MATH 5531 Statistical Methods I HW 5

Russell Land

November 12, 2018

Table 1 shows plasma inorganic phosphate levels (mg/dl) one hour after a standard glucose tolerance test for obese subjects, with or without hyperinsulinemia, and controls (data from Jones, 2017).

(a) Perform a one-factor analysis of variance to test the hypotheses that there are no mean differences among the three groups. What conclusions can you draw?

(b) Obtain a 95% confidence interval for the differences in means between the two obese groups.

Table 1: *Plasma Phosphate Levels in Obese and Control Subjects*

Hyperinsulinemic Obese	Non-hyperinsulinemic Obese	Control
2.3	3.0	3.0
4.1	4.1	2.6
4.2	3.9	3.1
4.6	3.3	2.1
3.8	3.3	2.8
5.2	3.9	3.4
3.1		2.9
3.7		2.6
3.8		3.1
		3.2

(c) Apply Fisher's LSD and Tukey multiple comparison procedures to determine where the differences lie (if you reject the null hypothesis).

(d) Using an appropriate model examine the standardized residuals for all the observations to look for any systematic effects and to check the Normality assumptions.

To answer part (a)–(d) we will be using the SAS Studio Software, University Edition. In this document we will implement a LaTeX package named StatRep. This will allow us to display our SAS code in our TeX document neatly, and will also allow us to compile our SAS results straight into the LaTeX document. Once we present our data below, I will follow up with an analysis of the data. We will begin by creating a data set in the program, and then run the actually procedures for each test. Below shows the data set `PhosphateLevels` being created:

```
title 'Inorganic Phosphate Levels';

data PhosphateLevels;
input IDNumber Levels;
datalines;
1 2.3
1 4.1
1 4.2
```

```
... more data lines ...

3 2.6
3 3.1
3 3.2
;

/* IDNumber 1, 2, and 3 correspond to the following:
1 - Hyperinsulinemic Obese
2 - Non-hyperinsulinemic Obese
3 - Control */

run;
```

Now that our data set is stored as `PhosphateLevels`, we can run the appropriate procedures, including `means`,`anova`, and `reg`. We will use the `means` procedure to analyze the means and variance to determine an appropriate test. Since we will be using an `anova` procedure we have to check if our underlying assumptions hold. That is, we must make sure the samples have a normal distribution, have equal variance, and are independent random samples. Below is the SAS code for our analyses:

```
proc means;
by IDNumber;

proc anova;
class IDNumber;
model Levels = IDNumber;
means IDNumber/ lsd tukey;

proc reg;
id IDNumber;
model Levels=IDNumber/ cli;

proc print;
run;
```

Figure 1: Statistics for ANOVA, LSD, Tukey, and Residual Analysis

Inorganic Phosphate Levels

The MEANS Procedure

IDNumber=1

Analysis Variable : Levels				
N	**Mean**	**Std Dev**	**Minimum**	**Maximum**
11	3.9454545	0.7776421	2.3000000	5.2000000

IDNumber=2

Analysis Variable : Levels				
N	**Mean**	**Std Dev**	**Minimum**	**Maximum**
8	3.4375000	0.4627171	2.9000000	4.1000000

IDNumber=3

Analysis Variable : Levels				
N	**Mean**	**Std Dev**	**Minimum**	**Maximum**
12	2.7833333	0.4086193	2.1000000	3.4000000

Figure 1: *continued*

Inorganic Phosphate Levels

The ANOVA Procedure

Class Level Information		
Class	**Levels**	**Values**
IDNumber	3	1 2 3

Number of Observations Read	31
Number of Observations Used	31

Inorganic Phosphate Levels

The ANOVA Procedure

Dependent Variable: Levels

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	7.80827835	3.90413917	11.65	0.0002
Error	28	9.38268939	0.33509605		
Corrected Total	30	17.19096774			

R-Square	Coeff Var	Root MSE	Levels Mean
0.454208	17.20529	0.578875	3.364516

Source	DF	Anova SS	Mean Square	F Value	Pr > F
IDNumber	2	7.80827835	3.90413917	11.65	0.0002

