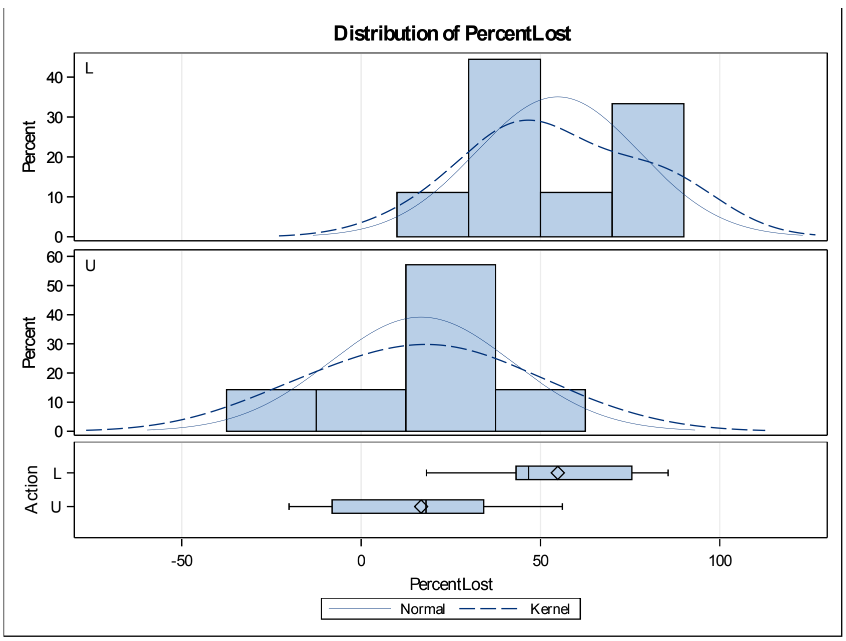
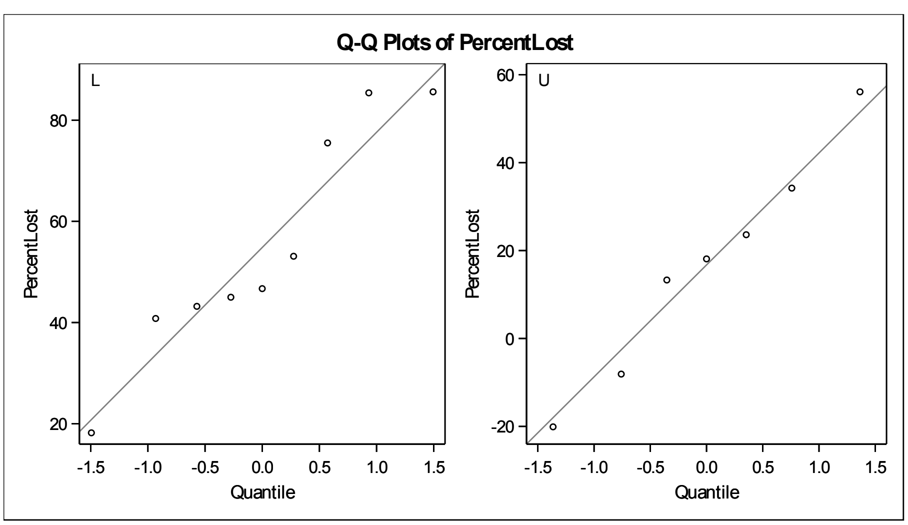
<u>Question 2)</u>

Assumptions

Assuming one knows little enough about the **normalcy** and **variance** of the population distribution to warrant using a non-parametric test like the Rank-Sum, a potential outlier in the logged data (value = 18.2) may also make Rank-Sum appropriate in this case. An initial t-test—boxplots and qqplots for which are produced below (Figure 1)—was run to preliminarily determine normalcy and variance. Furthermore, the assumption of observation **independence** should be adequately satisfied by the "randomly located transect patterns".

Figure 1:





Analysis

To test for the impact of logging on fire-affected plots, a Wilcoxon Rank-Sum test was run to test the ***null-hypothesis (H_0)*** that: there is **no difference** between logged and unlogged plots **in the mean percentage of seedlings lost**:

$$\mu_L - \mu_U = 0$$

The ***alternative-hypothesis (H_A)*** would be that: the mean percentage of seedlings lost is **greater** in logged plots than in unlogged plots:

$$\mu_L - \mu_U > 0$$

Assuming the null-hypothesis, statistical software determined that, for this test, **S_{stat} = 36**. The corresponding Z-scores (**including continuity correction**) and p-values were as follows:

Z = -2.4346 → p-value (one-sided) = **.0075**

On this basis, one would **reject the null-hypothesis *(H_0)***, determining that: there is statistically significant evidence to suggest that the **mean percentage of seedlings lost is greater in logged plots** than in unlogged plots affected by fire.

The associated **99% confidence interval,** according to **Monte Carlo estimates** for the Exact Test, is as follows: **[.0043, .0083]**.

