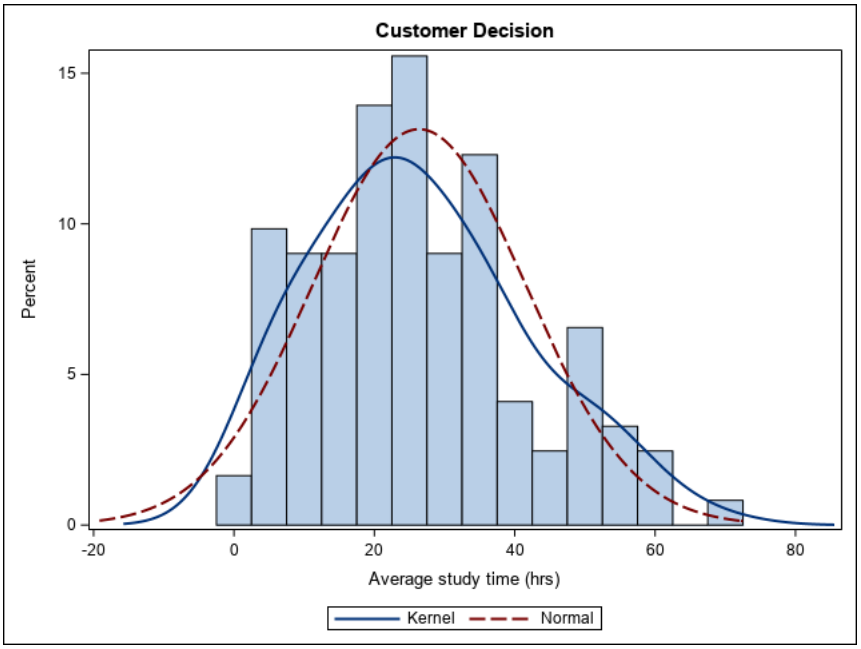
Homework 2

1a)

The histogram shows the average study time along with kernel density and the best fitting normal curve. Using the eyeballing approach, the hours of study distribution look similar to the normal distribution with a slight skew towards the right.



b)

We use the proc univariate to test the normality statistically. The output shows the test result for Normality. The question statement mentions that $\alpha = 0.05$ and here since the sample size is very small (N=122), the Shapiro test is used. The p-value is $0.0079 < 0.05$, so we reject the null hypothesis and conclude that the average study time does not follow the normal distribution.

The UNIVARIATE Procedure
Variable: AveTime (Average study time (hrs))

Moments			
N	122	Sum Weights	122
Mean	26.3651016	Sum Observations	3216.54239
Std Deviation	15.1768129	Variance	230.335649
Skewness	0.48016013	Kurtosis	-0.307681
Uncorrected SS	112675.08	Corrected SS	27870.6135
Coeff Variation	57.564022	Std Error Mean	1.37404408

Basic Statistical Measures			
Location		Variability	
Mean	26.36510	Std Deviation	15.17681
Median	24.82108	Variance	230.33565
Mode	.	Range	68.23397
		Interquartile Range	21.45978

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	19.18796	Pr > |t|	<.0001
Sign	M	61	Pr >= |M|	<.0001
Signed Rank	S	3751.5	Pr >= |S|	<.0001

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.969928	Pr < W	0.0079
Kolmogorov-Smirnov	D	0.067436	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.113863	Pr > W-Sq	0.0765
Anderson-Darling	A-Sq	0.853444	Pr > A-Sq	0.0279

Quantiles (Definition 5)	
Level	Quantile
100% Max	69.006831
99%	60.021982
95%	52.654466
90%	49.437782
75% Q3	36.602812
50% Median	24.821082
25% Q1	15.143034
10%	6.191171
5%	4.790607
1%	1.677311
0% Min	0.772862

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.772862	96	56.3370	80
1.677311	37	59.7214	38
2.847294	29	59.8352	61
3.317500	28	60.0220	35
3.926174	42	69.0068	41

c)

Proc corr is used to get the correlation between Units, AvgTime, and GPA.
H0: Correlation coefficient = 0
H1: Correlation coefficient ≠ 0