Inorganic Phosphate Levels

The ANOVA Procedure

t Tests (LSD) for Levels

note	This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	28
Error Mean Square	0.335096
Critical Value of t	2.04841

Comparisons significant at the 0.05 level are indicated by ***.				
IDNumber Comparison	**Difference Between Means**	**95% Confidence Limits**		
1 - 2	0.5080	−0.0430	1.0589	
1 - 3	1.1621	0.6672	1.6571	***
2 - 1	−0.5080	−1.0589	0.0430	
2 - 3	0.6542	0.1129	1.1954	***
3 - 1	−1.1621	−1.6571	−0.6672	***
3 - 2	−0.6542	−1.1954	−0.1129	***

Figure 1: *continued*

Inorganic Phosphate Levels

The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for Levels

note	This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	28
Error Mean Square	0.335096
Critical Value of Studentized Range	3.49918

Comparisons significant at the 0.05 level are indicated by ***.				
IDNumber Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
1 - 2	0.5080	−0.1576	1.1735	
1 - 3	1.1621	0.5642	1.7600	***
2 - 1	−0.5080	−1.1735	0.1576	
2 - 3	0.6542	0.0004	1.3079	***
3 - 1	−1.1621	−1.7600	−0.5642	***
3 - 2	−0.6542	−1.3079	−0.0004	***

Inorganic Phosphate Levels

The REG Procedure

Model: MODEL1

Dependent Variable: Levels

Number of Observations Read	31
Number of Observations Used	31

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	7.77657	7.77657	23.95	<.0001
Error	29	9.41440	0.32463		
Corrected Total	30	17.19097			

Root MSE	0.56977	**R-Square**	0.4524
Dependent Mean	3.36452	**Adj R-Sq**	0.4335
Coeff Var	16.93459		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	4.54705	0.26239	17.33	<.0001
IDNumber	1	−0.58188	0.11889	−4.89	<.0001

Figure 1: *continued*

Inorganic Phosphate Levels

The REG Procedure

Model: MODEL1

Dependent Variable: Levels

Output Statistics							
Obs	**IDNumber**	**Dependent Variable**	**Predicted Value**	**Std Error Mean Predict**	**95% CL Predict**		**Residual**
1	1	2.3	3.9652	0.1598	2.7549	5.1754	−1.6652
2	1	4.1	3.9652	0.1598	2.7549	5.1754	0.1348
3	1	4.2	3.9652	0.1598	2.7549	5.1754	0.2348
4	1	4.0	3.9652	0.1598	2.7549	5.1754	0.0348
5	1	4.6	3.9652	0.1598	2.7549	5.1754	0.6348
6	1	4.6	3.9652	0.1598	2.7549	5.1754	0.6348
7	1	3.8	3.9652	0.1598	2.7549	5.1754	−0.1652
8	1	5.2	3.9652	0.1598	2.7549	5.1754	1.2348
9	1	3.1	3.9652	0.1598	2.7549	5.1754	−0.8652
10	1	3.7	3.9652	0.1598	2.7549	5.1754	−0.2652
11	1	3.8	3.9652	0.1598	2.7549	5.1754	−0.1652
12	2	3.0	3.3833	0.1024	2.1993	4.5673	−0.3833
13	2	4.1	3.3833	0.1024	2.1993	4.5673	0.7167
14	2	3.9	3.3833	0.1024	2.1993	4.5673	0.5167
15	2	3.1	3.3833	0.1024	2.1993	4.5673	−0.2833
16	2	3.3	3.3833	0.1024	2.1993	4.5673	−0.0833
17	2	2.9	3.3833	0.1024	2.1993	4.5673	−0.4833
18	2	3.3	3.3833	0.1024	2.1993	4.5673	−0.0833
19	2	3.9	3.3833	0.1024	2.1993	4.5673	0.5167
20	3	3.0	2.8014	0.1540	1.5943	4.0085	0.1986
21	3	2.6	2.8014	0.1540	1.5943	4.0085	−0.2014
22	3	3.1	2.8014	0.1540	1.5943	4.0085	0.2986
23	3	2.2	2.8014	0.1540	1.5943	4.0085	−0.6014
24	3	2.1	2.8014	0.1540	1.5943	4.0085	−0.7014
25	3	2.4	2.8014	0.1540	1.5943	4.0085	−0.4014
26	3	2.8	2.8014	0.1540	1.5943	4.0085	−0.001404
27	3	3.4	2.8014	0.1540	1.5943	4.0085	0.5986
28	3	2.9	2.8014	0.1540	1.5943	4.0085	0.0986
29	3	2.6	2.8014	0.1540	1.5943	4.0085	−0.2014
30	3	3.1	2.8014	0.1540	1.5943	4.0085	0.2986
31	3	3.2	2.8014	0.1540	1.5943	4.0085	0.3986

