Yizhan Ao

Homework 3

April 4th, 2022

```
DATA HW3;
INFILE '/home/u58594663/my_shared_file_links/schimiak/OldClassData.csv' delimiter=',' dsd;
INPUT Subject $
Gender $
Phone $
Campus $
Grade $
Car $
Optimist
Math
Siblings
Pets
Credit_Hours
Social_Media
Extra_Curricular
Height
HS_GPA
Exercise
Time_To_Get_Ready
Distance;
```

1. Q1: Do a PROC FORMAT to change the following (answer to parts a and c put IN CODE.):

```
PROC FORMAT;
VALUE $New_Gender ‘M’=‘Male’ ‘F’ =‘Female’;
RUN;
```

```
PROC FORMAT;
VALUE NEW_MATH 1 ="I really like math."
2="I somewhat like math."
3="I could take math or leave it."
4="I really don't like math."
5="I'd rather have a root canal.";
RUN;

PROC PRINT;
FORMAT Gender $New_Gender.;
FORMAT Math NEW_MATH.;
RUN;
```

Obs	Subject	Gender	Phone	Campus	Grade	Car	Optimist	Math	Siblings	Pets	Credit_Hours	Social_Media	Extra_Curricular	Height	HS_GPA	Exercise	Time_To_Get_Ready	Distance
1	1	F	Ap	N	A	N	3	I somewhat like math.	1	1	3	3	2	67.0	3.500	3.5	60	10.0
2	3	F	Ap	N	A	Y	3	I'd rather have a root canal.	1	1	3	5	2	64.0	2.900	2.0	35	2.0
3	5	F	Ap	Y	B	Y	2	I could take math or leave it.	6	4	17	5	.	67.0	3.900	6.0	40	8.0
4	7	F	Ap	N	B	Y	2	I really don't like math.	3	2	6	4	3	66.0	3.400	4.0	30	50.0
5	10	F	Ap	Y	C	N	2	I could take math or leave it.	.	1	3	2	3	62.0	2.770	4.5	75	40.0
6	11	F	Ap	Y	B	Y	4	I could take math or leave it.	2	6	2	4	1	65.0	3.640	5.0	10	0.0
7	12	F	Ap	N	A	N	2	I could take math or leave it.	1	1	9	6	2	66.0	3.800	3.0	30	23.5
8	14	F	Ap	N	A	Y	2	I'd rather have a root canal.	2	4	6	3	2	65.0	3.570	3.0	15	12.0
9	15	F	Ap	Y	A	N	3	I could take math or leave it.	1	.	6	5	2	64.0	3.400	7.0	30	24.0
10	19	F	Ap	N	B	Y	2	I'd rather have a root canal.	2	1	3	4	3	64.0	4.120	1.0	35	3.0
11	24	F	Ap	N	A	N	3	I somewhat like math.	1	1	3	3	2	67.0	3.500	3.5	60	10.0
12	26	F	Ap	N	A	Y	3	I'd rather have a root canal.	1	1	3	5	2	64.0	2.900	2.0	35	2.0
13	28	F	Ap	Y	B	Y	2	I could take math or leave it.	6	4	17	5	3	67.0	3.900	6.0	40	8.0
14	30	F	Ap	N	B	Y	2	I really don't like math.	3	2	6	4	3	66.0	3.400	4.0	30	50.0
15	33	F	Ap	Y	C	N	2	I could take math or leave it.	2	1	3	2	3	62.0	2.770	4.5	75	40.0
16	34	F	Ap	Y	B	Y	4	I could take math or leave it.	2	6	2	4	1	65.0	3.640	5.0	10	0.0
17	35	F	Ap	N	A	N	2	I could take math or leave it.	1	1	9	6	2	66.0	3.800	3.0	30	23.5
18	37	F	Ap	N	A	Y	2	I'd rather have a root canal.	2	4	6	3	2	65.0	3.570	3.0	15	12.0
19	38	F	Ap	Y	A	N	3	I could take math or leave it.	1	.	6	5	2	64.0	3.400	7.0	30	24.0
20	42	F	Ap	N	B	Y	2	I'd rather have a root canal.	2	1	3	4	3	64.0	4.120	1.0	35	3.0
21	2	M	Ap	N	C	Y	2	I'd rather have a root canal.	1	1	6	3	1	67.0	.	5.0	60	1.0
22	4	M	Ap	N	B	Y	3	I'd rather have a root canal.	1	1	4	4	4	71.0	3.400	0.0	30	20.0
23	6	M	An	N	B	N	2	I could take math or leave it.	2	1	3	2	1	63.0	3.300	3.0	30	10.0
24	8	M	Ap	Y	A	N	3	I really don't like math.	1	1	6	3	1	69.0	2.800	7.0	30	0.0
25	9	M	Ap	Y	A	Y	1	I somewhat like math.	2	2	6	5	3	70.0	2.300	4.0	60	1.0
26	13	M	Ap	Y	A	Y	3	I really don't like math.	1	1	12	4	.	74.0	4.875	5.0	60	14.0
27	16	M	Ap	N	A	N	2	I really like math.	1	1	6	2	2	84.0	.	7.0	35	0.0
28	17	M	Ap	N	A	Y	2	I somewhat like math.	4	1	3	3	2	69.0	3.000	3.0	30	20.0
29	18	M	Ap	Y	A	Y	2	I really don't like math.	2	1	3	2	3	68.0	3.700	3.0	90	15.0
30	20	M	An	Y	A	N	2	I really don't like math.	2	1	3	5	2	68.5	3.600	5.0	20	1.0
31	21	M	Ap	Y	B	Y	3	I somewhat like math.	2	1	6	4	3	73.0	3.400	3.0	60	10.9
32	22	M	Ap	N	A	N	1	I'd rather have a root canal.	2	1	3	3	1	71.0	2.500	7.0	45	3.0
33	23	M	Ap	N	C	N	1	I could take math or leave it.	.	1	3	3	3	69.0	4.200	3.0	60	1.0
34	25	M	Ap	N	C	Y	2	I'd rather have a root canal.	1	1	6	3	1	67.0	.	5.0	60	1.0
35	27	M	Ap	N	B	Y	3	I'd rather have a root canal.	1	1	4	4	4	71.0	3.400	0.0	30	20.0
36	29	M	An	N	B	N	2	I could take math or leave it.	2	1	3	2	1	63.0	3.300	3.0	30	10.0
37	31	M	Ap	Y	A	N	3	I really don't like math.	1	1	6	3	1	69.0	2.800	7.0	30	0.0
38	32	M	Ap	Y	A	Y	1	I somewhat like math.	2	2	6	5	3	70.0	2.300	4.0	60	1.0
39	36	M	Ap	Y	A	Y	3	I really don't like math.	1	1	12	4	2	74.0	4.875	5.0	60	14.0
40	39	M	Ap	N	A	N	2	I really like math.	1	1	6	2	2	84.0	.	7.0	35	0.0
41	40	M	Ap	N	A	Y	2	I somewhat like math.	4	1	3	3	2	69.0	3.000	3.0	30	20.0
42	41	M	Ap	Y	A	Y	2	I really don't like math.	2	1	3	2	3	68.0	3.700	3.0	90	15.0
43	43	M	An	Y	A	N	2	I really don't like math.	2	1	3	5	2	68.5	3.600	5.0	20	1.0
44	44	M	Ap	Y	B	Y	3	I somewhat like math.	2	1	6	4	3	73.0	3.400	3.0	60	10.9
45	45	M	Ap	N	A	N	1	I'd rather have a root canal.	2	1	3	3	1	71.0	2.500	7.0	45	3.0
46	46	M	Ap	N	C	N	1	I could take math or leave it.	4	1	3	3	3	69.0	4.200	3.0	60	1.0