Given the evidence for this observational study—where random samples are selected from existing distinct populations of logged and unlogged fire-affected plots—inferences regarding the difference in percentage of seedlings lost **can be drawn to the populations** being examined. However, as assignment to groups (logged and unlogged) was not randomized, inferences regarding **causal relations cannot be drawn.**

Tables and boxplots for the results are produced below (Figure 2):

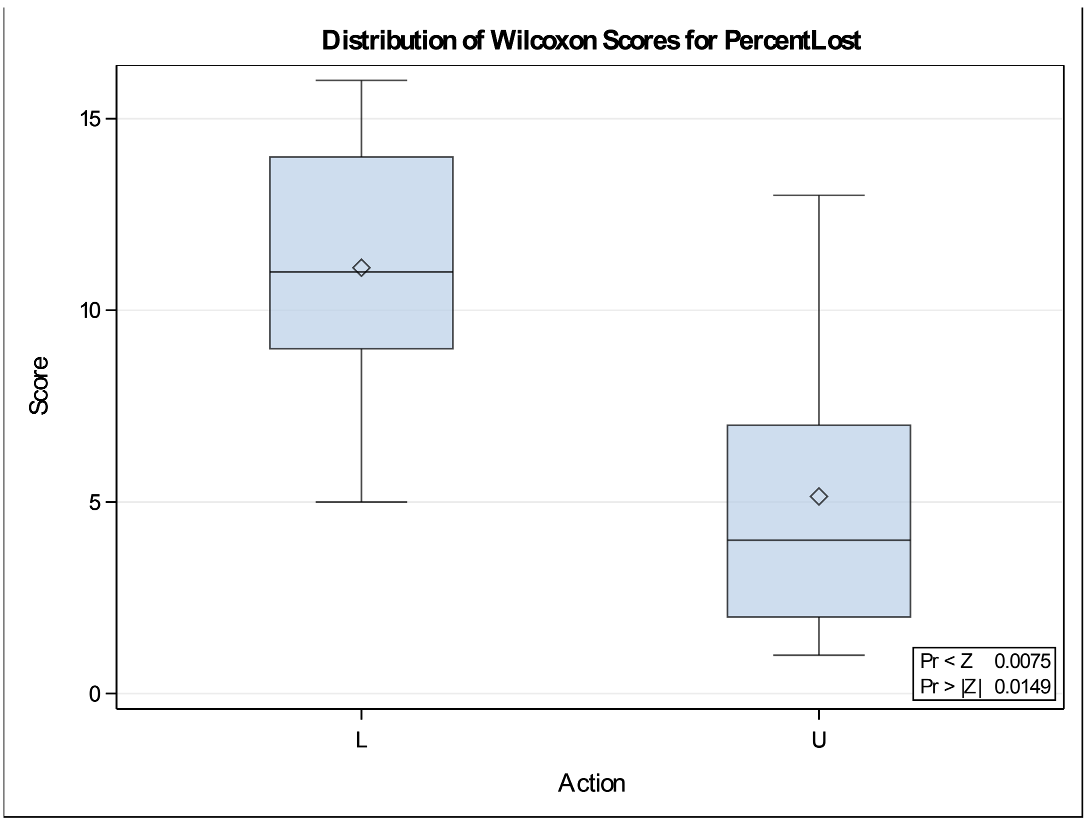
Figure 2:

Wilcoxon Scores (Rank Sums) for Variable PercentLost Classified by Variable Action

Action	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
L	9	100.0	76.50	9.447222	11.111111
U	7	36.0	59.50	9.447222	5.142857

Wilcoxon Two-Sample Test	
Statistic (S)	36.0000
Normal Approximation	
Z	-2.4346
One-Sided Pr < Z	0.0075
Two-Sided Pr > |Z|	0.0149
t Approximation	
One-Sided Pr < Z	0.0139
Two-Sided Pr > |Z|	0.0279
Z includes a continuity correction of 0.5.	

Monte Carlo Estimates for the Exact Test	
One-Sided Pr <= S	
Estimate	0.0063
99% Lower Conf Limit	0.0043
99% Upper Conf Limit	0.0083
Two-Sided Pr >= |S - Mean|	
Estimate	0.0120
99% Lower Conf Limit	0.0092
99% Upper Conf Limit	0.0148
Number of Samples	10000
Initial Seed	1234



Question 3)

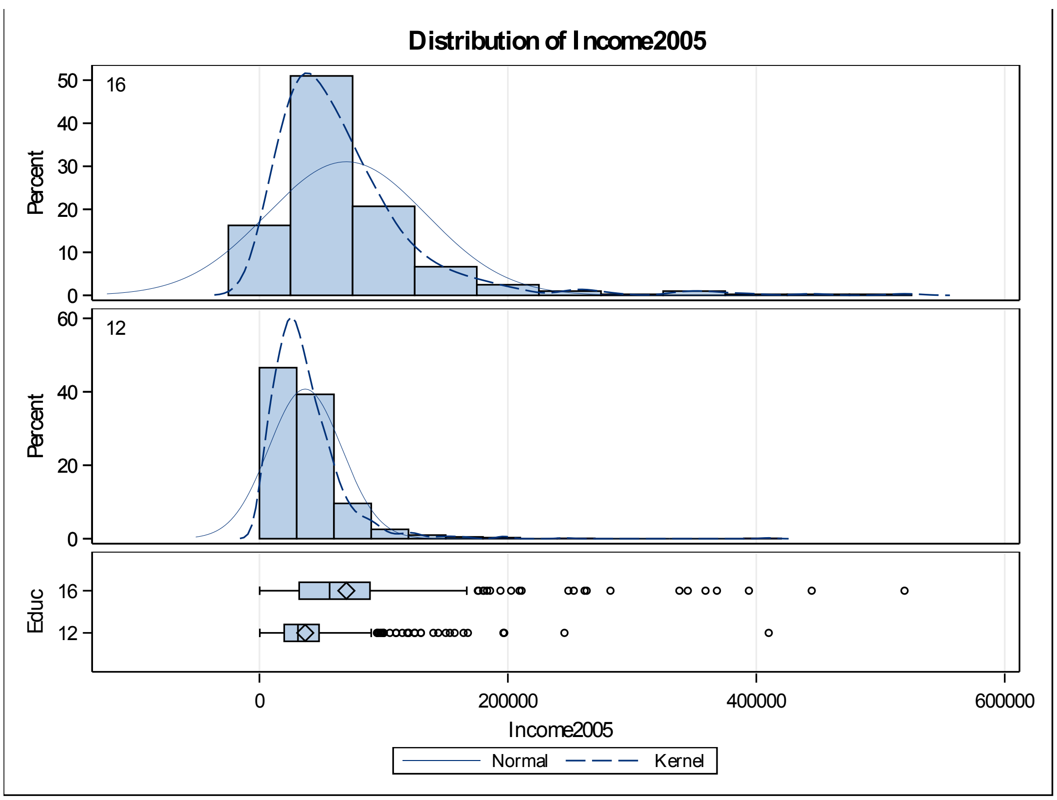
Parts A + B)