Summary Statistics for Customer Decision

The CORR Procedure

Section of course=01

3 Variables:	Units AveTime GPA

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
Units	58	13.79310	3.15538	800.00000	9.00000	19.00000	Number of units enrolled
AveTime	58	29.68670	14.46548	1722	0.77286	69.00683	Average study time (hrs)
GPA	58	3.30138	0.39409	191.48000	2.42000	3.94000	GPA

Pearson Correlation Coefficients, N = 58
Prob > |r| under H0: Rho=0

	Units	AveTime	GPA
Units Number of units enrolled	1.00000	0.42598 0.0009	-0.15327 0.2507
AveTime Average study time (hrs)	0.42598 0.0009	1.00000	-0.34324 0.0083
GPA GPA	-0.15327 0.2507	-0.34324 0.0083	1.00000

- The units enrolled and GPA has a weak negative relationship with r = -0.153 and correlation that is not statistically significant is with a p-value (0.2507) greater than alpha (0.05).
- The units enrolled and average study time has a moderate positive relationship with r= 0.476. and correlation that is statistically significant is with a p-value (0.0009) less than alpha (0.05)
- The GPA and average study time have a weak negative relationship with r= -0.343. and correlation that is statistically significant is with a p-value (0.0083) less than alpha (0.05)
- If students are enrolled in more units, they would spend more time for study. The output also shows that spending too much time may also lead to a decrease in their GPA. The reason for the decrease in GPA may also be due to a lack of time for each class because of too many courses.

2a)

Obs	SID	plaque1	plaque2	treat
126	126	0.7683	0.8089	1
127	127	0.8118	0.7580	1
128	128	0.8060	0.7981	1
129	129	0.8543	0.8570	1
130	130	0.6351	0.7407	1
131	131	0.7110	0.7235	1
132	132	0.8930	0.6294	1
133	133	0.6799	0.6769	1
134	134	0.9116	0.7369	1
135	135	0.8569	0.7063	1
136	136	0.7901	0.9245	1
137	137	0.7243	0.7551	1
138	138	0.8148	0.8010	1
139	139	0.8397	0.7527	1
140	140	0.6937	0.6699	1
141	141	0.7270	0.9575	1

