CSCI 688
Homework 3

Megan Rose Bryant
Department of Mathematics
William and Mary

September 29, 2014

10.6

A study was performed on wear of a bearing y and in relationship to x_1 = oil viscosity and x_2 = load. The following data were obtained:

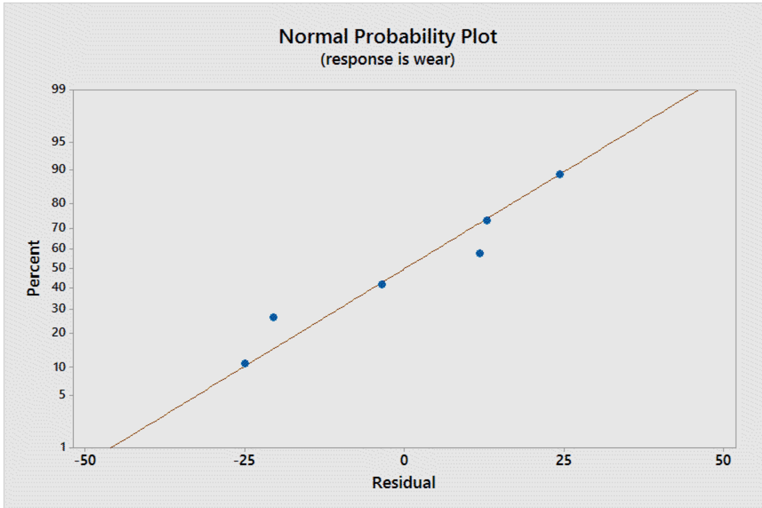
y	x_1	x_2
193	1.6	851
230	15.5	816
172	22.0	1058
91	43.0	1201
113	33.0	1357
125	40.0	1115

a.)Fit a multiple linear regression model to this data.

```
[fontsize=\small]
Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
25.4979  86.18%     76.96%      10.03%


Coefficients

Term          Coef  SE Coef  T-Value  P-Value   VIF
Constant     351.0     74.8     4.70    0.018
viscosity    -1.27     1.17    -1.09    0.356  2.64
load       -0.1539   0.0895    -1.72    0.184  2.64


Regression Equation

wear = 351.0 - 1.27 viscosity - 0.1539 load
```

b.) Test for significance of regression.

First, we will examine the normal probability plot to ensure that there is no reason to doubt the normality assumption. Based on the graph below, we would accept the assumption of normality and proceed to test for significance.



Regression Analysis: wear versus viscosity, load

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	2	12162	86.18%	12161.6	6080.8	9.35	0.051
viscosity	1	10240	72.56%	769.6	769.6	1.18	0.356
load	1	1921	13.61%	1921.2	1921.2	2.96	0.184
Error	3	1950	13.82%	1950.4	650.1		
Total	5	14112	100.00%				

Model Summary

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)
25.4979	86.18%	76.96%	12696.7	10.03%

c.) Compute t statistics for each model parameter. What conclusions can ou draw?

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
25.4979	86.18%	76.96%	10.03%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	351.0	74.8	4.70	0.018	
viscosity	-1.27	1.17	-1.09	0.356	2.64
load	-0.1539	0.0895	-1.72	0.184	2.64

Regression Equation

wear = 351.0 - 1.27 viscosity - 0.1539 load

Since the t-values are relatively small for both viscosity and wear, but the t-value of the constant is high, we can conclude that that the two variables probably have a linearly dependent relationship.

10.9

The yield of a chemical process is related to the concentration of the reactant and the operating temperature. An experiment has been conducted with the following results.