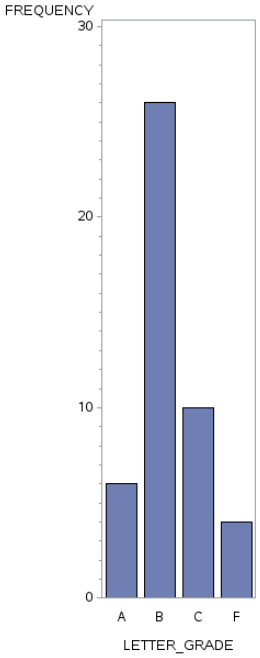
1. Q2: Create a new variable for the letter grade (LETTER_GRADE) for the HS_GPA

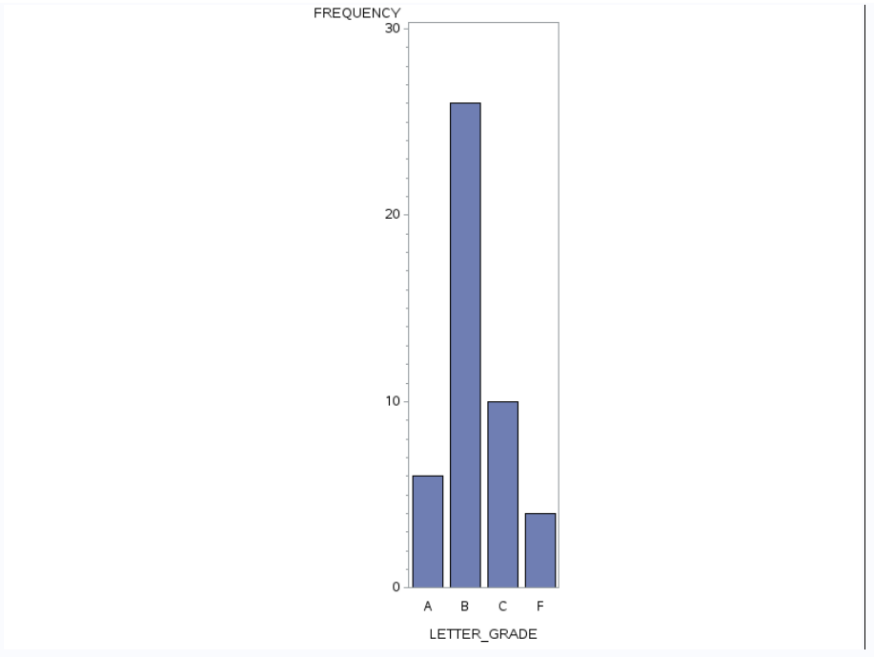
```
if HS_GPA>= 4.0 THEN LETTER_GRADE = 'A';
else if HS_GPA >= 3.0 AND HS_GPA <=4.0 THEN LETTER_GRADE = 'B';
else if HS_GPA >= 2.0 AND HS_GPA <=3.0 THEN LETTER_GRADE = 'C';
else if HS_GPA >= 2.0 AND HS_GPA <=1.0 THEN LETTER_GRADE = 'D';
else if HS_GPA LT 1.0 THEN LETTER_GRADE = 'F';
else if HS_GPA = . THEN LETTER_GRADE = '.';
```

Q3:

```
PROC GCHART DATA = HW3;
VBAR LETTER_GRADE;
```

RUN;





Q4: Do a hypothesis test: At a 0.05 significance level, test if there is a difference in the between the Male and Female High School GPA? State your conclusion based on your SAS results in the context of the question.

PROC TTEST ALPHA=0.05;

CLASS Gender;

VAR HS_GPA;

```
RUN;

PROC NPAR1WAY WILCOXON;

CLASS Gender;

VAR HS_GPA;

RUN;
```

Null Hypothesis (H0): GPA(males) = GPA(females).

Alternative Hypothesis (Ha): GPA(males) $\neq$ GPA(females).

We can't use a typical T-test since there's an outlier in the sample. For our scenario, we'll need to employ a Wilcoxon. Wilcoxon's p-value is 0.2288, which is larger than 0.05. This indicates that we do not reject the null hypothesis and that the alternative hypothesis is not supported by sufficient evidence.