The SAS System

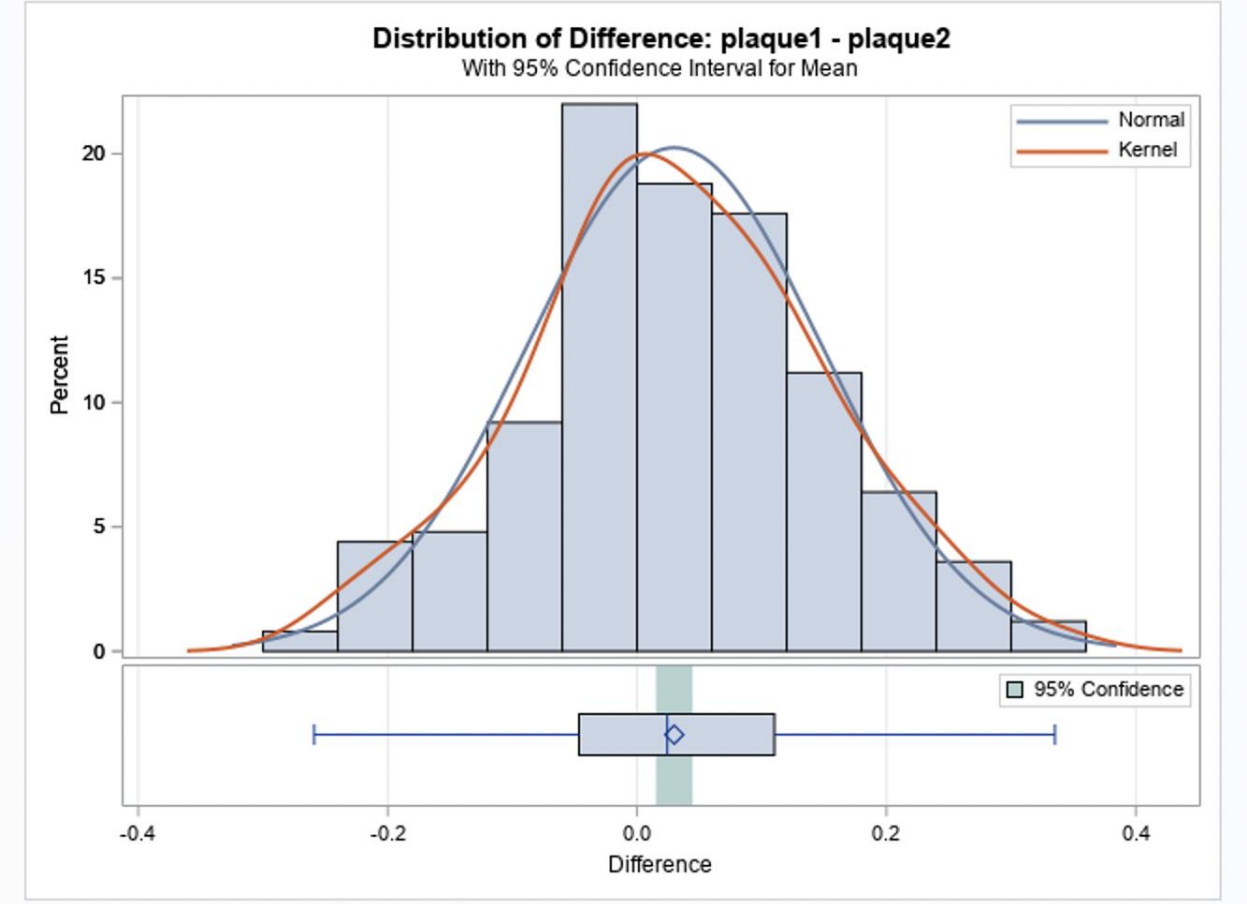
The TTEST Procedure

Difference: plaque1 - plaque2

N	Mean	Std Dev	Std Err	Minimum	Maximum
250	0.0298	0.1182	0.00748	-0.2590	0.3351

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0.0298	0.0150	0.0445	0.1182	0.1087	0.1296

DF	t Value	Pr > |t|
249	3.98	<.0001



After performing paired t-test on 250 observations containing plaque1 and plaque2, we came up with the following hypothesis:

H0: plaque1 - plaque2 = 0

H1: plaque1 - plaque2 ≠ 0

The test static shows that the plaque level plummeted by 0.0298 and p-value being 0.001, there is strong evidence against the null hypothesis. Hence, we can conclude that there is a statistical difference in the plaque level before and after the treatment.

b)