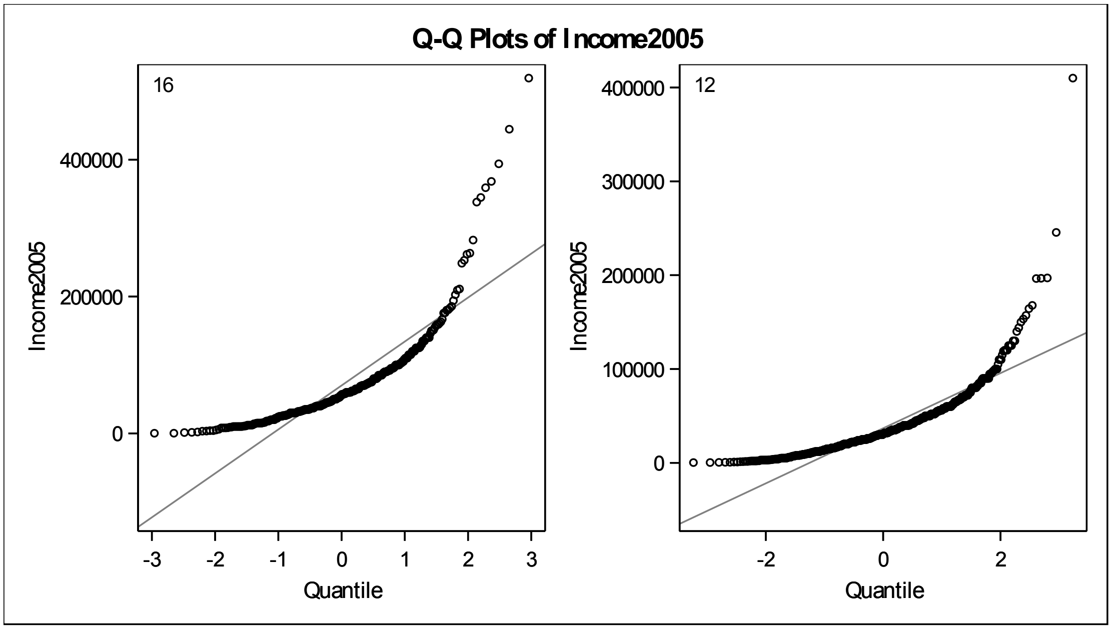
Assumptions

An initial t-test—boxplots and qqplots for which are produced below (Figure 3)—was run to preliminarily determine normalcy and variance. The data are clearly **right-skewed** and have **unequal variances**; of the two groups, the one with the larger mean also has the larger variance. Moreover, judging from the description of the sampling procedure for the study's design, the assumption of observation **independence** should be adequately satisfied. Although these conditions are suited for a logarithmic transformation, knowledge of the CLT tells us that the study's adequate sample size (n = 1426) should ensure reasonable results, compensating somewhat for skewedness and variance. Figures for assumption testing (Figure 3) are produced below:

Figure 3:

Educ	N	Mean	Std Dev	Std Err	Minimum	Maximum
16	406	69997.0	64256.8	3189.0	200.0	519340
12	1020	36864.9	29369.7	919.6	300.0	410008
Diff (1-2)		33132.1	42326.9	2483.8		





Analysis

To test for the impact of education on income level, a one-sided Welch's two-sample t-test was run to test the ***null-hypothesis (H₀)*** that: there is **no difference** in mean income level between those with 16 and those with 12 years of education:

$$\mu_{16} - \mu_{12} = 0$$

The ***alternative-hypothesis (H_A)*** would be that: mean income level is **greater** for those with 16 years of education than it is for those with 12 years:

$$\mu_{16} - \mu_{U12} > 0$$

Assuming the null-hypothesis and unequal variance (**Satterthwaite**), with a sample size of *n* = 1426, with *df* = 1424, and $\propto =$ .05, the critical value would be: $\mathbf{t_{crit}} = \mathbf{1.961.}$

Statistical software determined that, for this test $\mathbf{t_{stat}} \approx \mathbf{9.98}$.

Correspondingly, the **p-value (one-sided)** = **.0001**.