The TTEST Procedure

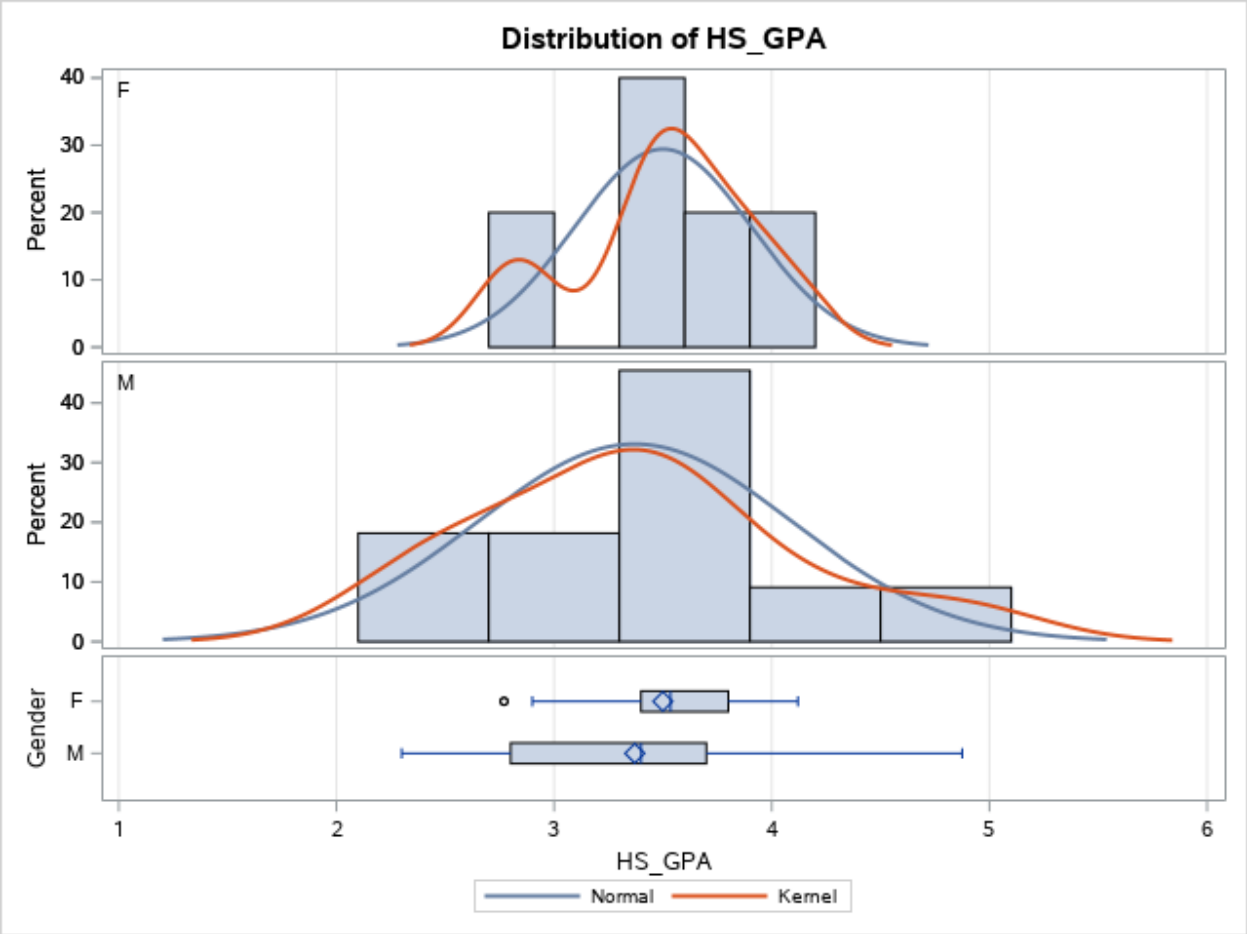
Variable: HS_GPA

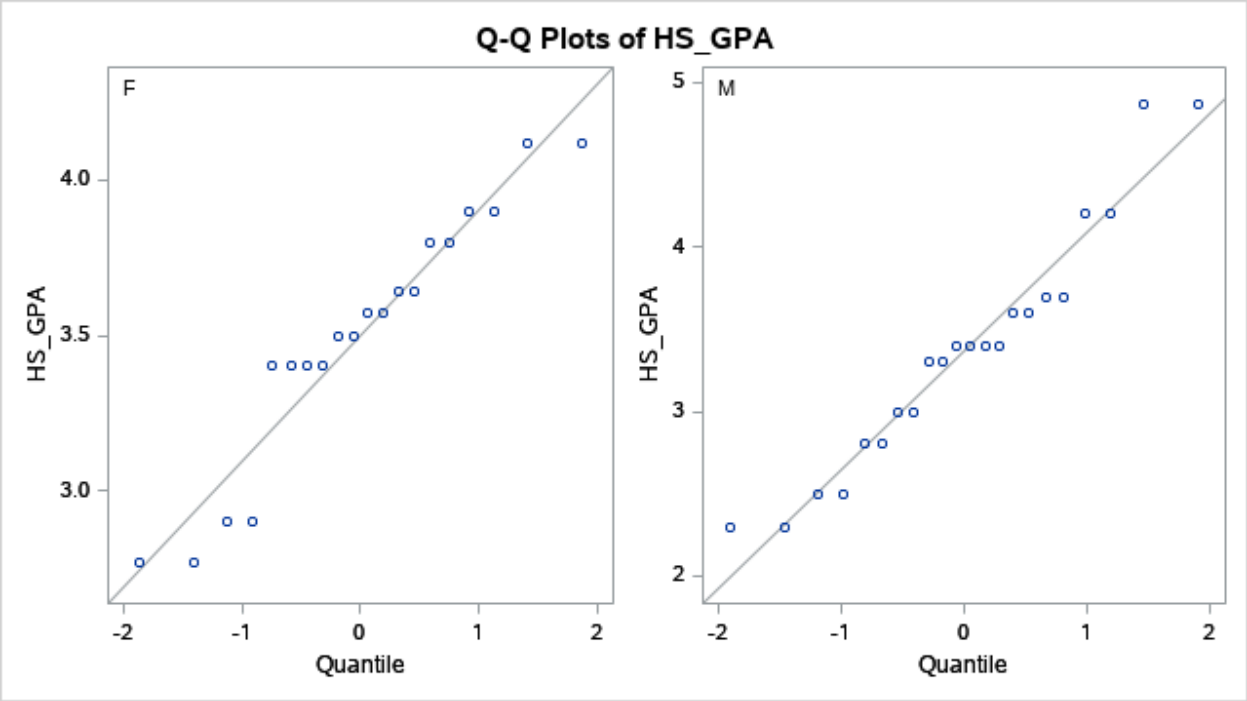
Gender	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
F		20	3.5000	0.4068	0.0910	2.7700	4.1200
M		22	3.3705	0.7228	0.1541	2.3000	4.8750
Diff (1-2)	Pooled		0.1295	0.5940	0.1835		
Diff (1-2)	Satterthwaite		0.1295		0.1789		

