Final Exam Part II

You have three problems to complete. Please answer the questions first. Then upload your program and your results after you have answered the question for each problem.

PROBLEM 1:

A research study was conducted to examine the effectiveness of different fertilizers. A particular scrub was randomly assigned to one of three groups: Fertilizer A, Fertilizer B, and The Control (no fertilizer). After four weeks the height of the scrub was recorded:

Fertilizer A	Fertilizer B	Control
38	22	14
47	19	16
39	18	11
35	23	18
42	21	15

1. Show, using SAS, that all assumptions are met to do an ANOVA. Then do the ANOVA.
    a. State your conclusions making sure to speak to all pertinent information and assumptions.
        i. Assumptions
            1. Check for Equal variances (show output) (3 pts)

Variable: _Difference_ (Difference: fa - control)

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.96203	Pr < W	0.7917
Kolmogorov-Smirnov	D	0.189853	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.030866	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.207978	Pr > A-Sq	>0.2500

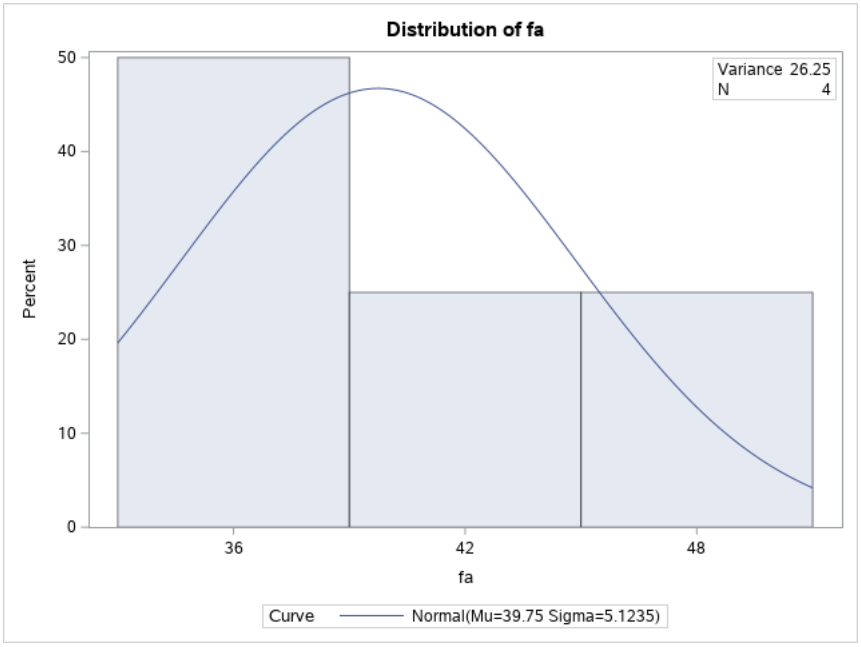
Variable: _Difference_ (Difference: fa - fb)

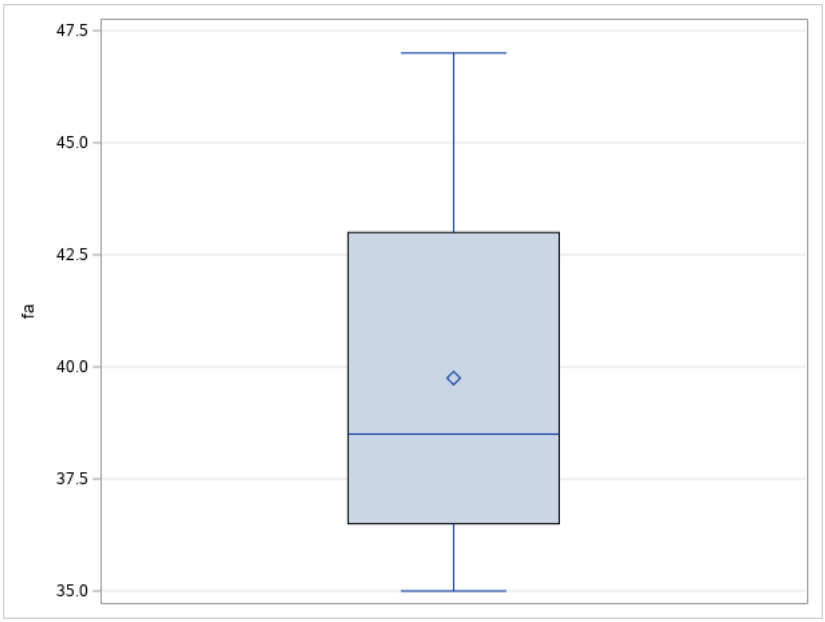
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.980031	Pr < W	0.9022
Kolmogorov-Smirnov	D	0.181232	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.025596	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.179262	Pr > A-Sq	>0.2500

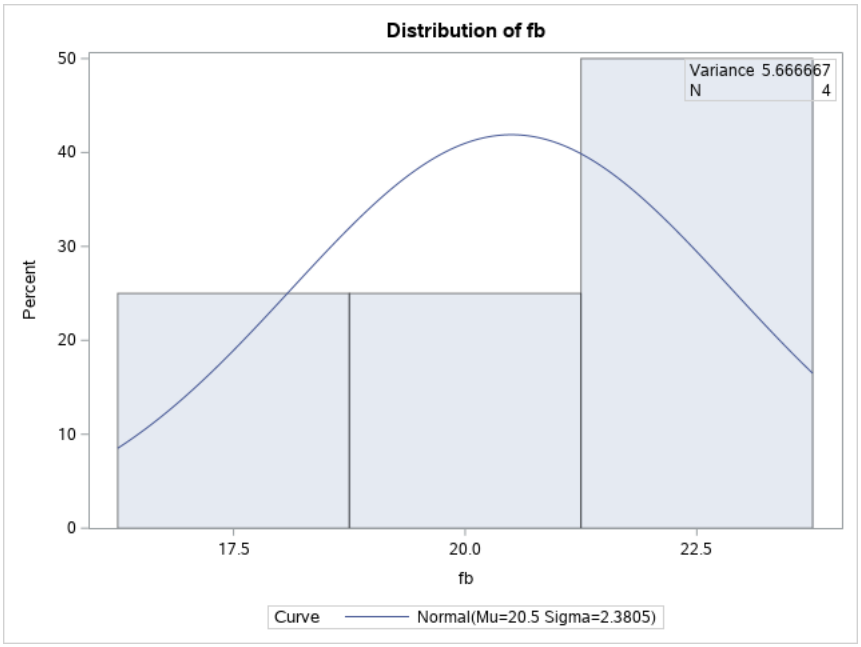
Variable: _Difference_ (Difference: control - fb)

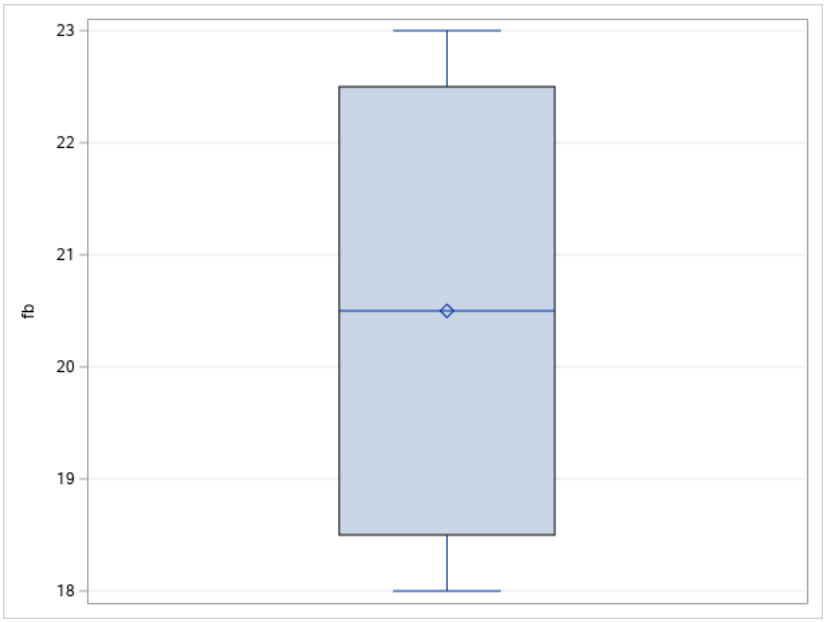
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.963072	Pr < W	0.7982
Kolmogorov-Smirnov	D	0.213533	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.029941	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.200349	Pr > A-Sq	>0.2500