On this basis, one would **reject the null-hypothesis *(H_0)***, determining that: there is statistically significant evidence to suggest that the mean income for those with 16 years of education is **greater** than it is for those with 12 years.

The associated **95% confidence interval** for the difference in means, providing **two-sides** for intuitive clarity, is as follows: [**26610.40**, **39653.80]**.

Part C)

Given the design description and the evidence for this observational study—in which neither sampling nor allocation is randomized, relying on a subset of those that voluntarily responded to a survey—inferences **can only be drawn to the population of 41-49 year old survey respondents**, though this limited scope of inference may still be informative. However, inferences regarding **causal relations cannot be drawn.**

Further tables for the results (Figure 4) are produced below:

Figure 4:

Educ	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
16		69997.0	63727.9	76266.1	64256.8	60120.1	69009.5
12		36864.9	35060.4	38669.4	29369.7	28148.2	30702.9
Diff (1-2)	Pooled	33132.1	28259.8	38004.3	42326.9	40828.0	43940.9
Diff (1-2)	Satterthwaite	33132.1	26610.4	39653.8			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	1424	13.34	<.0001
Satterthwaite	Unequal	473.85	9.98	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	405	1019	4.79	<.0001

Part D)

Comparing these results with the numbers, tables, and figures produced by the log-transformed analysis prepared for HW3, it is arguable that the **log-transformed analysis is more appropriate**. Although the assumption of independence holds for both kinds of tests, one should note that the data are very right-skewed, and the variance is higher for the group with the higher mean. As the **median** is more resistant to such skewedness than the **mean**, a log transformation—which provides inference on the former parameter—is especially suitable for this kind of data.

Question 4)

Part A)

Rank Sum Figures

All	Group	Order	Rank			
18.8	N	1	1			
20	N	2	2			
20.1	N	3	3			
20.9	N	4	4.5			
20.9	N	5	4.5			
21.4	N	6	6			
22	T	7	7			
22.7	N	8	8			
22.9	N	9	9			
23	T	10	10			
24.5	T	11	11			
25.8	T	12	12			
30	T	13	13			
37.6	T	14	14			
38.5	T	15	15			
				T-Stat	82	
				n1	7	
				R-bar	8	
				Mean(T)	56	
				n2	8	
				S(R)	4.468141	
				SD(T)	8.633269	
				Cont. Correction	-0.5	
				Z-Stat	2.95369	
				P-val (1-sided)	(1-.99843) = .00157	
				Crit Val	1.771	
				df	13	
				alpha	0.05	

Conclusion

On the basis of the above evidence, one would **reject the null-hypothesis** that there is no difference in the mean metabolic expenditures (kcal/kg/day) of patients with and without trauma:

[**P-val** (one-sided) = .00157]

Part B)

Statistical software was used to verify the rank-sum and the Z-statistic above, producing the following tables (Figure 5):

Figure 5:

Wilcoxon Scores (Rank Sums) for Variable Measure Classified by Variable Trauma					
Trauma	**N**	**Sum of Scores**	**Expected Under H0**	**Std Dev Under H0**	**Mean Score**
N	8	38.0	64.0	8.633269	4.750000
T	7	82.0	56.0	8.633269	11.714286
Average scores were used for ties.					

Wilcoxon Two-Sample Test	
Statistic	82.0000
Normal Approximation	
Z	2.9537
One-Sided Pr > Z	0.0016
Two-Sided Pr > |Z|	0.0031
t Approximation	
One-Sided Pr > Z	0.0052
Two-Sided Pr > |Z|	0.0105
Z includes a continuity correction of 0.5.	

Kruskal-Wallis Test	
Chi-Square	9.0698
DF	1
Pr > Chi-Square	0.0026

Rounding variations aside, the statistical software, which uses a continuity correction of (.5) by default, found a **one-sided p-value (.0016)** roughly equal to that which was identified by hand **(.00157)**. Furthermore, the software verified the value of the **Z-statistic (82)**.