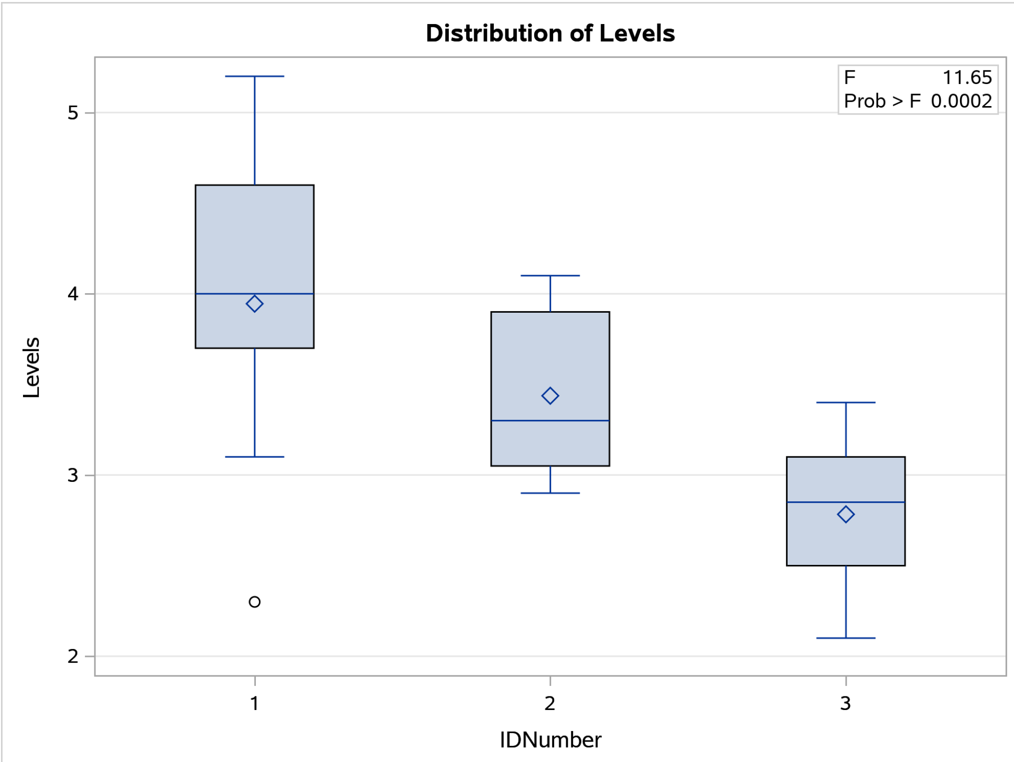
Figure 1: *continued*

Sum of Residuals	0
Sum of Squared Residuals	9.41440
Predicted Residual SS (PRESS)	10.89647

Inorganic Phosphate Levels

Obs	IDNumber	Levels
1	1	2.3
2	1	4.1
3	1	4.2
4	1	4.0
5	1	4.6
6	1	4.6
7	1	3.8
8	1	5.2
9	1	3.1
10	1	3.7
11	1	3.8
12	2	3.0
13	2	4.1
14	2	3.9
15	2	3.1
16	2	3.3
17	2	2.9
18	2	3.3
19	2	3.9
20	3	3.0
21	3	2.6
22	3	3.1
23	3	2.2
24	3	2.1
25	3	2.4
26	3	2.8
27	3	3.4
28	3	2.9
29	3	2.6
30	3	3.1
31	3	3.2