Yield	Concentration	Temperature
81	1.00	150
89	1.00	180
83	2.00	150
91	2.00	180
79	1.00	150
87	1.00	180
84	2.00	150
90	2.00	180

a.)Suppose we wish to fit a main effects model to this data. Set up the $\boldsymbol{X}'\boldsymbol{X}$ *matrix using the data exactly as it appears in the table.*

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1.00 & 150 \\ 1 & 1.00 & 180 \\ 1 & 2.00 & 150 \\ 1 & 2.00 & 180 \\ 1 & 1.00 & 150 \\ 1 & 1.00 & 180 \\ 1 & 2.00 & 150 \\ 1 & 2.00 & 180 \end{pmatrix}$$

$$\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1.00 & 1.00 & 2.00 & 2.00 & 1.00 & 1.00 & 2.00 & 2.00 \\ 150 & 180 & 150 & 180 & 150 & 180 & 150 & 180 \end{pmatrix} \begin{pmatrix} 1 & 1.00 & 150 \\ 1 & 1.00 & 180 \\ 1 & 2.00 & 150 \\ 1 & 2.00 & 180 \\ 1 & 1.00 & 150 \\ 1 & 1.00 & 180 \\ 1 & 2.00 & 150 \\ 1 & 2.00 & 180 \end{pmatrix} = \begin{pmatrix} 8 & 12 & 1320 \\ 12 & 20 & 1980 \\ 1320 & 1980 & 219600 \end{pmatrix}$$

b.)Is the matrix you obtain in part (a) diagonal? Discuss your response.

For a matrix to be diagonal, every value that is not on the diagonal should be 0. Since $\vec{X}'\vec{X}$ does not have 0's for any entry, it is not diagonal. This is because we did not orthogonally code our variables before constructing the $\boldsymbol{X}$ matrix.

c.)Suppose we write our model in terms of the "usual" coded variables

$$x_1 = \frac{\text{Conc} - 1.5}{0.5} x_2 = \frac{\text{Temp} - 165}{15}$$

Set up the $\boldsymbol{X}'\boldsymbol{X}$ *matrix for the model in terms of this set of coded variables. Is this matrix diagonal. Discuss your response.*

$$\boldsymbol{X} = \begin{pmatrix} 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 8 \end{pmatrix}$$

This matrix is diagonal because all of the nondiagonal values of the matrix are 0. It is diagonal because instead of using the natural factor levels, we orthogonally coded the variables.

d.)Define a new set of coded variables

$$x_1 = \frac{\text{Conc} - 1.5}{0.5} x_2 = \frac{\text{Temp} - 165}{15}$$

Set up the $\boldsymbol{X}'\boldsymbol{X}$ matrix for the model in terms of this set of coded variables. Is this matrix diagonal. Discuss your response.

$$\boldsymbol{X} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

$$\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 4 & 4 \\ 4 & 4 & 2 \\ 4 & 2 & 4 \end{pmatrix}$$

This matrix is also not diagonal because the values outside the diagonal are not zeroes. This is because the variables were not coded orthogonally (I believe that they were coded dummy, since the resultings matrix entries have all 0's or 1's).

e.)Summarize what you have learned from this problem about coding the variables.

Since the original experiment used an orthogonal desgin (each factor level has its only column in the design matrix and the columns are linearly independent), I would conclude based on our calculations that orthogonal coding is best paired with orthogonal design. This is because the resultant matrix is much simpler and easier to use.

4.3

A chemist wishes to test the effect of four chemical agents on the strength of a particular type of cloth. Because there might be variability from one bolt to another, the chemist ecides to use a randomized block design, with the bolts of cloth considered as blocks. SHe selects five bolts and applies all four chemicals in random order to each bolt. The resulting tensile strengths follow. Analyze the data from this experiment (use $\alpha = 0.05$*) and draw appropriate conclusions.*

	Bolt				
Chemical	1	2	3	4	5
1	73	68	74	71	67
2	73	67	75	72	70
3	75	68	78	73	68
4	73	71	75	75	69