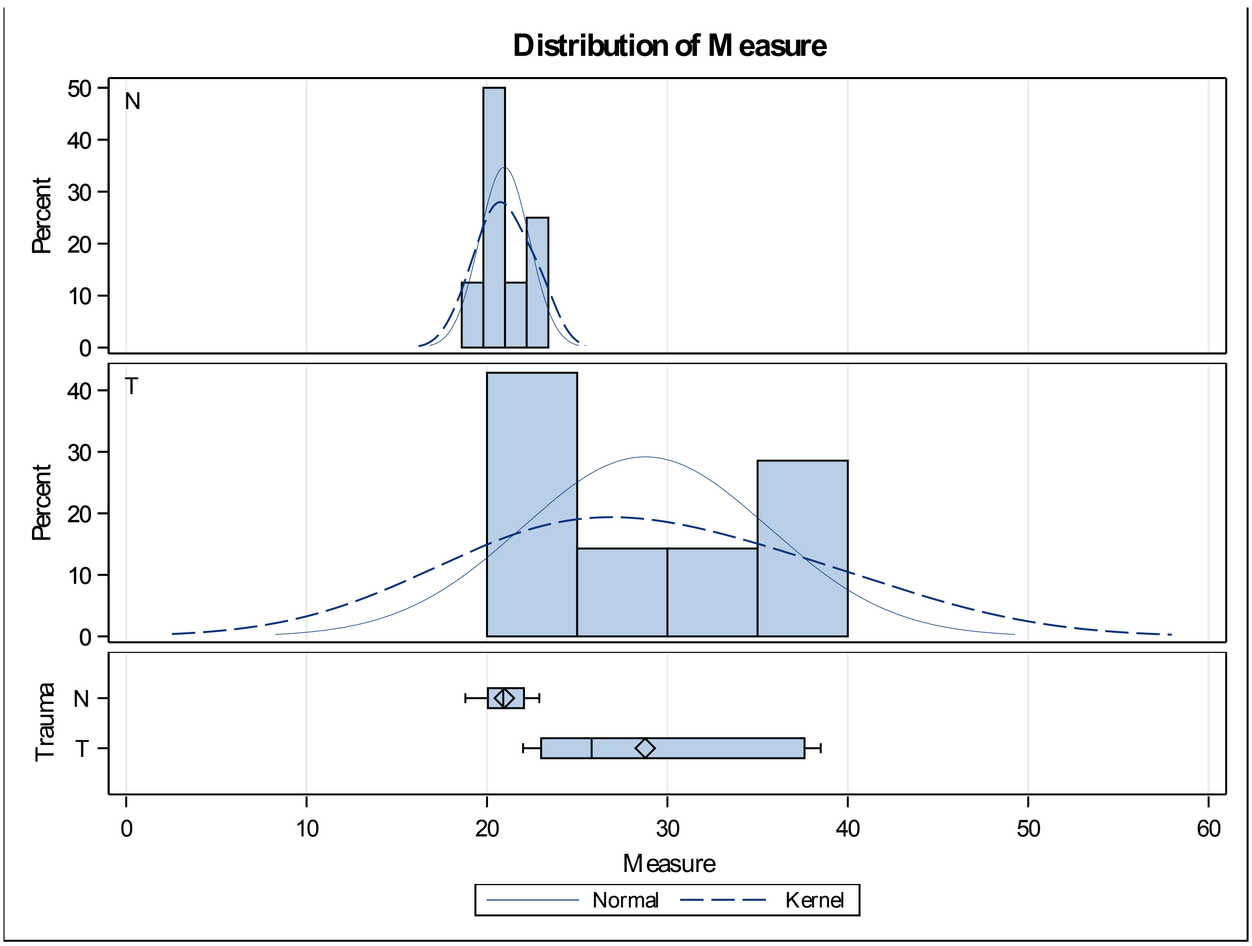
Parts C.i and C.ii)

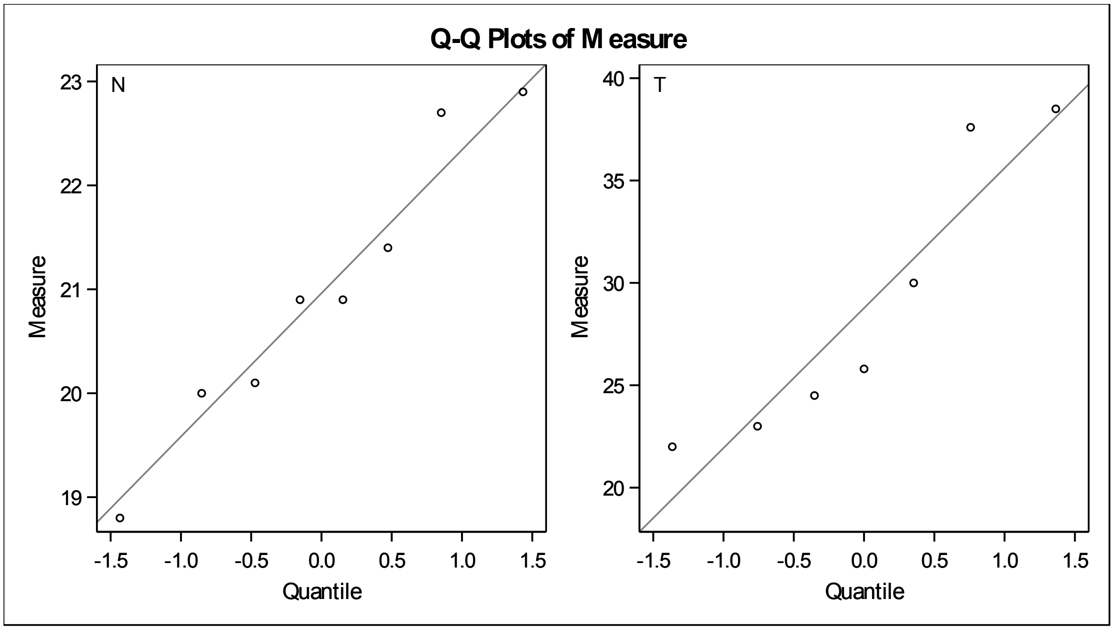
Assumptions

An initial t-test—histograms, box-plots and qq-plots for which are produced below (Figure 6)—was run to preliminarily determine normalcy and variance. Although the data appear **roughly normal**, they have substantially **unequal variances**; of the two groups, the one with the larger mean also has the larger variance. Moreover, judging from the sparse description of study's design, one can assume that observation **independence** is adequately satisfied. However, since the sample size is not especially large (n = 15), it is difficult to make assumptions regarding the normalcy of the distribution for the population being examined. For such reasons, a rank-sum test would be appropriate in this case.