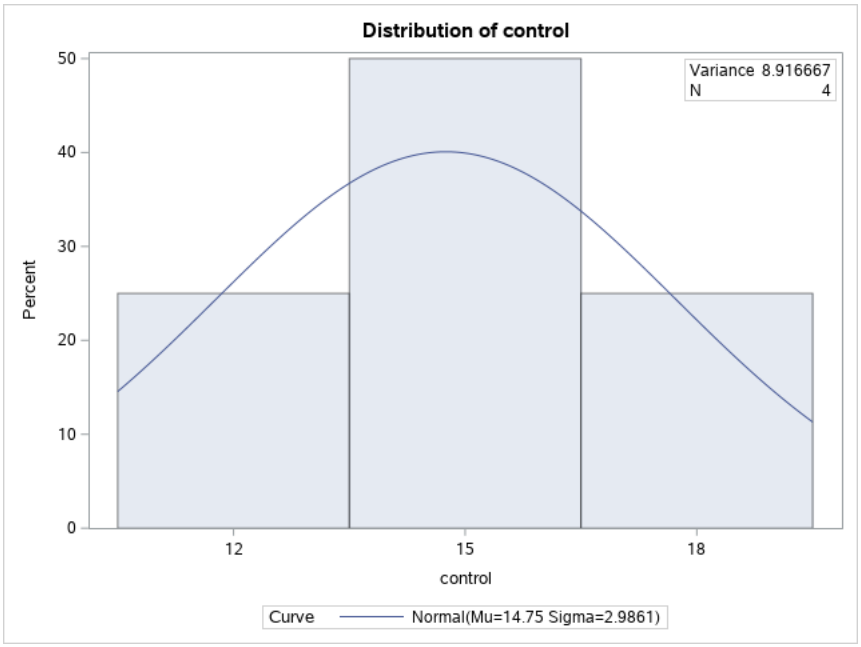
2. Check for Normality (show output: Normal probability plot and boxplot) (3 pts)

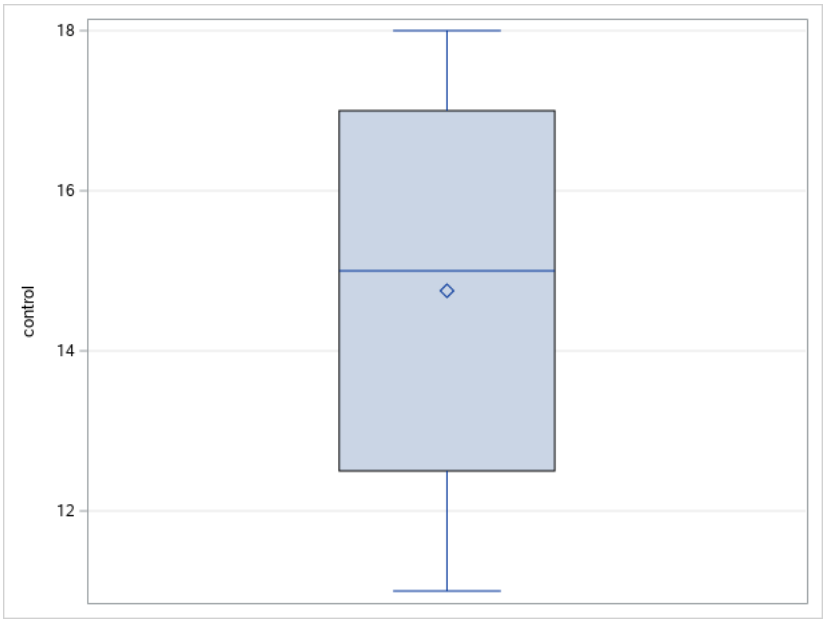



47.5
45.0
42.5
40.0
37.5
35.0
fa


Distribution of fb
Variance 5.666667
N 4
50
40
30
20
10
0
Percent
17.5
20.0
22.5
fb
Curve
Normal(Mu=20.5 Sigma=2.3805)


23
22
21
20
19
18
fb


Distribution of control
Variance 8.916667
N 4
50
40
30
20
10
0
Percent
12
15
18
control
Curve
Normal(Mu=14.75 Sigma=2.9861)



ii. State conclusion in the context of the problem. Just saying you reject or not reject or that the means are not equal, you will receive no points. (including why you made that conclusion based on your SAS output (show output). (3 pts)

Fitted Normal Distribution for fa

Goodness-of-Fit Tests for Normal Distribution				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.30819127	**Pr > D**	>0.150
Cramer-von Mises	**W-Sq**	0.05932455	**Pr > W-Sq**	>0.250
Anderson-Darling	**A-Sq**	0.33478805	**Pr > A-Sq**	>0.250

Fitted Normal Distribution for fb

Goodness-of-Fit Tests for Normal Distribution				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	**D**	0.23569394	**Pr > D**	>0.150
Cramer-von Mises	**W-Sq**	0.04629105	**Pr > W-Sq**	>0.250
Anderson-Darling	**A-Sq**	0.27833812	**Pr > A-Sq**	>0.250

Fitted Normal Distribution for control

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.16224911	Pr > D	>0.150
Cramer-von Mises	W-Sq	0.02347987	Pr > W-Sq	>0.250
Anderson-Darling	A-Sq	0.16795867	Pr > A-Sq	>0.250

I will reject the means are not equal. Firstly, the pre-assumptions are met to do a hypothesis ANOVA. Secondly, the p-value of the test results are all less than 0.05. In this case we can conclude that we have 95% confidence to say that there are differences between those means.

iii. If you reach the conclusion that the means are different, do a Tukey to analyze which means are different. Otherwise say that based on the ANOVA analysis, there was insufficient evidence to say the means are different. (3 pts)

DF	t Value	Pr > |t|
3	5.58	0.0114

DF	t Value	Pr > |t|
3	8.26	0.0037

DF	t Value	Pr > |t|
3	5.19	0.0139

Both these 3 p-values are all smaller than 0.05. Therefore, their means are all significant different from each other.

2. Copy your code here (Remember to put in comments and titles, points will be taken off if you do not do this). (1 code, 1 pts comments (should have at least one comment))

```
data work.ferti;
input
fa 1-2
fb 4-5
control 7-9;
datalines;
38 22 14
47 19 16
39 18 11
35 23 18
42 21 15;

ods noproctitle;
ods graphics / imagemap=on;

/* Exploring Data */
proc univariate data=WORK.FERTI;
     ods select Histogram;
     var fa fb control;
     histogram fa fb control;
run;

proc univariate data=WORK.FERTI;
     ods select Histogram GoodnessOfFit ProbPlot QQPlot;
     var fa fb control;

     /* Checking for Normality */
     histogram fa fb control / normal(mu=est sigma=est);
     inset var n / position=ne;
     probplot fa fb control / normal(mu=est sigma=est);
     inset var n / position=nw;
     qqplot fa fb control / normal(mu=est sigma=est);
     inset var n / position=nw;
run;

proc sgplot data=WORK.FERTI;
```

3. Paste your output. (1 pts)

Data

Obs	fa	fb	control
1	38	22	14
2	47	19	16
3	39	18	11
4	35	23	18

Variable: _Difference_ (Difference: control - fb)