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
F		3.5000	3.3096	3.6904	0.4068	0.3093	0.5941
M		3.3705	3.0500	3.6909	0.7228	0.5561	1.0330
Diff (1-2)	Pooled	0.1295	-0.2414	0.5005	0.5940	0.4877	0.7601
Diff (1-2)	Satterthwaite	0.1295	-0.2343	0.4933			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	40	0.71	0.4844
Satterthwaite	Unequal	33.664	0.72	0.4741

Equality of Variances				
Method	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Folded F	21	19	3.16	0.0146





The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable HS_GPA Classified by Variable Gender

Gender	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Average scores were used for ties.					
F	20	474.0	430.0	39.544444	23.70
M	22	429.0	473.0	39.544444	19.50

Wilcoxon Two-Sample Test

Statistic	Z	Pr > Z	Pr > |Z|	t Approximation	
				Pr > Z	Pr > |Z|
Z includes a continuity correction of 0.5.					
474.0000	1.1000	0.1357	0.2713	0.1389	0.2777

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
1.2380	1	0.2658

Q5; Do a hypothesis test:

PROC TTEST ALPHA=0.05 H0=3.5;

VAR Exercise;

RUN;

PROC TTEST ALPHA=0.05 H0=3.5 SIDES=U;

VAR Exercise;

RUN;

Null Hypothesis (H0): Students exercise 3.5 times a week.

Alternative Hypothesis (Ha): Students exercise over 3.5 times a week.

Since there is no outlier in the sample, we can use a traditional T-test. The p-value in the T-test is 0.0184, which is smaller than 0.05. This means that we reject the null, and there is enough evidence to support the alternative hypothesis.