Based on the included minitab results below which indicate $p = .121$ for the chemical factor, we would fail to reject the hypothesis at the 5% significance level. Therefore, we cannot conclude that different chemical agents have an effect on the tensile strength of the fabric.

```
General Linear Model: Tensile-Strength versus Chemical, Bolt

Method

Factor coding  (-1, 0, +1)


Factor Information

Factor    Type    Levels  Values
Chemical  Fixed        4  1, 2, 3, 4
Bolt      Random       5  1, 2, 3, 4, 5


Analysis of Variance

Source      DF  Adj SS  Adj MS  F-Value  P-Value
  Bolt       4  157.00  39.250    21.61    0.000
  Chemical   3   12.95   4.317     2.38    0.121
Error       12   21.80   1.817
Total       19  191.75


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
1.34784  88.63%     82.00%      68.42%


Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant  71.750    0.301   238.07    0.000
```

```
Bolt
  1        1.750    0.603     2.90    0.013     *
  2       -3.250    0.603    -5.39    0.000     *
  3        3.750    0.603     6.22    0.000     *
  4        1.000    0.603     1.66    0.123     *
Chemical
  1       -1.150    0.522    -2.20    0.048  1.50
  2       -0.350    0.522    -0.67    0.515  1.50
  3        0.650    0.522     1.25    0.237  1.50


Regression Equation

Tensile-Strength = 71.750 + 1.750 Bolt_1 - 3.250 Bolt_2 + 3.750 Bolt_3 + 1.000 Bolt_4
                   - 3.250 Bolt_5 - 1.150 Chemical_1 - 0.350 Chemical_2 + 0.650 Chemical_3
                   + 0.850 Chemical_4

Equation treats random terms as though they are fixed.


Expected Mean Squares, using Adjusted SS

             Expected Mean Square
   Source    for Each Term
1  Bolt      (3) + 4.0000 (1)
2  Chemical  (3) + Q[2]
3  Error     (3)


Error Terms for Tests, using Adjusted SS

                                   Synthesis
   Source    Error DF  Error MS  of Error MS
1  Bolt         12.00    1.8167  (3)
2  Chemical     12.00    1.8167  (3)


Variance Components, using Adjusted SS

Source  Variance  % of Total    StDev  % of Total
Bolt     9.35833      83.74%  3.05914      91.51%
Error    1.81667      16.26%  1.34784      40.32%
Total     11.175              3.34290
```

4.10

An article in the ***Fire Safety Journal*** *("The Effect of Nozzle Design on the Stability and Performance of Turblent Water Jets," Vol. 4, August 1981) describes an experiment in which a shape factor was determined for several different nozzle designs at six levels of jet efflux velocity. Interest focused on potential differences* *between nozzle designsm with velocity considered as a nuisance variable. The data are shown below:*

Nozzle	Jet Efflux Velocity (m/s)					
Design	11.73	14.37	16.59	20.43	23.46	28.74
1	0.78	0.80	0.81	0.75	0.77	0.78
2	0.85	0.85	0.92	0.86	0.81	0.83
3	0.93	0.92	0.95	0.89	0.89	0.83
4	1.14	0.97	0.98	0.88	0.86	0.83
5	0.97	0.86	0.78	0.76	0.76	0.75

a.)Does the nozzle design affect the shape factor? Compare the nozzles with a scatter plat and with an analysis of variance, using $\alpha = 0.05$.
Based on the minitab output below which shows that the p-value of the nozzle design factor is .0000, we can conclude that there is strong evidence to reject the null and conclude that nozzle design has a significant effect on the shape factor. This result is echoed in the scatterpot below because ranges don't overlap in the extreme.

```
General Linear Model: Shape Factor versus Nozzle Design, JEV

Method

Factor coding  (-1, 0, +1)


Factor Information

Factor         Type    Levels  Values
Nozzle Design  Fixed        5  1, 2, 3, 4, 5
JEV            Random       6  1, 2, 3, 4, 5, 6


Analysis of Variance

Source           DF   Adj SS    Adj MS  F-Value  P-Value
  Nozzle Design   4  0.10218  0.025545     8.92    0.000
  JEV             5  0.06287  0.012573     4.39    0.007
Error            20  0.05730  0.002865
Total            29  0.22235


Model Summary

        S    R-sq  R-sq(adj)  R-sq(pred)
0.0535257  74.23%     62.63%      42.02%


Coefficients

Term              Coef  SE Coef  T-Value  P-Value   VIF
Constant       0.85867  0.00977    87.87    0.000
Nozzle Design
  1            -0.0770   0.0195    -3.94    0.001  1.60
```