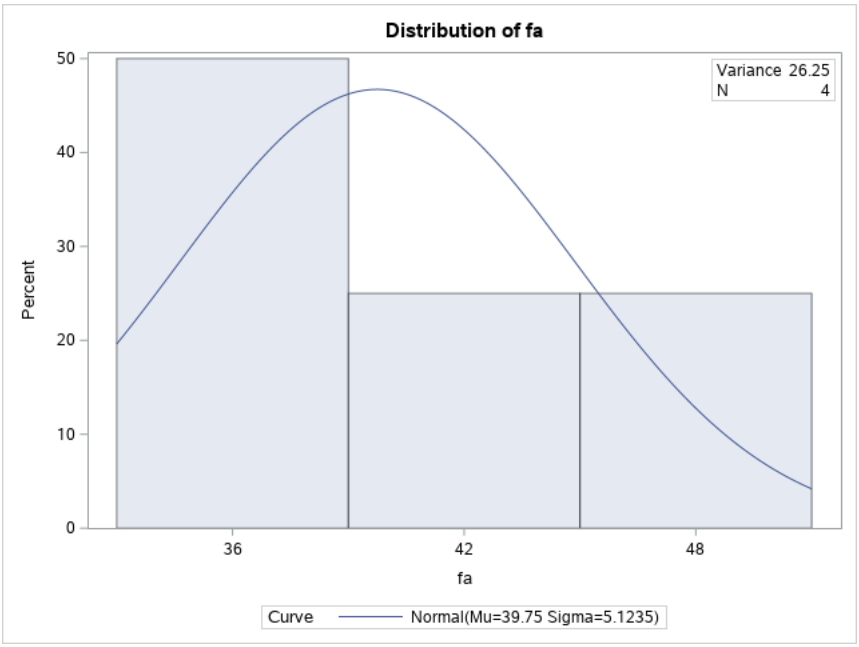
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.963072	Pr < W	0.7982
Kolmogorov-Smirnov	D	0.213533	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.029941	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.200349	Pr > A-Sq	>0.2500

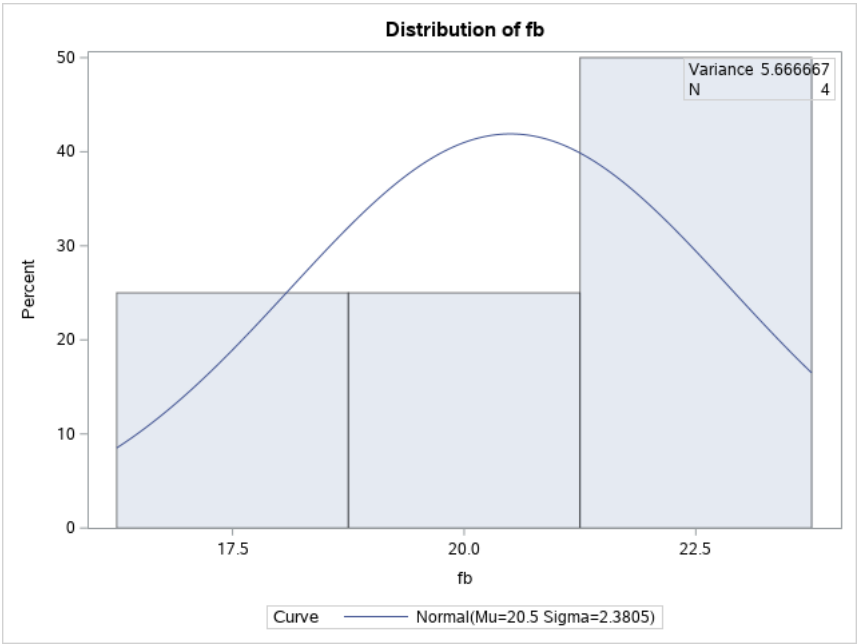
Variable: _Difference_ (Difference: fa - fb)

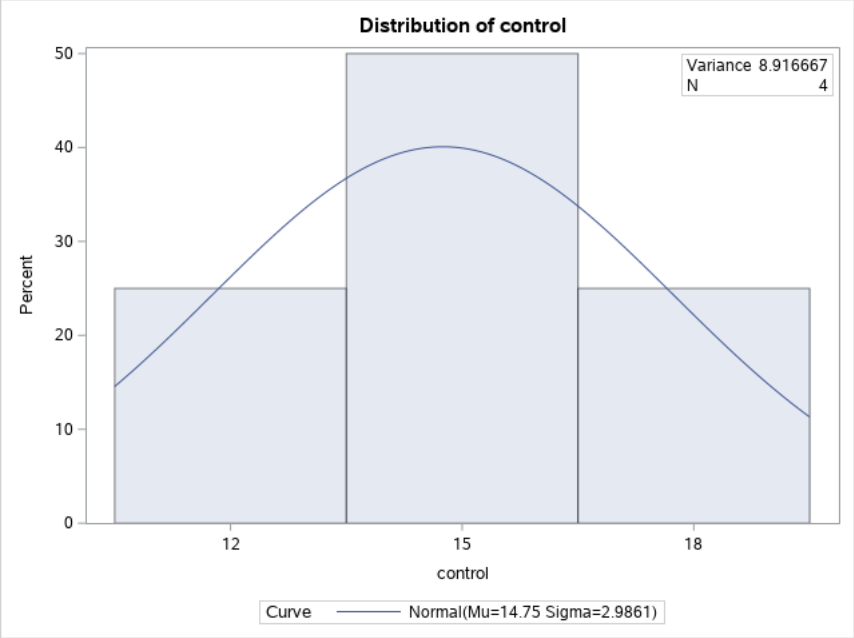
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.980031	Pr < W	0.9022
Kolmogorov-Smirnov	D	0.181232	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.025596	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.179262	Pr > A-Sq	>0.2500

Variable: _Difference_ (Difference: fa - control)

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.96203	Pr < W	0.7917
Kolmogorov-Smirnov	D	0.189853	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.030866	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.207978	Pr > A-Sq	>0.2500


Distribution of fa
Variance 26.25
N 4
Percent
50
40
30
20
10
0
36
42
48
fa
Curve
Normal(Mu=39.75 Sigma=5.1235)


Distribution of fb
Variance 5.666667
N 4
Percent
50
40
30
20
10
0
17.5
20.0
22.5
fb
Curve
Normal(Mu=20.5 Sigma=2.3805)



Variable: _Difference_ (Difference: control - fb)

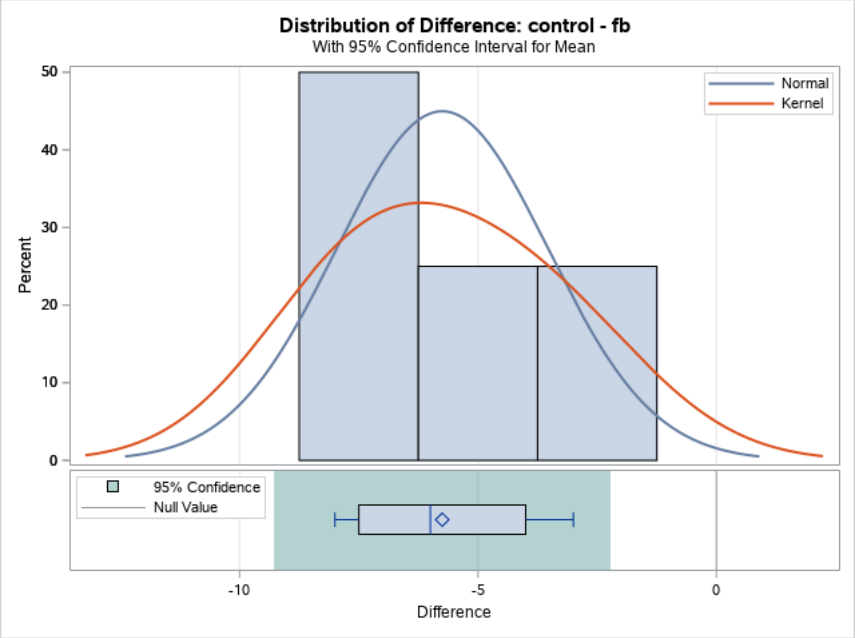
Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.963072	**Pr < W**	0.7982
Kolmogorov-Smirnov	**D**	0.213533	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.029941	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.200349	**Pr > A-Sq**	>0.2500