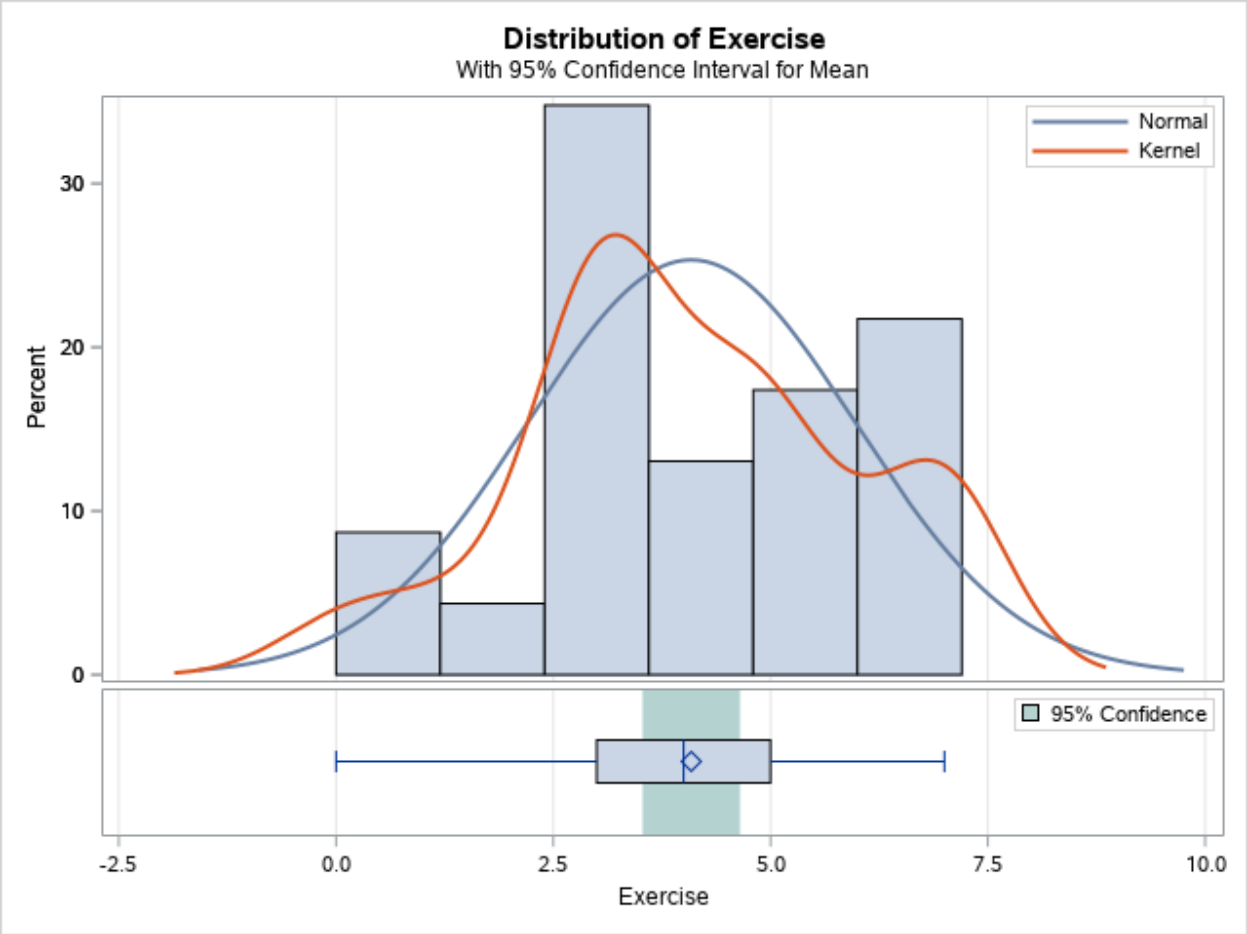
The TTEST Procedure

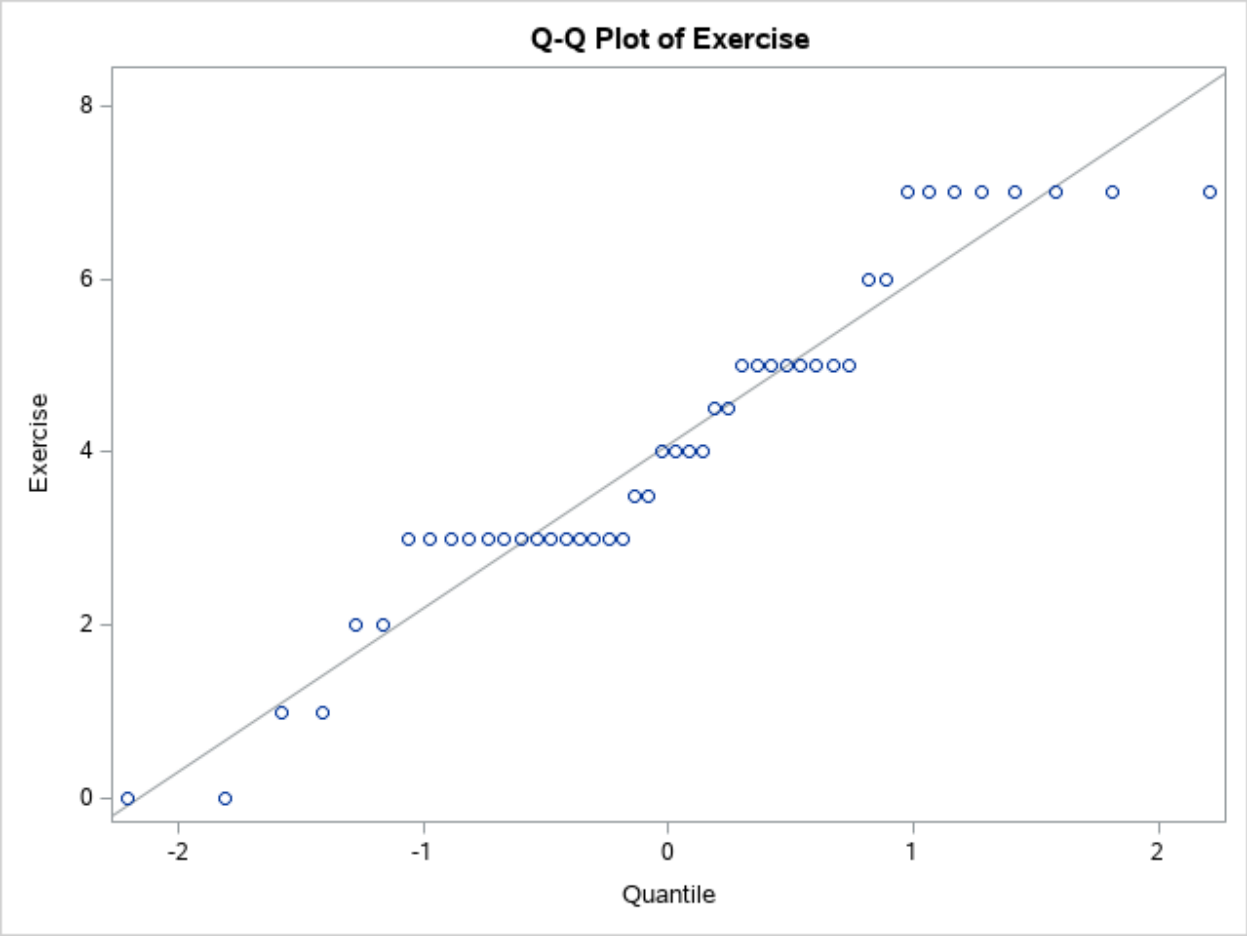
Variable: Exercise

N	Mean	Std Dev	Std Err	Minimum	Maximum
46	4.0870	1.8895	0.2786	0	7.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
4.0870	3.5259	4.6481	1.8895	1.5672	2.3798