Figure 2: Comparative Box Plots for Normality



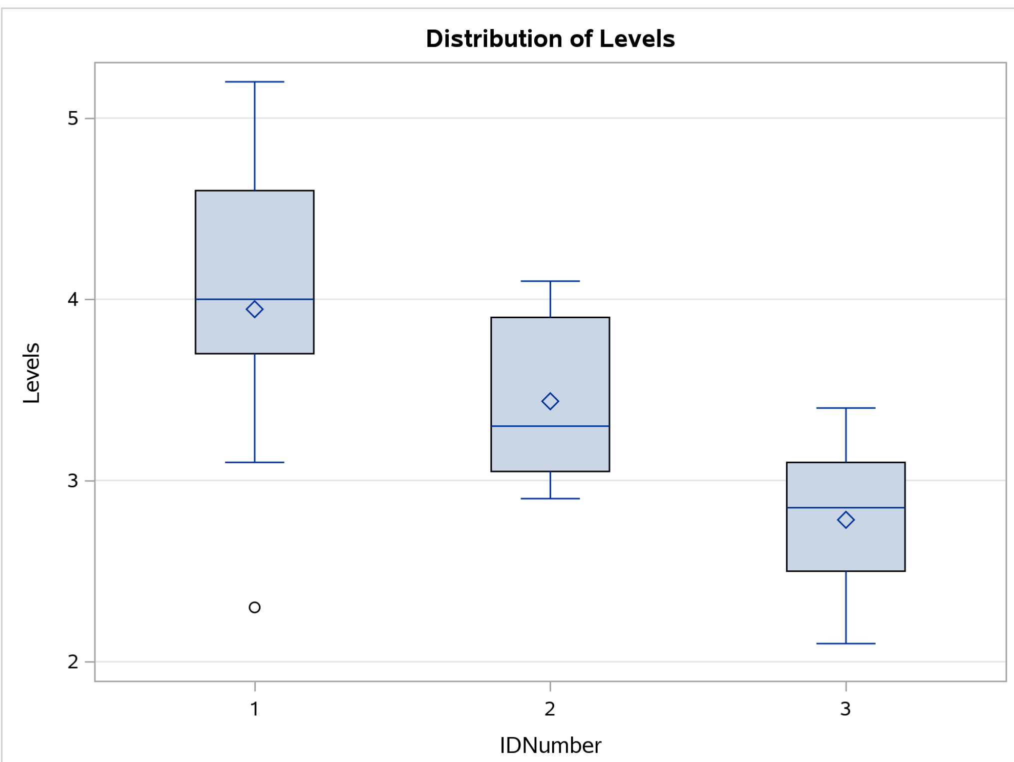