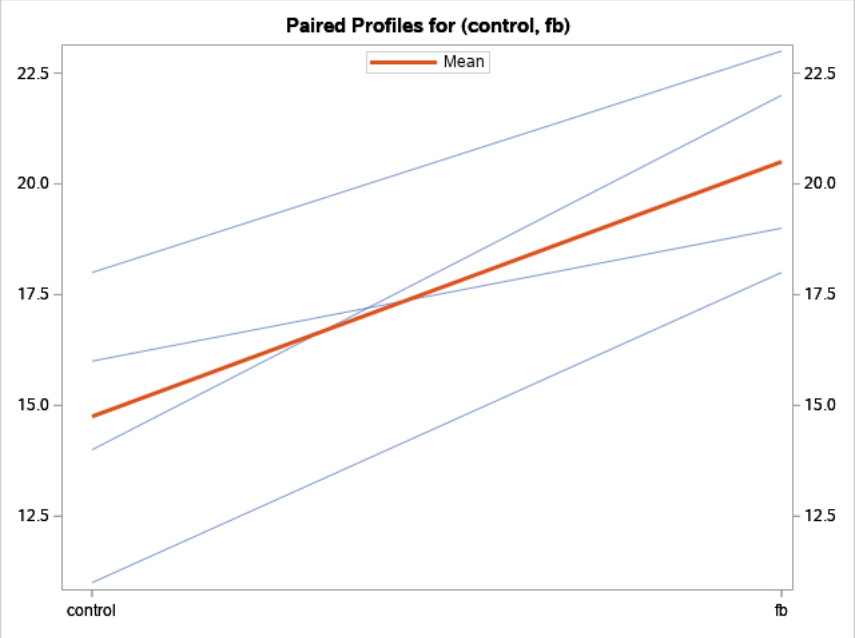
Difference: control - fb

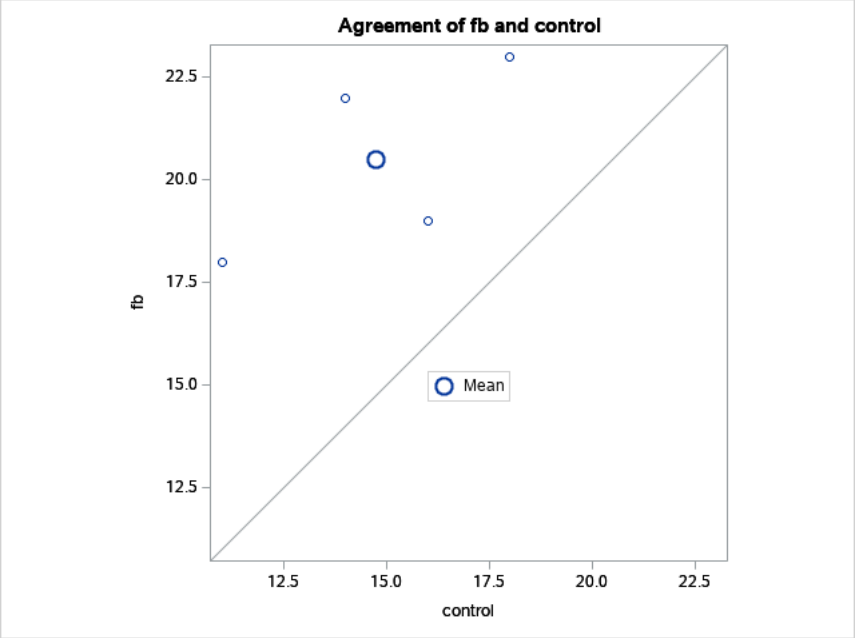
N	Mean	Std Dev	Std Err	Minimum	Maximum
4	-5.7500	2.2174	1.1087	-8.0000	-3.0000

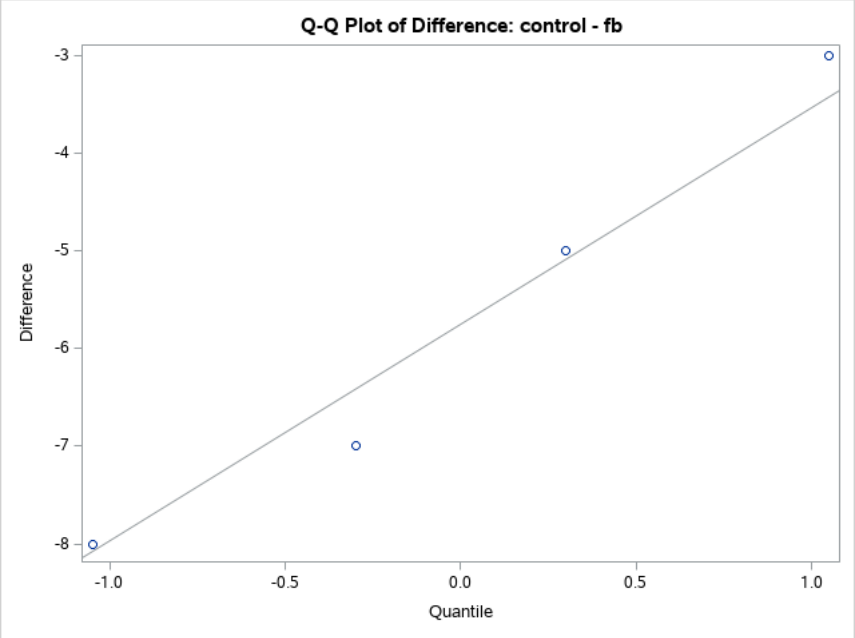
Mean	95% CL Mean		Std Dev	95% CL Std Dev	
-5.7500	-9.2783	-2.2217	2.2174	1.2561	8.2675

DF	t Value	Pr > |t|
3	-5.19	0.0139


Distribution of Difference: control - fb
With 95% Confidence Interval for Mean
Normal
Kernel
Percent
50
40
30
20
10
0
95% Confidence
Null Value
-10
-5
0
Difference


Paired Profiles for (control, fb)
Mean
22.5
20.0
17.5
15.0
12.5
control
fb


Agreement of fb and control
22.5
20.0
17.5
15.0
12.5
fb
Mean
12.5
15.0
17.5
20.0
22.5
control


Q-Q Plot of Difference: control - fb
-3
-4
-5
-6
-7
-8
Difference
-1.0
-0.5
0.0
0.5
1.0
Quantile

PROBLEM 2:

A random sample of 5 scrubs were measured. Their heights were recorded below.

Scrub Height
38
47
39
35
42

If your **FIRST** name ends with A -K Do number 1 (Sue ends in "e", I would do Number 1)
If your **FIRST** name ends with L- Q Do number 2
If your **FIRST** name ends with R-Z Do number 3
(example: My first name is Sue, so I would do number 1)
Part A is worth 3points
Part B is worth 3 points
Part C is worth 3 points
Upload Code with comments 2points
SAS Outpoint 1 points

1. (People whose first name end in A-K do this problem): Test to see if the mean height of the scrub has increased from 36.9 at a 0.05 significance level.
    a. Check to see if all assumptions are met.
    b. State your conclusion in the context of the problem. No points given if you do not answer in the context of the problem.
    c. Upload code and results.
2. (People whose first name end in L-Q do this problem): Test to see if the mean height of the scrub has decreased from 44.6 at a 0.05 significance level.
    a. Check to see if all assumptions are met.
    b. State your conclusion in the context of the problem. No points given if you do not answer in the context of the problem.
    c. Upload code and results.
3. (People whose first name end in R-Z do this problem):Test to see if the mean height of the scrub has changed from 37.2 at a 0.05 significance level.
    a. Check to see if all assumptions are met.
    Yes, they are met. According to the following table, the distribution of the data is significant enough to be normal distribution.

Variable: sh

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.898884	Pr < W	0.4255
Kolmogorov-Smirnov	D	0.308191	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.059325	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.334788	Pr > A-Sq	>0.2500

b. State your conclusion in the context of the problem. No points given if you do not answer in the context of the problem.