DF	t Value	Pr > |t|
45	2.11	0.0407


Distribution of Exercise
With 95% Confidence Interval for Mean
Normal
Kernel
Percent
30
20
10
0
95% Confidence
-2.5
0.0
2.5
5.0
7.5
10.0
Exercise



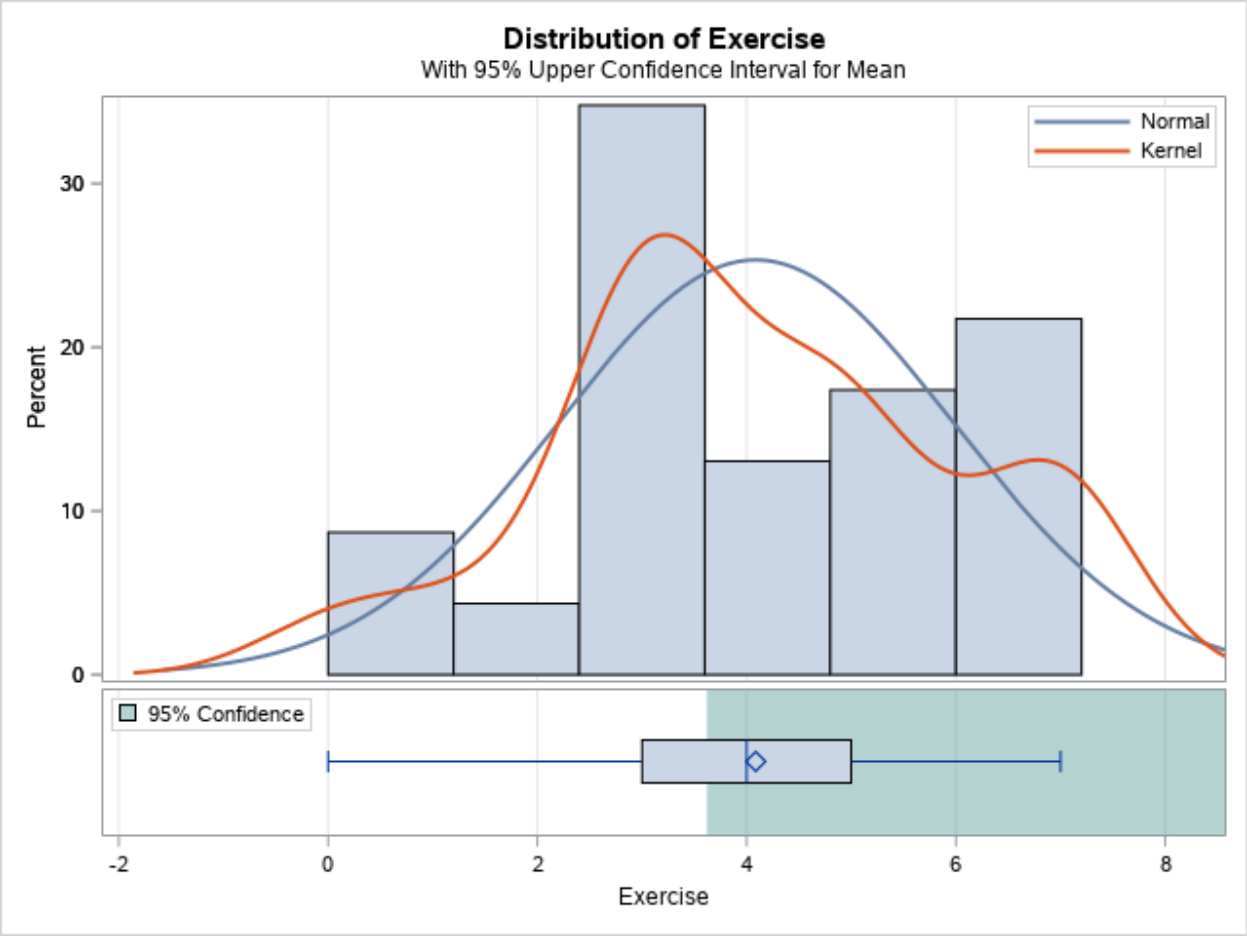
The TTEST Procedure

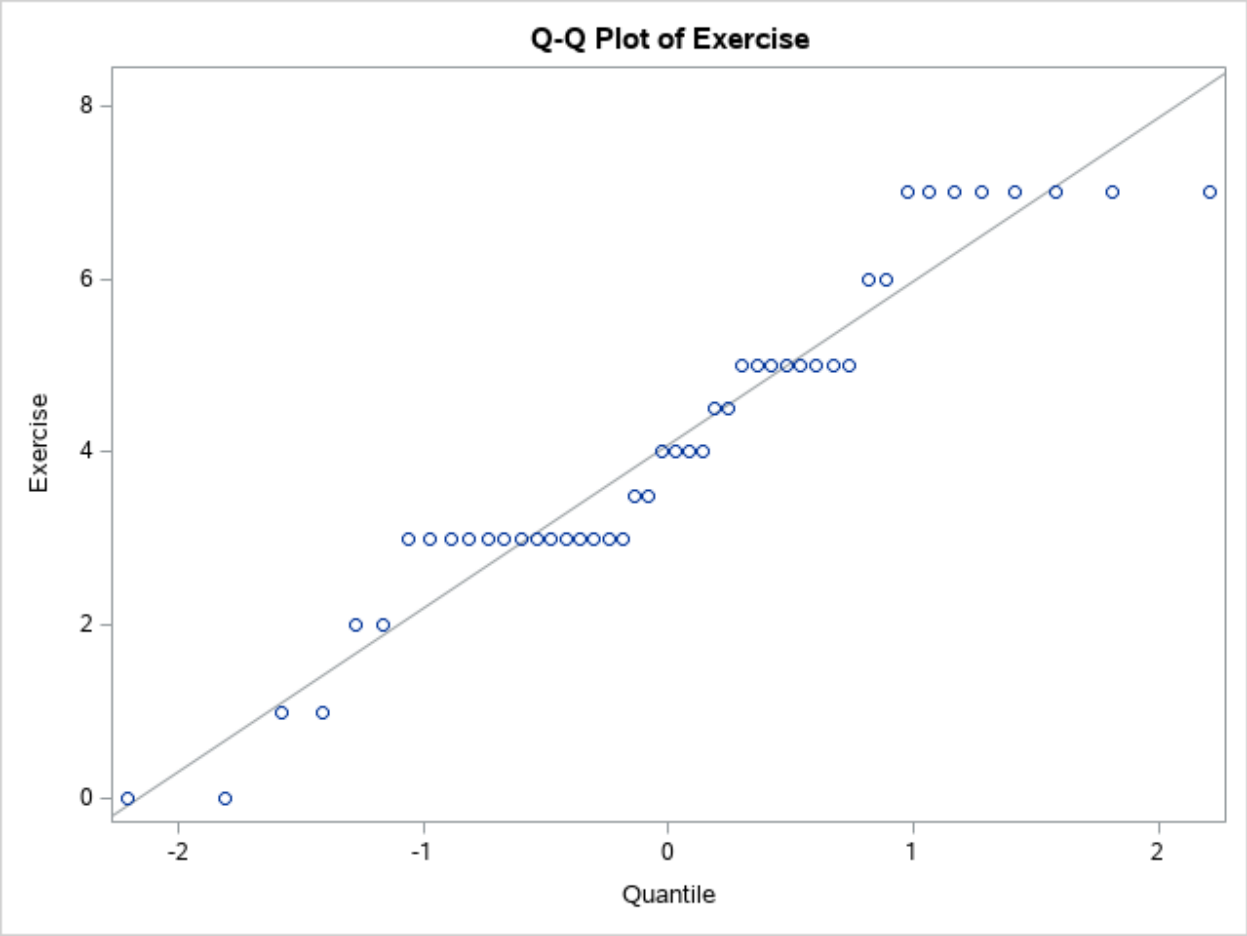
Variable: Exercise

N	Mean	Std Dev	Std Err	Minimum	Maximum
46	4.0870	1.8895	0.2786	0	7.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
4.0870	3.6191	Infty	1.8895	1.5672	2.3798

DF	t Value	Pr > t
45	2.11	0.0204


Distribution of Exercise
With 95% Upper Confidence Interval for Mean
Normal
Kernel
Percent
30
20
10
0
95% Confidence
-2
0
2
4
6
8
Exercise



Q6:

PROC MEANS CLM ALPHA=0.05 MAXDEC=2;

VAR TIME_TO_GET_READY;

RUN;

People take between 38.77 and 55.14 minutes to get ready on average given a 95% confidence interval.

The MEANS Procedure

Analysis Variable : Time_To_Get_Ready	
Lower 95% CL for Mean	Upper 95% CL for Mean
36.32	48.03