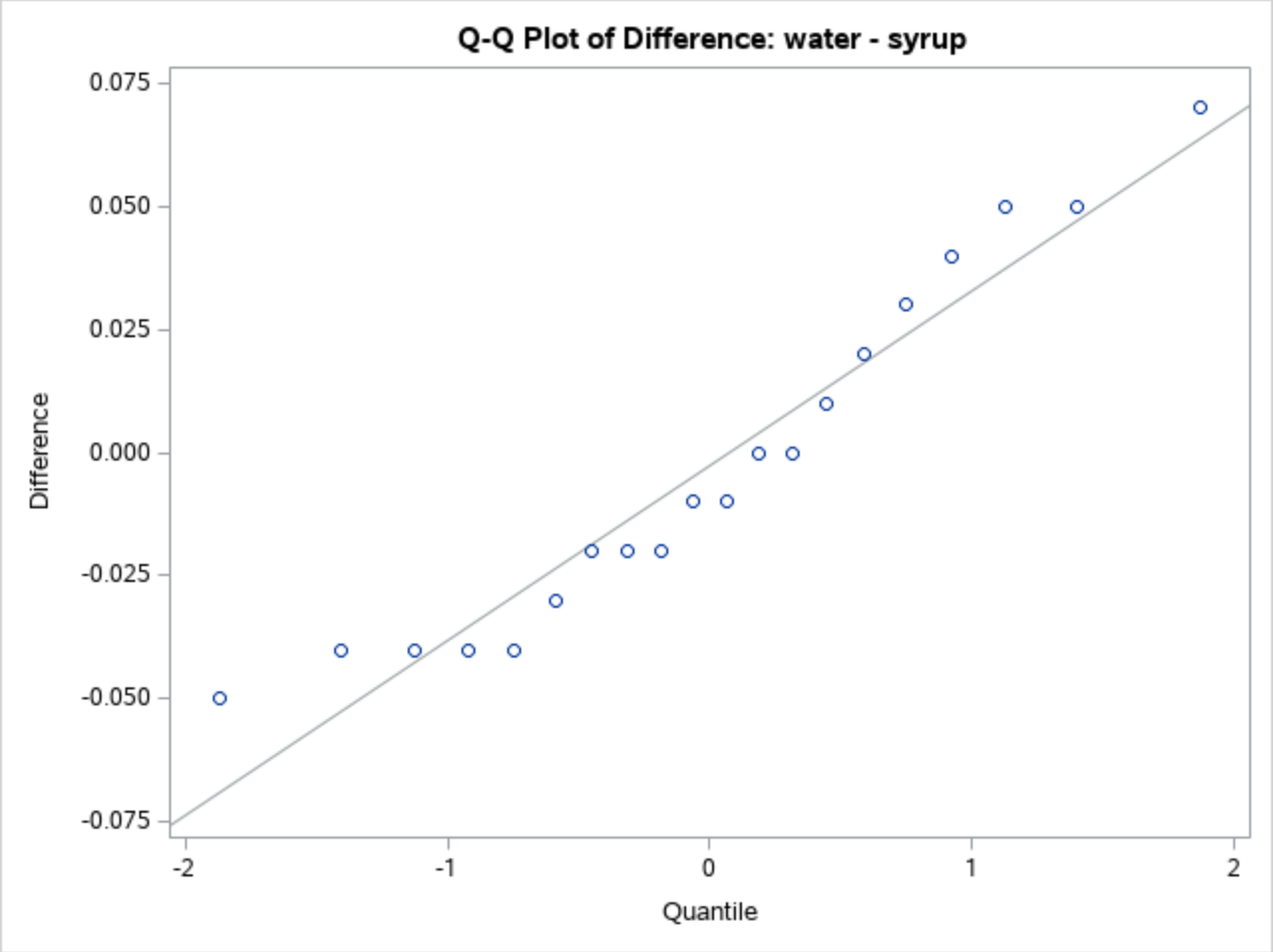
Mean	95% CL Mean		Std Dev	95% CL Std Dev	
-0.00250	-0.0191	0.0141	0.0355	0.0270	0.0519

DF	t Value	Pr > |t|
19	-0.31	0.7564


Distribution of Difference: water - syrup
With 95% Confidence Interval for Mean
Normal
Kernel
Percent
30
20
10
0
95% Confidence
-0.10
-0.05
0.00
0.05
0.10
0.15
Difference


Paired Profiles for (water, syrup)
Mean
2.0
1.8
1.6
1.4
1.2
1.0
water
syrup


Agreement of syrup and water
Mean
syrup
water
2.0
1.8
1.6
1.4
1.2
1.0
1.0
1.2
1.4
1.6
1.8
2.0



The UNIVARIATE Procedure
Variable: difference

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	-0.31474	**Pr > |t|**	0.7564
Sign	**M**	-2	**Pr >= |M|**	0.4807
Signed Rank	**S**	-9	**Pr >= |S|**	0.7098