The SAS System

The TTEST Procedure

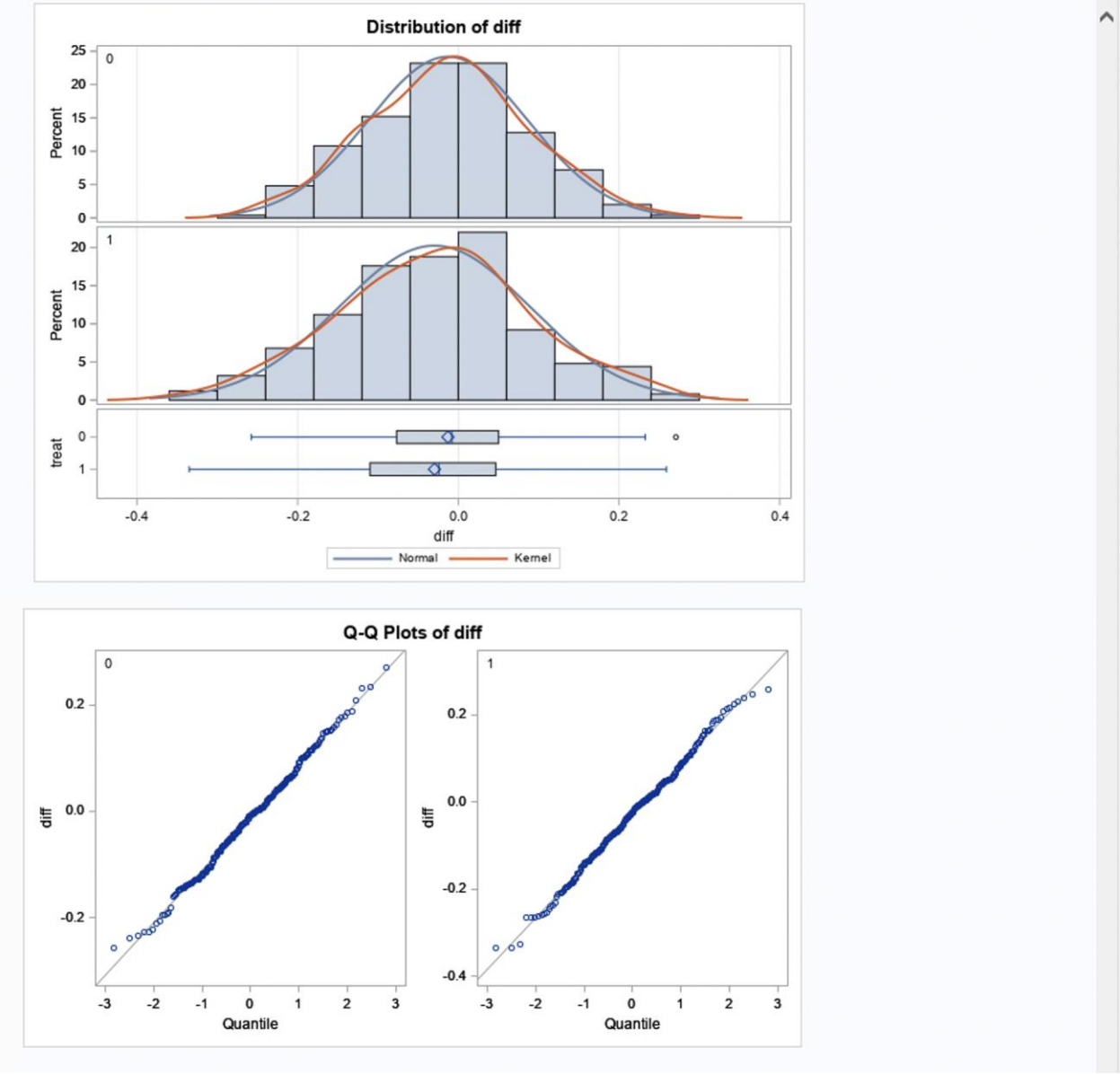
Variable: diff

treat	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		250	-0.0130	0.0990	0.00626	-0.2577	0.2709
1		250	-0.0298	0.1182	0.00748	-0.3351	0.2590
Diff (1-2)	Pooled		0.0168	0.1091	0.00975		
Diff (1-2)	Satterthwaite		0.0168		0.00975		

treat	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		-0.0130	-0.0253	-0.00063	0.0990	0.0911	0.1086
1		-0.0298	-0.0445	-0.0150	0.1182	0.1087	0.1296
Diff (1-2)	Pooled	0.0168	0.000735	Infty	0.1091	0.1027	0.1163
Diff (1-2)	Satterthwaite	0.0168	0.000734	Infty			

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	498	1.72	0.0427
Satterthwaite	Unequal	483.15	1.72	0.0427

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	249	249	1.43	0.0053



To conduct this test, we perform a difference-in-difference study on the plaque level of the two groups.
From the t-test data, we can see that there is a drop in the plaque level in both groups. However, the drop is higher in the group who went through the treatment (0.0298) compared to the control group (0.0130).

Hence, our hypothesis is as follows:
H0: diffplacebo –difftreatment >= 0
H1: diffplacebo –difftreatment < 0

Next, we perform ttest for equal variance to validate our hypothesis:

H0: variance if both the groups are equal
H1: variance if both the groups are different

The equality of variance test between placebo and treatment groups yields a p-value of 0.0053. There is statistically good evidence to reject the hypothesis which implies that both groups have different variances.