DF	t Value	Pr > |t|
3	1.00	0.3929

According to the p-value, which is 0.3929 > 0.05, I cannot reject the Null Hypothesis. Therefore, the mean height of the scrub has not changed from 37.2.

c. Upload code and results.

```
data work.scrub_height;
input
sh 1-2;
datalines;
38
47
39
35
42;

ods noproctitle;
ods graphics / imagemap=on;

proc reg data=WORK.SLEEP alpha=0.05 plots(only)=(diagnostics residuals
                rstudentbypredicted dffits fitplot observedbypredicted);
        model react=sleep /;
        run;
quit;
```

Results:

data

Obs	sh
1	38
2	47
3	39
4	35

Variable: sh

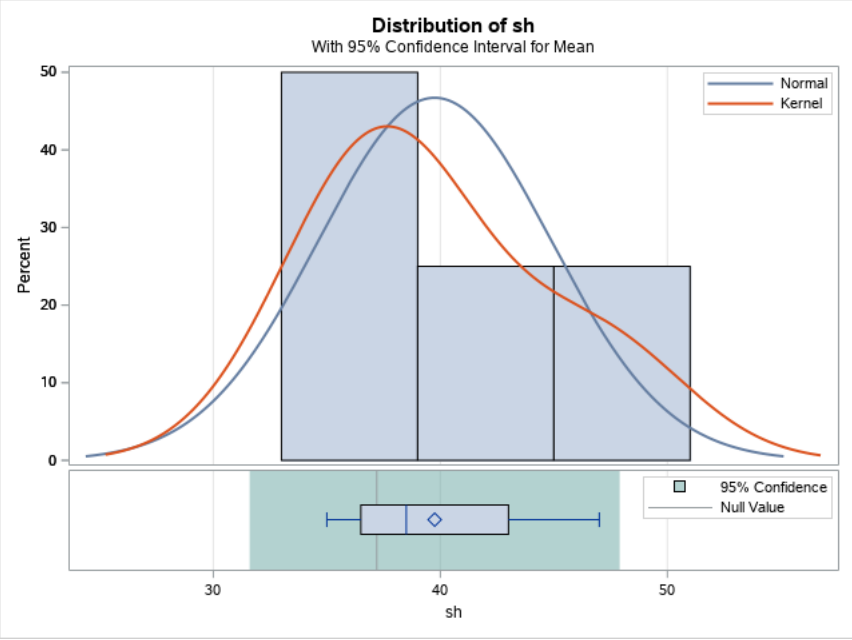
Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	W	0.898884	Pr < W	0.4255
Kolmogorov-Smirnov	D	0.308191	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.059325	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.334788	Pr > A-Sq	>0.2500

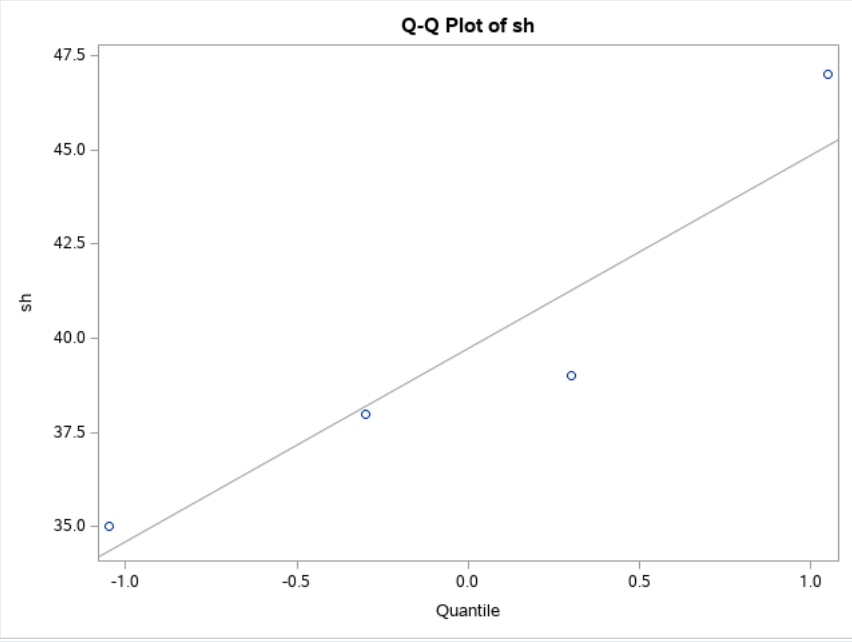
Variable: sh

N	Mean	Std Dev	Std Err	Minimum	Maximum
4	39.7500	5.1235	2.5617	35.0000	47.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
39.7500	31.5974	47.9026	5.1235	2.9024	19.1031

DF	t Value	Pr > |t|
3	1.00	0.3929





PROBLEM 3:

If the letter of your LAST name begins with A-M do the following input procedure plus as a seed for the number generator use the first 4 letters of your last name using the following code to find your unique seed (**xxxxxxxx**). The number in the seed will vary in length depending on the first four letters of your last name:
A-1, B-2, C-3, D-4,E-5, F-6, G-7,H-8,I-9, J-10, K-11, L-12, M-13, N-14, O-15, P-16, Q-17, R-18, S-19, T-20 U-21, V-22, W-23, X-24, Y-25, Z-26 (if your name is less than 4 letters just use as many letters as is in your last name.)

Example: So for my last name Chimiak, I would put 38913 for the seed part of this input statement RANNOR(38913).

```
DATA FINAL;

   DO STRAIN = 'A','B','C';
      DO I = 1 to 7 ;
            SPEED = RANNOR(XXXXXXXX)*5 + 20 + 4*(Strain EQ 'B')+ 5*(Strain EQ 'C') -
2*(STRAIN EQ 'A');
            React = (INT(RANNOR(XXXXXXXX)*5 + 10 + 2*(Strain EQ 'B')+ 3*(Strain EQ
'C') - (STRAIN EQ 'A')));
            Sleep = 12 - ((STRAIN EQ 'A')+ 3.5*(Strain EQ 'C')+ 2*(Strain EQ
'B'))*rannor(XXXXXXXX);
            OUTPUT;
         END;
      END;
   DROP I;
RUN;
```

1. Do a regression analysis that predicts speed (in minutes) using number of hours of sleep.
    a. Write you findings making sure you speak to all pertinent information and assumptions.
        - The p-value to the Analysis of Variance table from the PROC REG.(3pts)

- The R^2 (3pts)
- The p-value to the intercept and slope (2pts each)
- The residual analysis (iid random variable)
  - speaking to the plot predictor vs residual (3pts)
  - speaking to the normal probability plot (3pts)
  - speaking to the boxplot (3pts) PLOTS = DIAGNOSTICS(STATS=NONE);
- Answer the question, is this equation reliable, that is, should it be used or not based on the above information. (3pts)

b. Copy your code here in part (a) of number 1 (Remember to put in comments and titles). (1 code, 1 pts comments (should have at least one comment)

c. Paste your output. (1pts)

2. Give an electric signed statement: I affirm that I will not receive help from ANYONE on this project. This project is my work and my work only. If there is any indication that I have received help, I understand that I will receive a zero for my Final Exam. (1point)

<u>If the letter of your LAST name begins with N-Z</u> do the following input procedure plus as a seed for the number generator use the first 4 letters of your last name using the following code to find your unique seed (**xxxxxxxx**). The number in the seed will vary in length depending on the first four letters of your last name:
A-1, B-2, C-3, D-4,E-5, F-6, G-7,H-8,I-9, J-10, K-11, L-12, M-13, N-14, O-15, P-16, Q-17, R-18, S-19, T-20 U-21, V-22, W-23, X-24, Y-25, Z-26 (if your name is less than 4 letters just use as many letters as is in your last name.)

Example: So for my last name Chimiak, I would put 38913 for the seed part of this input statement RANNOR(38913).

GOU(71521)

```
DATA FINAL;

   DO STRAIN = 'A','B','C';
      DO I = 1 to 7 ;
            SPEED = RANNOR(XXXXXXXX)*5 + 20 + 4*(Strain EQ 'B')+ 5*(Strain EQ 'C') -
2*(STRAIN EQ 'A');
            React = (INT(RANNOR(XXXXXXXX)*5 + 10 + 2*(Strain EQ 'B')+ 3*(Strain EQ
'C') - (STRAIN EQ 'A')));
            Sleep = 12 - ((STRAIN EQ 'A')+ 3.5*(Strain EQ 'C')+ 2*(Strain EQ
'B'))*rannor(XXXXXXXX);
            OUTPUT;
         END;
```

```
    END;
  DROP I;
RUN;
```

1. Do a regression analysis that predicts Reaction time (React) (in minutes) using number of hours of sleep.

    a. Write you findings making sure you speak to all pertinent information and assumptions.