Figure 2: *continued*

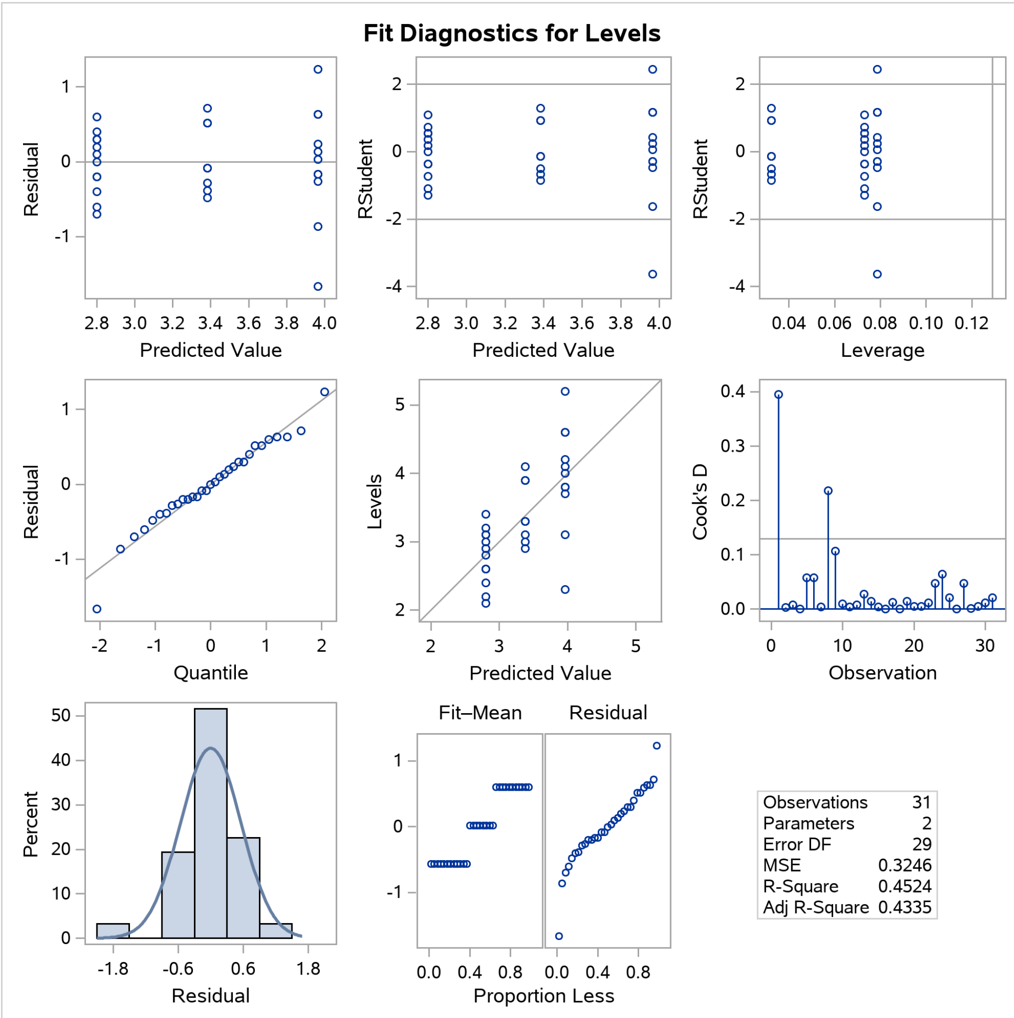