Our hypothesis is as follows:
H0: diffplacebo –difftreatment >= 0
H1: diffplacebo –difftreatment < 0

We perform satterthwaite unequal test. one tail value of p-value = 0.0855/2 = 0.04275. This shows that there is good evidence against the null hypothesis. Hence, with a 95% confidence we can conclude that there is a difference in plaque level before and after treatment.

c)

Tests used in part b are more reliable over the test used in part a. Dataset used in part b allowed us to consider that external factors that could affect the plaque level with the aid of placebo group. Whereas dataset in part a doesn't consider these external factors making tests used in part b more reliable in the real world.

d)

T-test for smoking:

we perform t-test to validate if the average number cigarettes smoked by both placebo and treatment group is equal or not.

H0: cig_smokedplacebo – cig_smokedtreatment = 0
H1: cig_smokedplacebo – cig_smokedtreatment ≠ 0

The SAS System

The TTEST Procedure

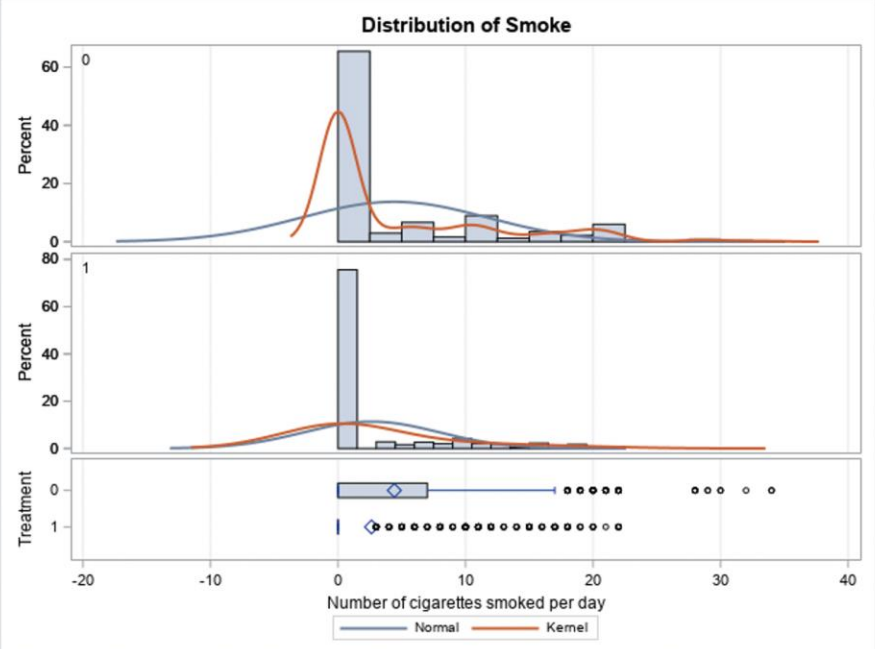
Variable: Smoke (Number of cigarettes smoked per day)

Treatment	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		750	4.4200	7.2705	0.2655	0	34.0000
1		750	2.6267	5.2679	0.1924	0	22.0000
Diff (1-2)	Pooled		1.7933	6.3487	0.3278		
Diff (1-2)	Satterthwaite		1.7933		0.3278		

Treatment	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		4.4200	3.8988	4.9412	7.2705	6.9203	7.6584
1		2.6267	2.2490	3.0043	5.2679	5.0141	5.5490
Diff (1-2)	Pooled	1.7933	1.1502	2.4364	6.3487	6.1293	6.5845
Diff (1-2)	Satterthwaite	1.7933	1.1502	2.4365			