- The p-value to the Analysis of Variance table from the PROC REG.(3pts)

  0.6309

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	8.15610	8.15610	0.24	0.6309
Error	19	649.65342	34.19229		
Corrected Total	20	657.80952			

- The R^2 (3pts)

  0.0124

Root MSE	5.84742	R-Square	0.0124
Dependent Mean	9.90476	Adj R-Sq	-0.0396
Coeff Var	59.03642		

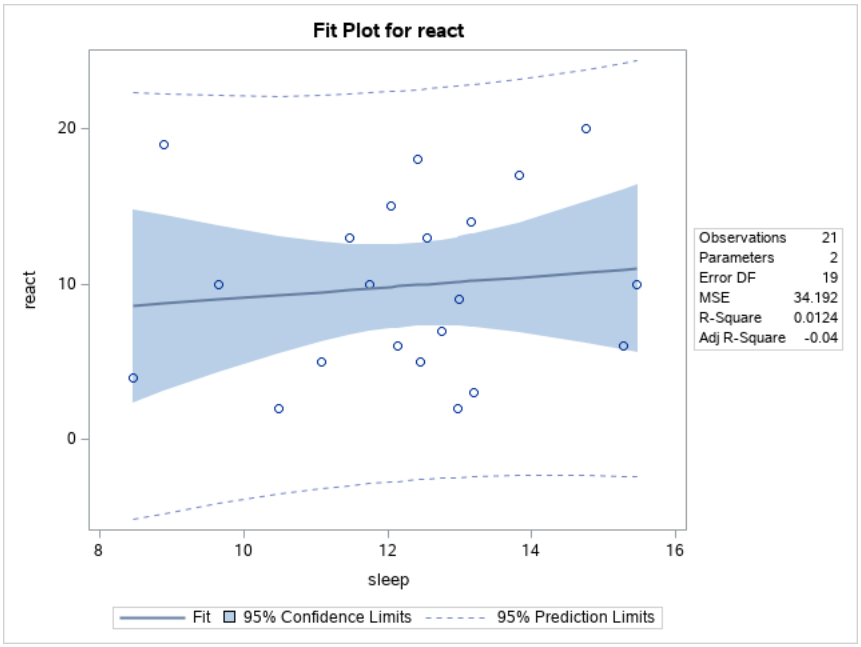
- The p-value to the intercept and slope (2pts each)

  p-value to intercept: 0.5234.

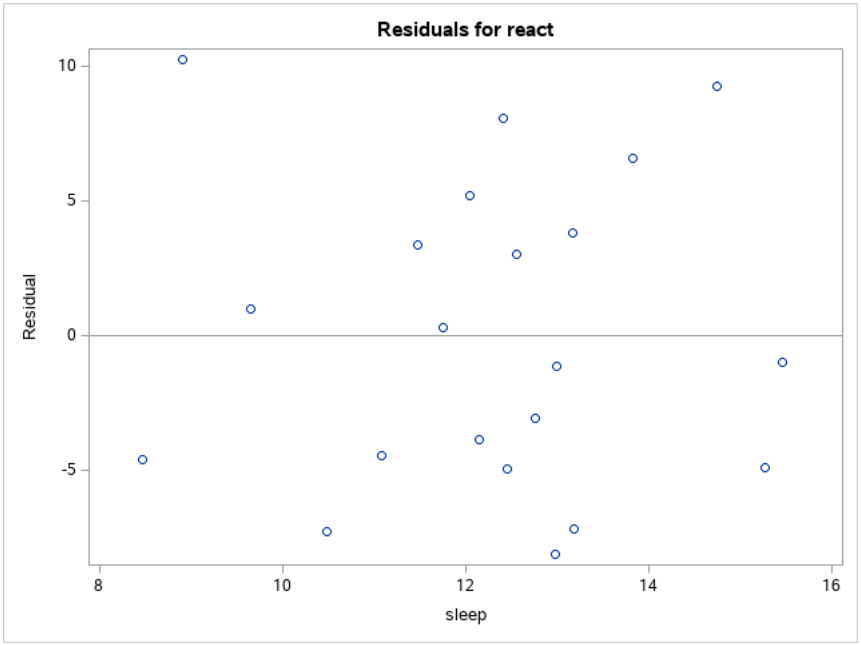
  p-value to slope: 0.6309.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	5.68192	8.73990	0.65	0.5234
sleep	1	0.34400	0.70434	0.49	0.6309

- The residual analysis (iid random variable)
  - speaking to the plot predictor vs residual (3pts)

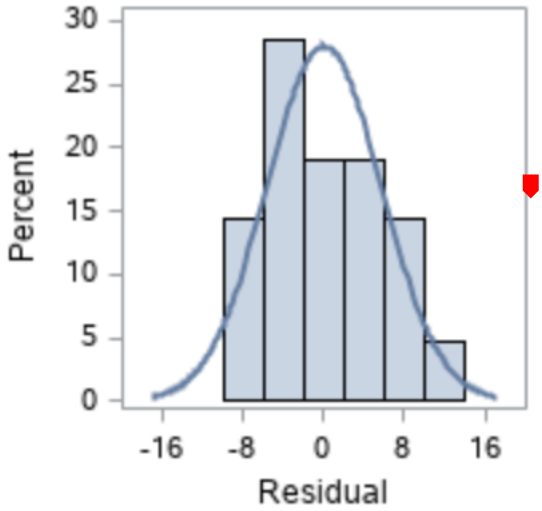


According to the predictor plot, we could conclude that the regression model fits data points badly. Most of data are away from the regression line.



The residuals plot shows that the residuals distribute randomly. And the they vary a lot. The variation of them must be large.

- speaking to the normal probability plot (3pts)

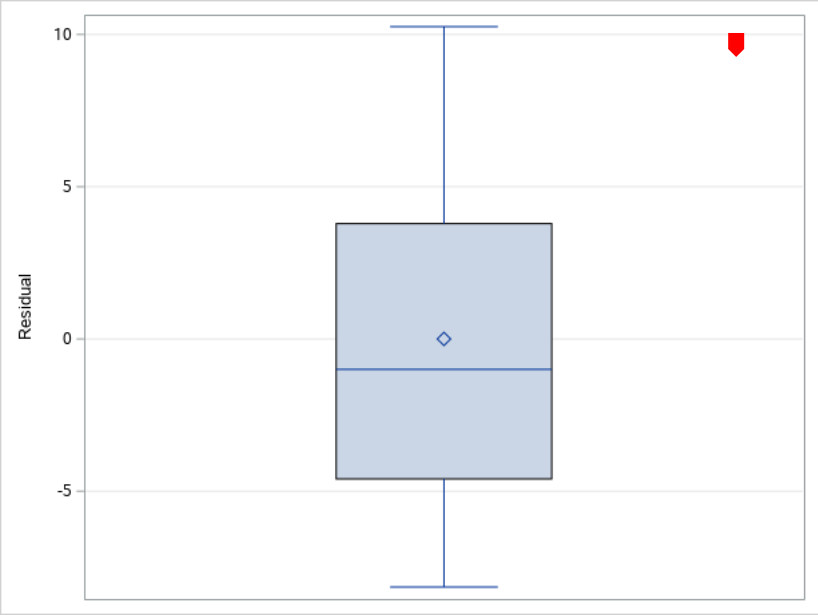


The residuals are not following a normal distribution. The shape of it is asymmertric.