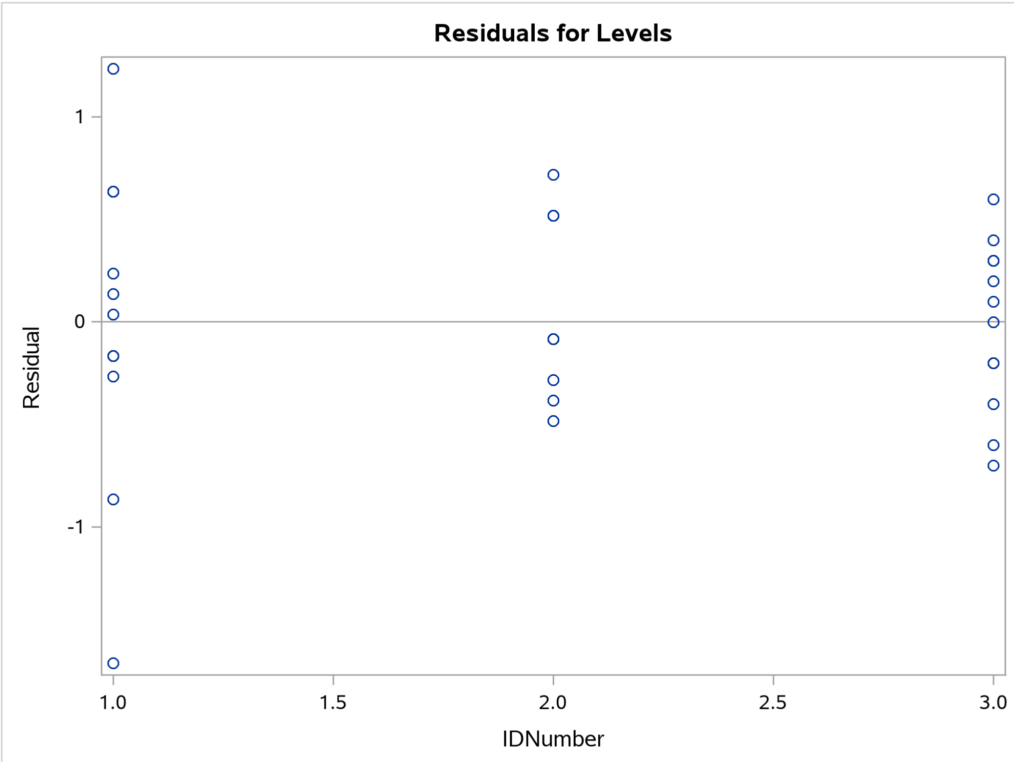
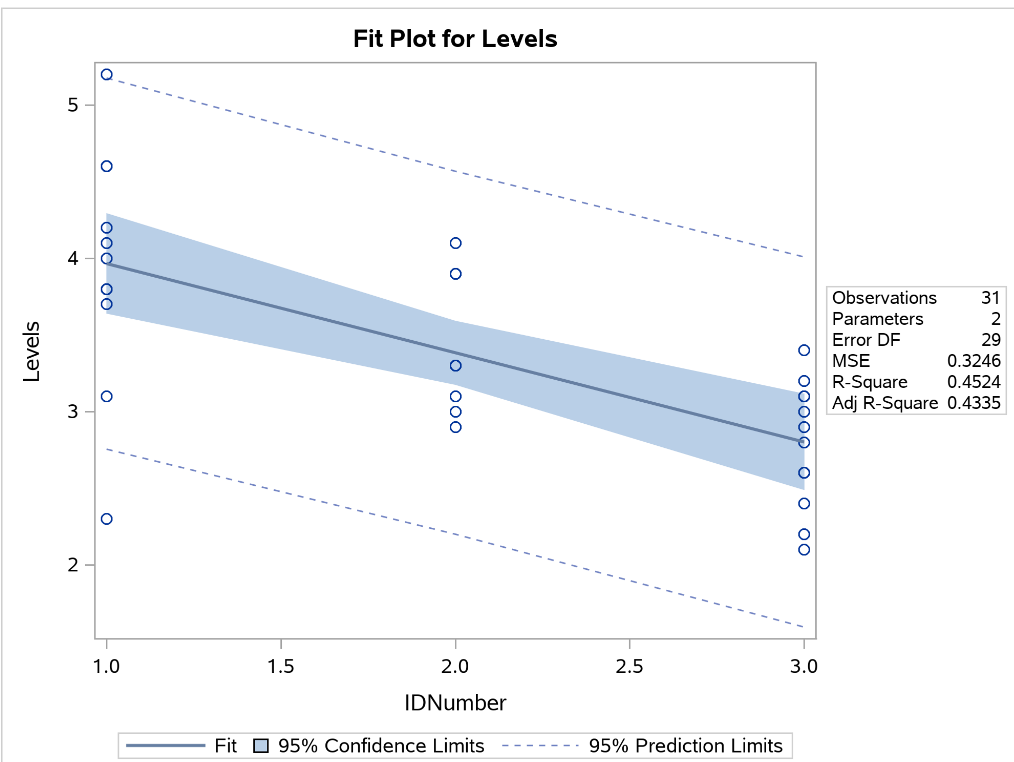