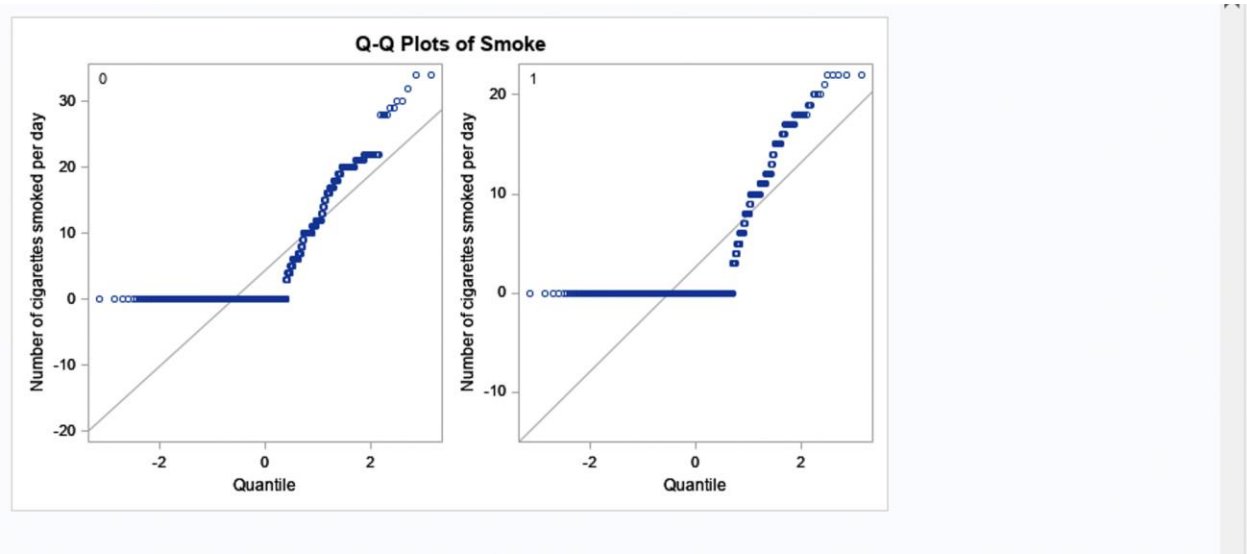
Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	1498	5.47	<.0001
Satterthwaite	Unequal	1365.5	5.47	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	749	749	1.90	<.0001



The above data concludes that average number of cigarettes smoked per day by the treatment group (2.6267) is less than the cigarettes smoked per day by the placebo group(4.42).

The equality of variance test gives a p-value less than 0.0001. Hence, there is statistical evidence to reject the hypothesis that both groups have equal variances.



T-test for alcohol:

we perform t-test to validate if the average number of alcoholic drinks consumed per day by both placebo and treatment group is equal.

H0: drinks_countplacebo – drinks_counttreatment = 0
H1: drinks_countplacebo – drinks_counttreatment ≠ 0

The SAS System

The TTEST Procedure

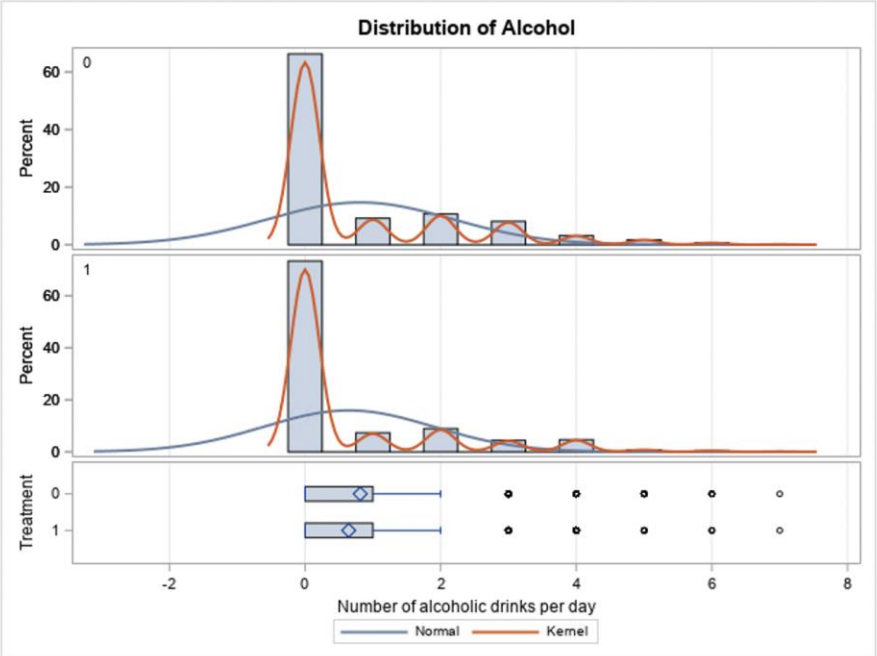
Variable: Alcohol (Number of alcoholic drinks per day)