- speaking to the boxplot (3pts) PLOTS = DIAGNOSTICS(STATS=NONE);
-

- Answer the question, is this equation reliable, that is, should it be used or not based on the above information. (3pts)

The following boxplot also shows that the distribution of residuals is skewed. Therefore, it is not normal distribution.



b. Copy your code here in part (a) of number 1 (Remember to put in comments and titles).
(1 code, 1 pts comments (should have at least one comment)

```
data work.sleep;
   do strain = 'A','B','C';
        Do I = 1 to 7;
             speed = rannor(71521)*5+20+4*(Strain EQ 'B')
+5*(Strain EQ 'C')-2*(Strain EQ 'A');
             react = (INT(rannor(71521)*5+10+2*(Strain EQ 'B')
+3*(Strain EQ 'C')-(Strain EQ 'A')));
             sleep = 12-(2*(Strain EQ 'B')+3.5*(Strain EQ 'C')+
(Strain EQ 'A'))*rannor(71521);
             output;
        end;
   end;
   drop I;
run;

ods noproctitle;
ods graphics / imagemap=on;

proc reg data=WORK.SLEEP alpha=0.05 plots(only)=(diagnostics
residuals
        rstudentbypredicted dffits fitplot
observedbypredicted);
   model react=sleep /;
   output out=work.Reg_stats dffits=dffits_ p=p_ lcl=lcl_
ucl=ucl_ lclm=lclm_
        uclm=uclm_ press=press_ r=r_ student=student_
rstudent=rstudent_;
   run;
quit;

ods graphics / reset width=6.4in height=4.8in imagemap;

proc sgplot data=WORK.REG_STATS;
   vbox r_ /;
   yaxis grid;
run;

ods graphics / reset;
```

c. Paste your output. (1pts)

Data:

Obs	strain	speed	react	sleep
1	A	17.0943	6	12.1414
2	A	22.7370	15	12.0487
3	A	21.8388	3	13.1883
4	A	26.6301	7	12.7553
5	A	17.2422	5	12.4493
6	A	20.4957	2	12.9783
7	A	15.0054	13	12.5518
8	B	20.0152	17	13.8180
9	B	24.2162	19	8.9002
10	B	24.1902	10	11.7507
11	B	27.7797	5	11.0759
12	B	24.3382	2	10.4851
13	B	22.4696	9	12.9939
14	B	12.0173	13	11.4791
15	C	24.7574	4	8.4728
16	C	27.1539	6	15.2700
17	C	25.2521	20	14.7458
18	C	28.2565	10	15.4574
19	C	28.2175	14	13.1667
20	C	27.4015	18	12.4077
21	C	15.5180	10	9.6523

Model: MODEL1
Dependent Variable: react

Number of Observations Read	21
Number of Observations Used	21

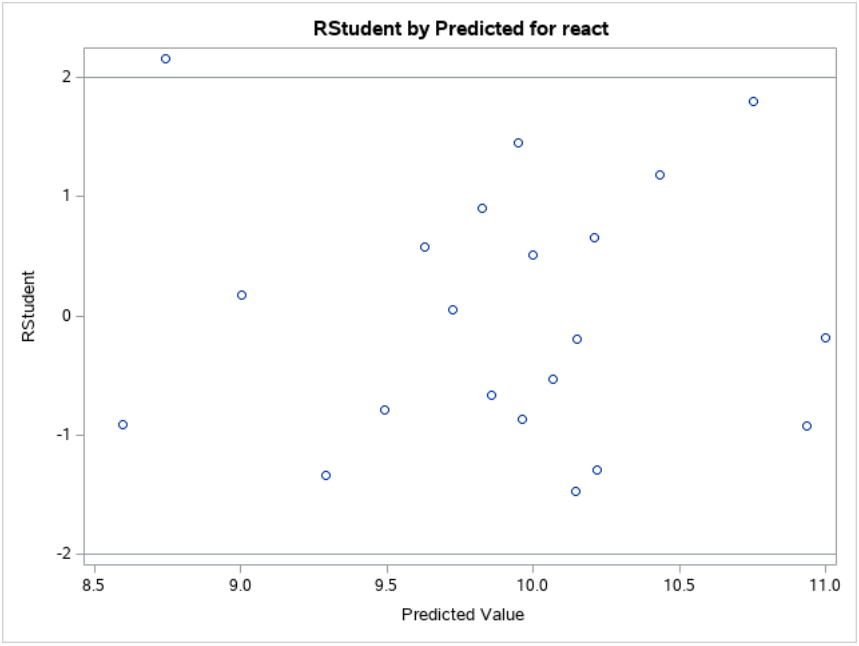
Analysis of Variance

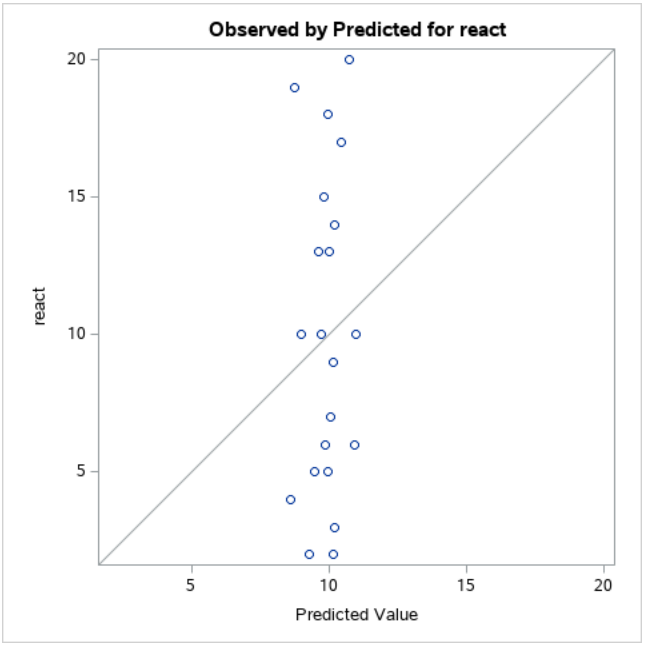
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	8.15610	8.15610	0.24	0.6309
Error	19	649.65342	34.19229		
Corrected Total	20	657.80952			

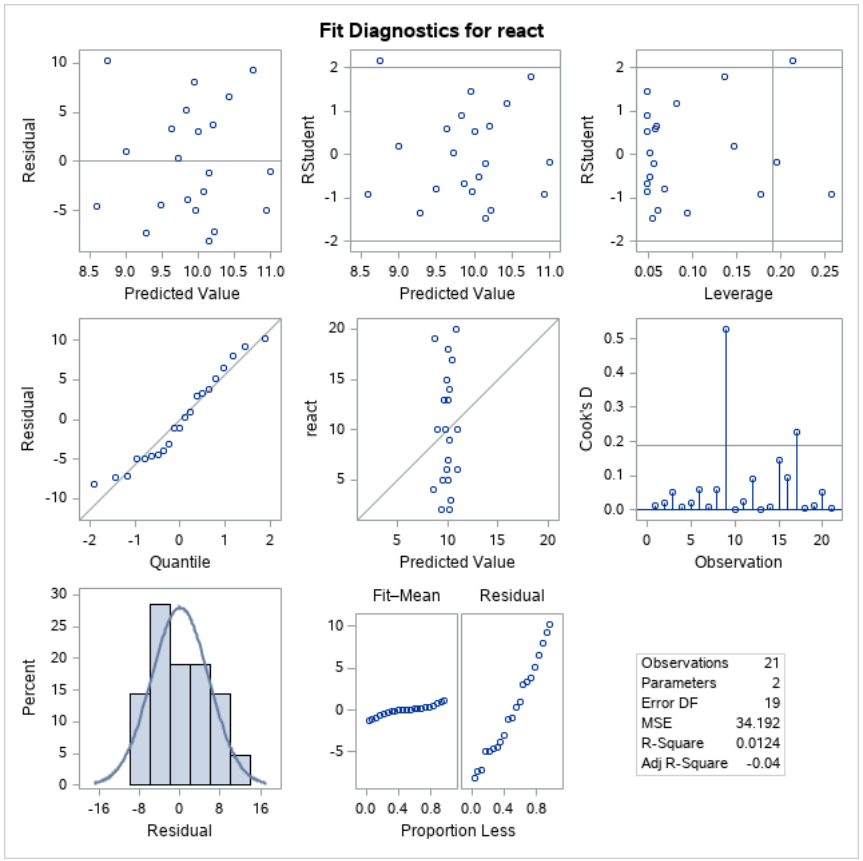
Root MSE	5.84742	R-Square	0.0124
Dependent Mean	9.90476	Adj R-Sq	-0.0396
Coeff Var	59.03642		