Figure 6:

Trauma	N	Mean	Std Dev	Std Err	Minimum	Maximum
N	8	20.9625	1.3794	0.4877	18.8000	22.9000
T	7	28.7714	6.8354	2.5835	22.0000	38.5000
Diff (1-2)		-7.8089	4.7528	2.4598		





Parts C.iii. and C.iv)

Analysis

To test for the impact of trauma on metabolic expenditure, a Wilcoxon Rank-Sum test was run to test the ***null-hypothesis (H_0)*** that: there is **no difference** in mean metabolic expenditure between those with and without trauma:

$$\mu_T - \mu_N = 0$$

The ***alternative-hypothesis (H_A)*** would be that: mean metabolic expenditure is **greater** for those with trauma than for those without:

$$\mu_T - \mu_N > 0$$

Assuming the null-hypothesis, with a sample size of *n* = 15, with *df* = 13, and $\propto = .05$, the critical value would be **S_{crit} = 8.**

Statistical software determined that, for this test, **S_{stat}** $= \mathbf{82}$.

The corresponding Z-scores (**including continuity correction**) and p-values were as follows:

Z = 2.9537→ p-value (one-sided) = **.0016**

On this basis, one would **reject the null-hypothesis *(H_0)***, determining that: there is statistically significant evidence to suggest that **mean metabolic expenditure is greater for those with trauma** than it is for those without.

The associated **95% confidence interval for Location Shift (Trauma - Nontrauma),** according to the **Hodges-Lehmann Estimation** for the Exact Test, is as follows: **[1.900, 16.7000]**, with a midpoint of **(9.3000)**.

Given the evidence for this sparsely described observational study—in which neither sampling nor allocation appear to be randomized, and where collections of available units are selected from existing distinct populations—inferences regarding the difference in mean metabolic expenditure **can only be drawn to the particular population of trauma patients examined by the study**, though the results are still informative within that limited scope of inference. However, inferences regarding **causal relations cannot be drawn.**

A table for the Hodges-Lehman Estimation (Figure 7) is included below:

Figure 7:

Hodges-Lehmann Estimation				
Location Shift (T - N) 5.3000				
Type	**95% Confidence Limits**		**Interval Midpoint**	**Asymptotic Standard Error**
Asymptotic (Moses)	1.9000	16.7000	9.3000	3.7756
Exact	1.9000	16.7000	9.3000	

Question 5) **Part A)** *Signed Rank Figures*

Child	Before	After	Differenc	Ordered Mag	Order	Rank	Ranks (+)	Ranks (-)		
1	85	75	10	5	1	1.5	1.5			
2	70	50	20	5	2	1.5		1.5		
3	40	50	-10	10	3	4		4		
4	65	40	15	10	4	4	4			
5	80	20	60	10	5	4	4			
6	75	65	10	15	6	6	6			
7	55	40	5	20	7	7	7			
8	20	25	-5	40	8	8	8			
9	70	30	40	60	9	9	9			
						SR-Stat	39.5		(SR-Stat) - H0 Expected	17
						n1	9			
						Mean(S) (norm approx	22.5			
						SD(S) (norm approx.)	8.440972			
						Cont. Corr	-0.5			
						Z-Stat	1.954751			
						P-val (norm approx.)	(1-.9747) = .0253			
						Crit-Val	1.86			
						df	8			
						alpha	0.05			
						Poss. Assignments	512			
						# as extreme	13			

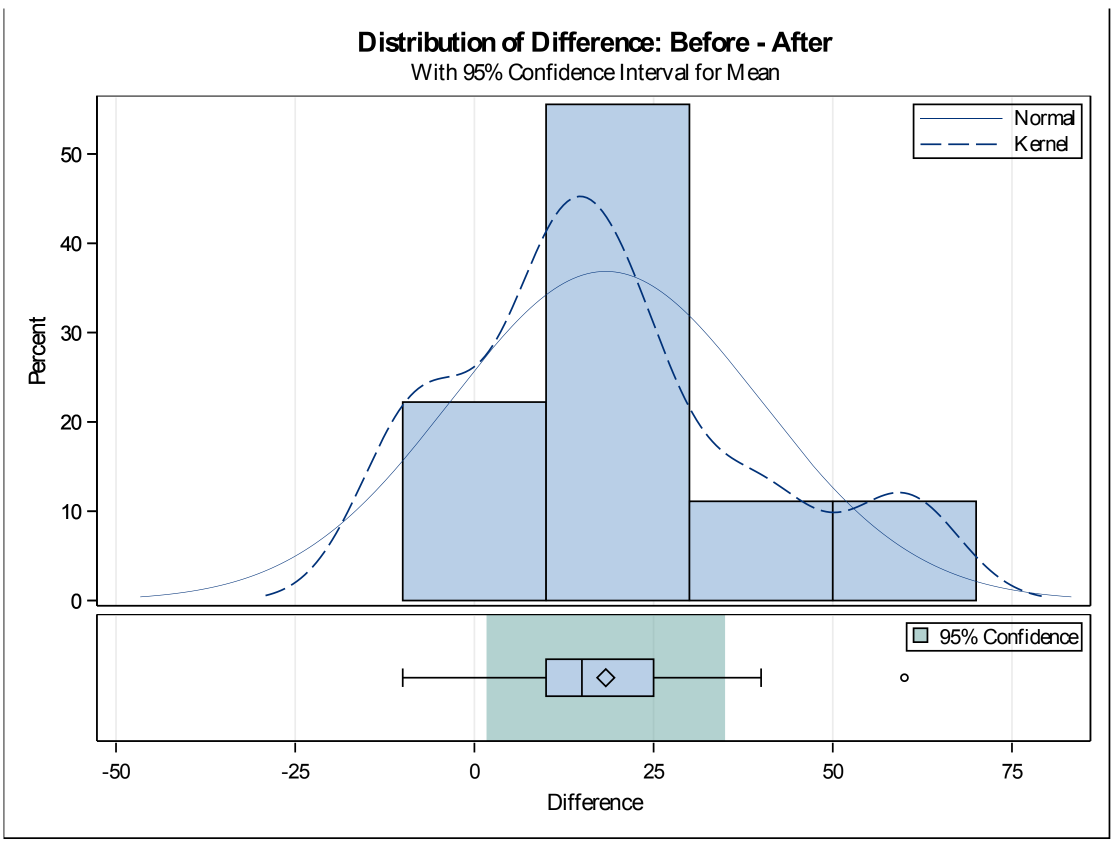
Part B)