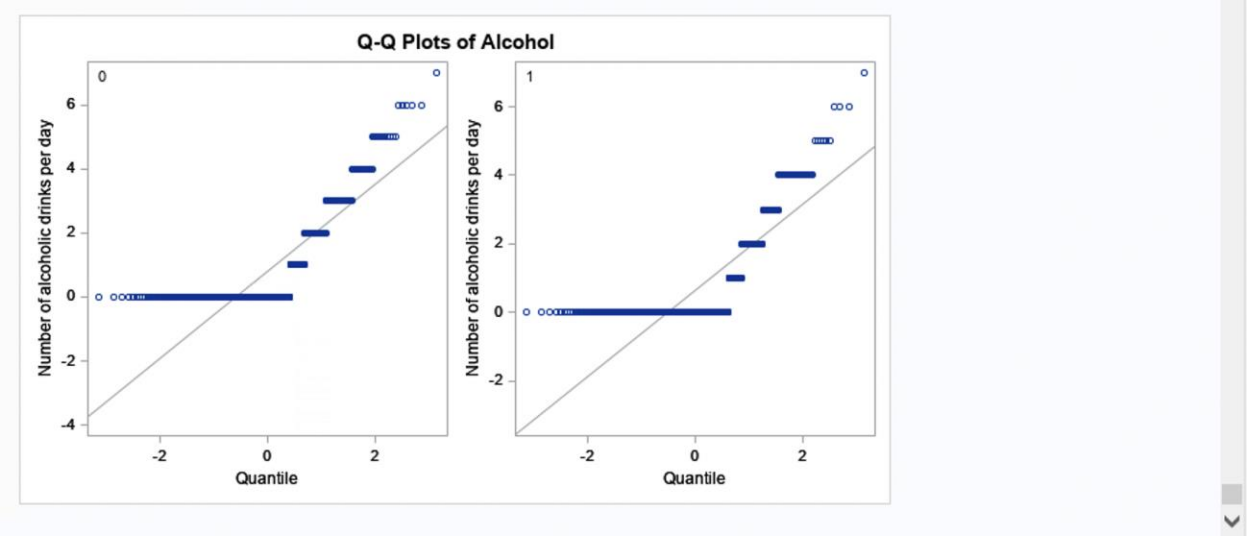
Treatment	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		750	0.8133	1.3573	0.0496	0	7.0000
1		750	0.6440	1.2530	0.0458	0	7.0000
Diff (1-2)	Pooled		0.1693	1.3062	0.0675		
Diff (1-2)	Satterthwaite		0.1693		0.0675		

Treatment	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		0.8133	0.7160	0.9106	1.3573	1.2919	1.4297
1		0.6440	0.5542	0.7338	1.2530	1.1926	1.3199
Diff (1-2)	Pooled	0.1693	0.0370	0.3016	1.3062	1.2611	1.3547
Diff (1-2)	Satterthwaite	0.1693	0.0370	0.3016			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	1498	2.51	0.0122
Satterthwaite	Unequal	1488.5	2.51	0.0122

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	749	749	1.17	0.0288





The above data concludes that average number of alcoholic drinks consumed per day by the treatment group (0.6440) is less than the average number of alcoholic drinks consumed per day by the placebo group (0.8133).

The equality of variance test gives a p-value less 0.0288. Hence, there is statistical evidence to reject the hypothesis that both groups have equal variances.

We can conclude that the treatment group is much healthier compared to placebo group. So, the randomization among the groups is not done perfectly.