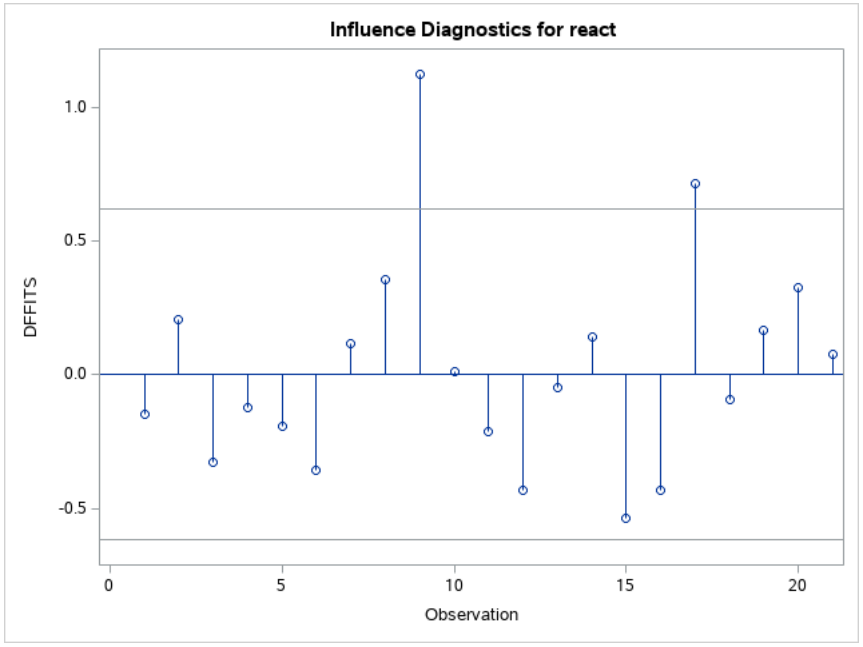
Parameter Estimates

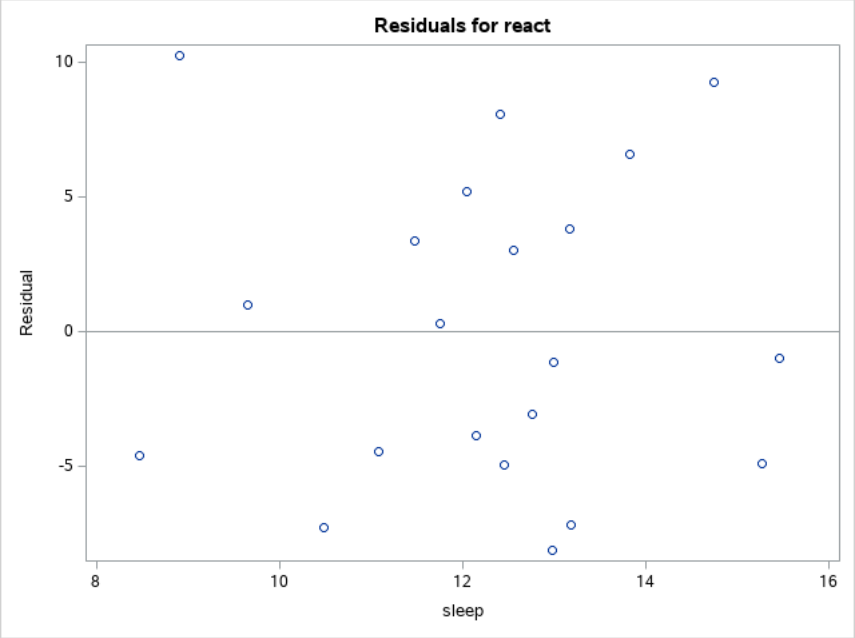
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	5.68192	8.73990	0.65	0.5234
sleep	1	0.34400	0.70434	0.49	0.6309

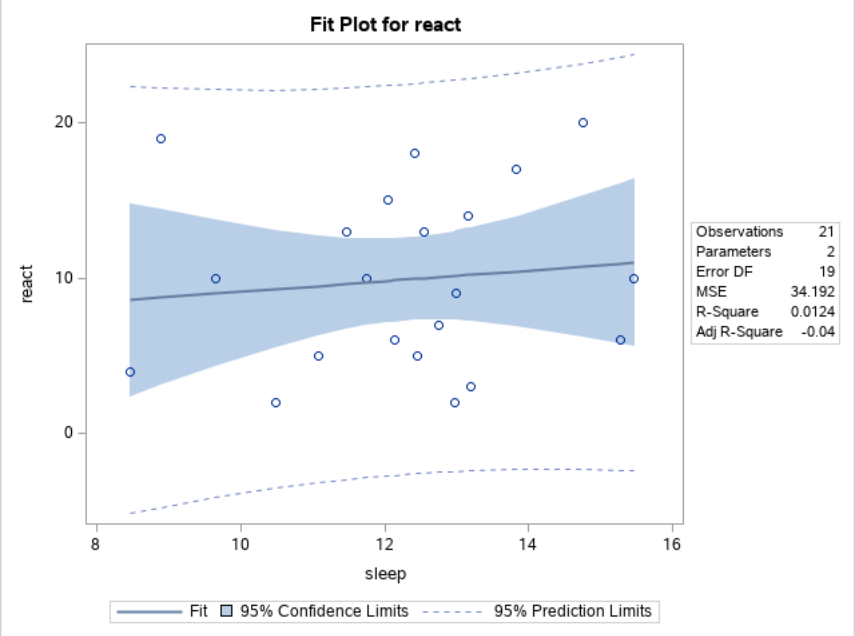

RStudent by Predicted for react
RStudent
2
1
0
-1
-2
8.5
9.0
9.5
10.0
10.5
11.0
Predicted Value

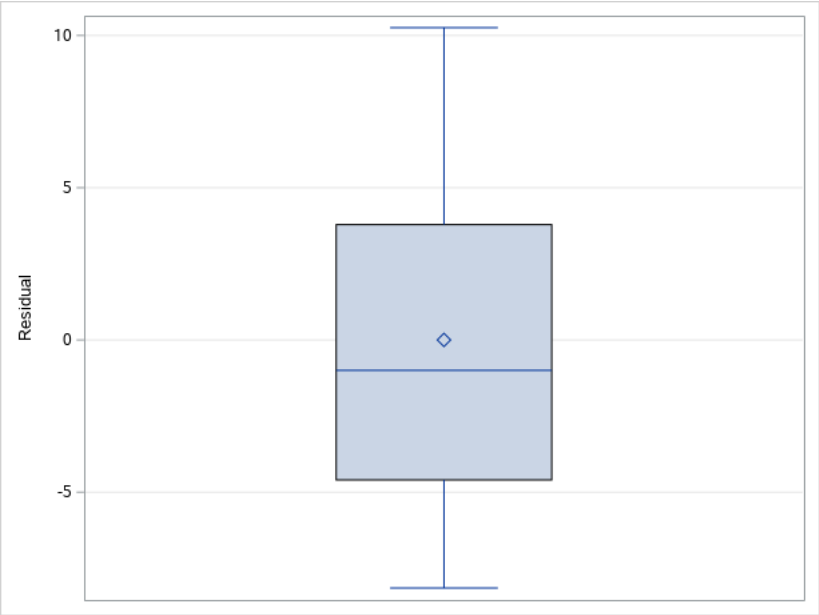

Observed by Predicted for react
react
20
15
10
5
5
10
15
20
Predicted Value


Fit Diagnostics for react
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
react
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 21
Parameters 2
Error DF 19
MSE 34.192
R-Square 0.0124
Adj R-Square -0.04


Influence Diagnostics for react
DFFITS
Observation


Residuals for react
Residual
sleep
10
5
0
-5
8
10
12
14
16


Fit Plot for react
react
sleep
20
10
0
8
10
12
14
16
Observations 21
Parameters 2
Error DF 19
MSE 34.192
R-Square 0.0124
Adj R-Square -0.04
Fit
95% Confidence Limits
95% Prediction Limits



2. Give an electric signed statement: I affirm that I will not receive help from ANYONE on this project. This project is my work and my work only. If there is any indication that I have received help, I understand that I will receive a zero for my Final Exam. (1point)

   YINGQIAO GOU