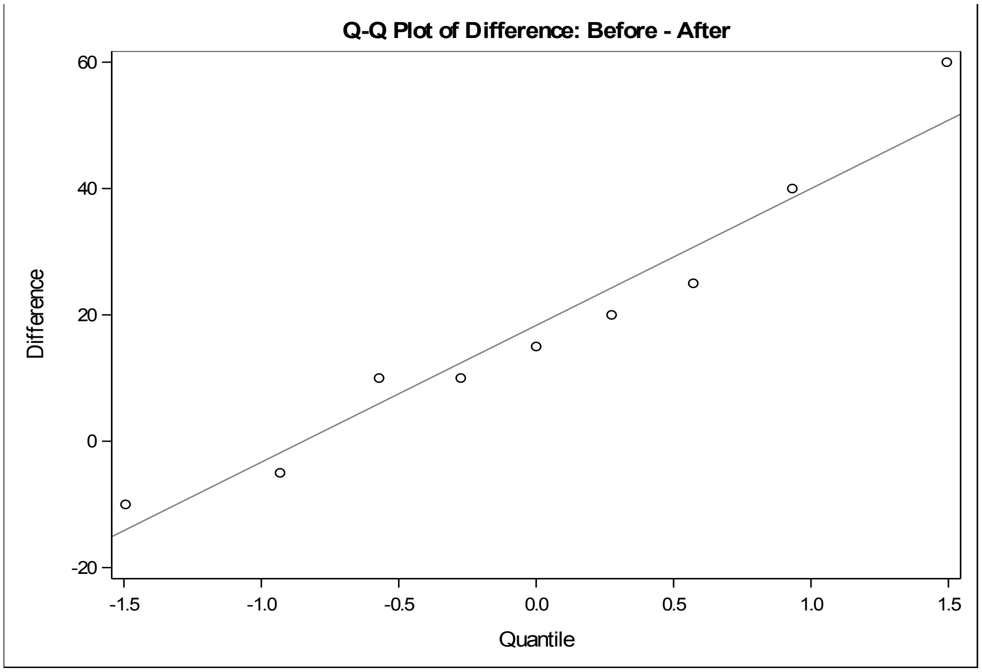
Assumptions

A paired t-test—histograms, box-plots, and qq-plots for which are produced below (Figure 8)—was run to preliminarily determine normalcy and variance. With the exception of a single outlier (differential value = 60), the data appears **roughly normal**, with a **variance** of (sd = 21.6506). Moreover, judging from the sparse description of study's design, one can assume that observation **independence** is adequately satisfied. However, since the sample size is not especially large (n = 9), it is difficult to make assumptions regarding the normalcy of the distribution for the population being examined. For such reasons, a signed-rank test would be appropriate in this case, especially for the sake of mitigating the influence of the outlying observation.

Figure 8:

N	Mean	Std Dev	Std Err	Minimum	Maximum
9	18.3333	21.6506	7.2169	-10.0000	60.0000





Analysis

To test for the impact of yoga on puzzle completion time, a **Signed-Rank test** was run to test the ***null-hypothesis (H₀)*** that: there is **no difference** in mean completion time before and after yoga treatment:

$$\mu_B - \mu_A = 0$$

The ***alternative-hypothesis (H_A)*** would be that: mean puzzle completion time is **greater** for patients before treatment:

$$\mu_B - \mu_A > 0$$

Hand-written analysis determined that, for this test $\mathbf{SR_{stat}} = \mathbf{39.5}$.

The SR-sum expected under H_0 would be: **Mean(S) = 22.5**, producing the following difference from a **test for location** shift: (39.5-22.5) = **17**

The corresponding Z-scores (**including continuity correction**) and p-values were as follows:

Z = 1.954751 → p-value (normal approx.) = **.0253**

On this basis, one would **reject the null-hypothesis *(H_0)***, determining that: there is moderate statistical evidence to suggest that **mean puzzle completion time is greater before treatment**.