Figure 2: *continued*





(a) We are trying to test the hypothesis that there are no mean differences among the three groups, Hyperinsulinemic Obese, Non-hyperinsulinemic Obese, and Control. We can define our hypotheses:

$H_0 : \mu_1 = \mu_2 = \mu_3$
vs.
H_A : At least one of the three sample means is different from the rest

From our SAS results, we can see that the `anova` procedure ran the test for us and returned a final p-value of 0.0002. Hence, we have significant evidence to reject the null hypothesis. Thus, there is significant evidence that at least one of the sample means differs from the rest.

(b) For our model it is best to use Tukey's simultaneous confidence intervals. Since we have different sample sizes we had to adjust our value of W to:

$$W^* = \frac{q_\alpha(t, \nu)}{\sqrt{2}} \sqrt{s_w^2 \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

However, using the SAS software we did not have to do any computations ourselves. We obtain a 95% confidence interval for the difference in means between the two obese groups to be $(-0.1576, 1.1735)$. Again, this was computed using the adjusted W value and the formula for the simultaneous confidence interval:

$$\left(\bar{y}_i - \bar{y}_j\right) \pm W^*$$

(c) From the data we obtained from running the `lsd` and `tukey` under the `means` options we were able to conclude the following results:

Fisher's Confidence Intervals

Difference in Means	95% C.I. for Difference	Conclusion
$\mu_2 - \mu_1$	(-1.0589,0.0430)	Not Significant
$\mu_3 - \mu_1$	(-1.6571,-0.6672)	Significant
$\mu_3 - \mu_2$	(-1.1954,-0.1129)	Significant

Tukey Confidence Intervals

Difference in Means	95% C.I. for Difference	Conclusion
$\mu_2 - \mu_1$	(-1.1735,0.1576)	Not Significant
$\mu_3 - \mu_1$	(-1.7600,-0.5642)	Significant
$\mu_3 - \mu_2$	(-1.3079,-0.0004)	Significant

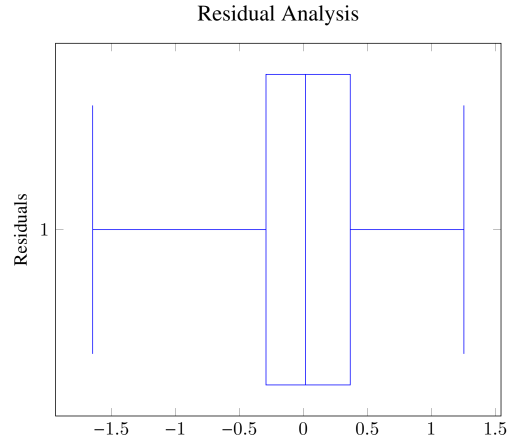
From these results we can see that between both the `lsd` and `tukey` results the difference between μ_2 and μ_1 are not significant enough to reject the null, because 0 is contained in the intervals. However, pairwise with the rest of the results, the data is significant enough to reject the null because 0 is not contained in the intervals of the other mean differences. Thus, we can conclude that when using Fisher's LSD and Tukey's LSD test, the majority of difference lies when means are pairwise with the control group, or μ_3. This leads us to believe that μ_3 may have a significant difference from the rest.

(d) To check our normality assumptions we can always analyze the box-plots that were printed when we ran our `anova` test. However, we can also check using the standardized residuals. These can be computed with the following formula:

$$e_{ij} = y_{ij} - \bar{y}_i$$

After applying this formula we obtain the following results for our data set:

Residuals for Phosphate Levels		
Hyperinsulinemic Obese	Non-hyperinsulinemic Obese	Control
-1.6455	-0.4375	0.2167
0.1545	0.6625	-0.1833
0.2545	0.4625	0.3167
0.0545	-0.3375	-0.5833
0.6545	-0.1375	-0.6833
0.6545	-0.5375	-0.3833
-0.1455	-0.1375	0.0167
1.2545	0.4625	0.6167
-0.8455		0.1167
-0.2455		-0.1833
-0.1455		0.3167
		0.4167



From the resulting table and graph we can conclude that our data is approximately normal. The median is just about the mean and the box-plot has a considerable symmetric shape. Thus, our normality assumptions were correct for the procedures we ran. Furthermore, along with the other SAS procedures we ran, we can add in the `proc reg` procedure with the `cli` option. This will give us similar results as above, but will support our argument more. The corresponding tables and graph will be shown with the rest of the SAS results above.