Given the evidence for this sparsely described observational study—in which neither sampling nor allocation appear to be randomized, since participation in the study was voluntary and all subjects underwent the treatment—inferences regarding the difference in mean puzzle completion time **can only be drawn to the particular population of autistic children examined by the study**, though the results are still informative within that limited scope of inference. However, as sampling was determined by voluntary participation, inferences regarding **causal relations cannot be drawn.**

Part C)

Statistical software was used to identify the **test statistic** (**t_{stat}**) and **p-value** for a **Paired T-test** of the data. Figures and **confidence intervals** for the following statistics are produced below (Figure 9):

t_{stat} = 2.54 → p-value = .0347

Figure 9:

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
18.3333	1.6912	34.9755	21.6506	14.6241	41.4777

DF	t Value	Pr > |t|
8	2.54	0.0347

Part D)

Analysis

To test for the impact of yoga on puzzle completion time, a paired t-test was run to test the ***null-hypothesis (H_0)*** that: there is **no difference** in mean completion time before and after yoga treatment:

$$\mu_B - \mu_A = 0$$

The ***alternative-hypothesis (H_A)*** would be that: mean puzzle completion time is **greater** for patients before treatment:

$$\mu_B - \mu_A > 0$$

Assuming the null-hypothesis, with a sample size of *n* = 9, with *df* = 8, and $\propto =$ .05, the critical value would be **t_{crit} = 1.860.**

Statistical software determined that, for this test, **t_{stat}** $=$ **2.54.**

Correspondingly, the **p-value = .0347**

On this basis, one would **reject the null-hypothesis *(H_0)***, determining that: there is moderate statistical evidence to suggest that **mean puzzle completion time is greater before treatment**.

The associated **95% confidence interval** for the difference in mean completion time, providing **two-sides** for intuitive clarity, is as follows: [**1.6912**, **34.9755]**.

As before, given the evidence for this sparsely described observational study—in which neither sampling nor allocation appear to be randomized, since participation in the study was voluntary and all subjects underwent the treatment—inferences regarding the difference in mean puzzle completion time **can only be drawn to the particular population of autistic children examined by the study**, though the results are still informative within that limited scope of inference. However, as sampling was determined by voluntary participation, inferences regarding **causal relations cannot be drawn.**

Further tables for the results (Figure 10) are produced below:

Figure 10:

Moments			
N	9	**Sum Weights**	9
Mean	18.3333333	**Sum Observations**	165
Std Deviation	21.6506351	**Variance**	468.75
Skewness	0.7310904	**Kurtosis**	0.5328254
Uncorrected SS	6775	**Corrected SS**	3750
Coeff Variation	118.094373	**Std Error Mean**	7.21687836

Basic Statistical Measures			
Location		**Variability**	
Mean	18.33333	**Std Deviation**	21.65064
Median	15.00000	**Variance**	468.75000
Mode	10.00000	**Range**	70.00000
		Interquartile Range	15.00000

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	2.540341	**Pr > |t|**	0.0347
Sign	**M**	2.5	**Pr >= |M|**	0.1797
Signed Rank	**S**	18.5	**Pr >= |S|**	0.0313

Part E)

Concluding Analysis

On the basis of the resulting evidence above, although a comparison of the Signed-Rank test and Paired T-test for this data produced roughly similar **p-values** (Hand Written Signed Rank = **.0253**; SAS Signed Rank = **.0313**; SAS Paired T-test = **.0347**), closer examination of the data and study design suggested that **Signed-Rank would be an appropriate test**, on account of its **resistance to outliers**, its **lack of assumptions** regarding normalcy and variance for this rather small sample size, and its **sensitivity to the magnitude** of observed differences, all of which would produce more reliable test results in the event of a non-normal and highly variable population distribution, presumably fraught with outliers.

<u>Bonus Question</u>

Part A)

A permutation distribution was built for the Rank-Sum statistic for the trauma data above, using 5000 permutations. Tables (Figure 11) comparing the results for the **mean**, **standard deviation**, and **p-values** with those identified in Question #3 are produced below.

Figure 11:

Rank Sum Permutation – Results

Data Scores for Variable Measure Classified by Variable Trauma					
Trauma	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
N	8	167.70	196.853333	11.790497	20.962500
T	7	201.40	172.246667	11.790497	28.771429

Data Scores Two-Sample Test	
Statistic (S)	201.4000
Z	2.4726
One-Sided Pr > Z	0.0067
Two-Sided Pr > |Z|	0.0134

Monte Carlo Estimates for the Exact Test	
One-Sided Pr >= S	
Estimate	0.0002
99% Lower Conf Limit	<.0001
99% Upper Conf Limit	0.0007
Two-Sided Pr >= |S - Mean|	
Estimate	0.0002
99% Lower Conf Limit	<.0001
99% Upper Conf Limit	0.0007
Number of Samples	5000
Initial Seed	2345

Rank Sum Test (from Question #3 above) – Results

Wilcoxon Scores (Rank Sums) for Variable Measure Classified by Variable Trauma					
Trauma	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
N	8	38.0	64.0	8.633269	4.750000
T	7	82.0	56.0	8.633269	11.714286
Average scores were used for ties.					

Wilcoxon Two-Sample Test	
Statistic	82.0000
Normal Approximation	
Z	2.9537
One-Sided Pr > Z	0.0016
Two-Sided Pr > |Z|	0.0031
t Approximation	
One-Sided Pr > Z	0.0052
Two-Sided Pr > |Z|	0.0105
Z includes a continuity correction of 0.5.	

Conclusion

Comparative analysis revealed a higher, but still significant p-value for the **permuted version** of the Rank-Sum test, where the **(one-sided) p-val = .0067**.