2	-0.0053	0.0195	-0.27	0.788	1.60
3	0.0430	0.0195	2.20	0.040	1.60
4	0.0847	0.0195	4.33	0.000	1.60
JEV					
1	0.0753	0.0219	3.45	0.003	*
2	0.0213	0.0219	0.98	0.341	*
3	0.0293	0.0219	1.34	0.195	*
4	-0.0307	0.0219	-1.40	0.176	*
5	-0.0407	0.0219	-1.86	0.078	*

Regression Equation

```
Shape Factor = 0.85867 - 0.0770 Nozzle Design_1 - 0.0053 Nozzle Design_2
               + 0.0430 Nozzle Design_3 + 0.0847 Nozzle Design_4 - 0.0453 Nozzle Design_5
               + 0.0753 JEV_1 + 0.0213 JEV_2 + 0.0293 JEV_3 - 0.0307 JEV_4 - 0.0407 JEV_5
               - 0.0547 JEV_6
```

Equation treats random terms as though they are fixed.

Fits and Diagnostics for Unusual Observations

Obs	Shape Factor	Fit	Resid	Std Resid	
19	1.1400	1.0187	0.1213	2.78	R

R Large residual

Expected Mean Squares, using Adjusted SS

	Source	Expected Mean Square for Each Term
1	Nozzle Design	(3) + Q[1]
2	JEV	(3) + 5.0000 (2)
3	Error	(3)

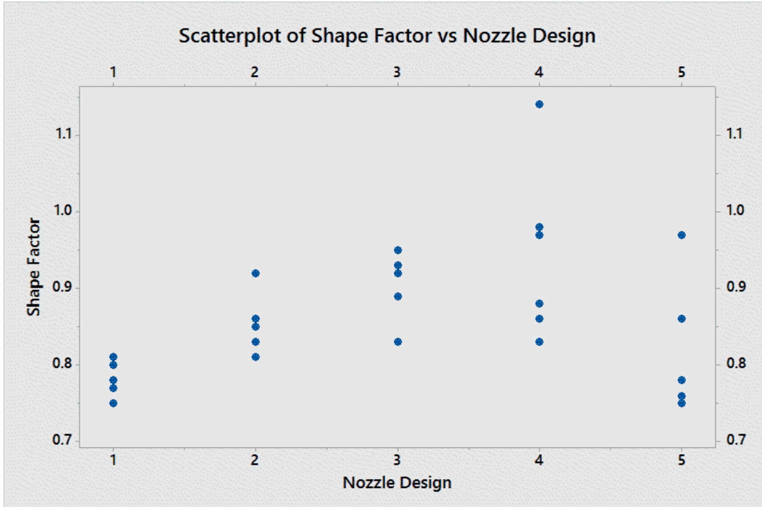
Error Terms for Tests, using Adjusted SS

	Source	Error DF	Error MS	Synthesis of Error MS
1	Nozzle Design	20.00	0.0029	(3)
2	JEV	20.00	0.0029	(3)

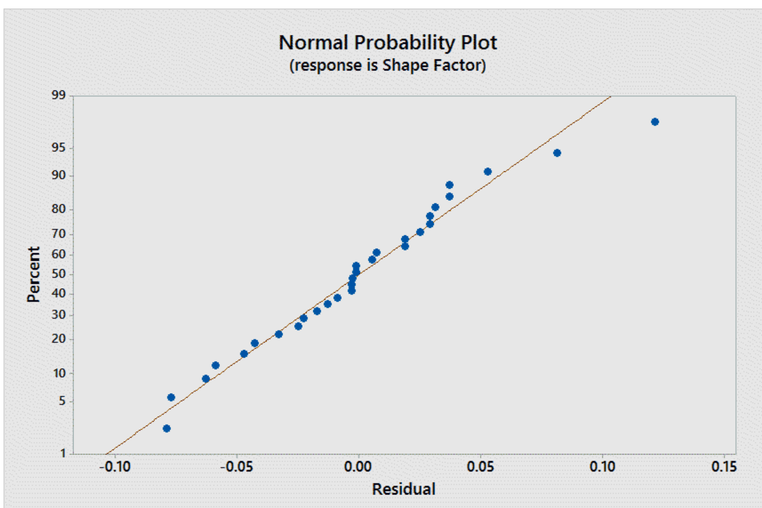
Variance Components, using Adjusted SS

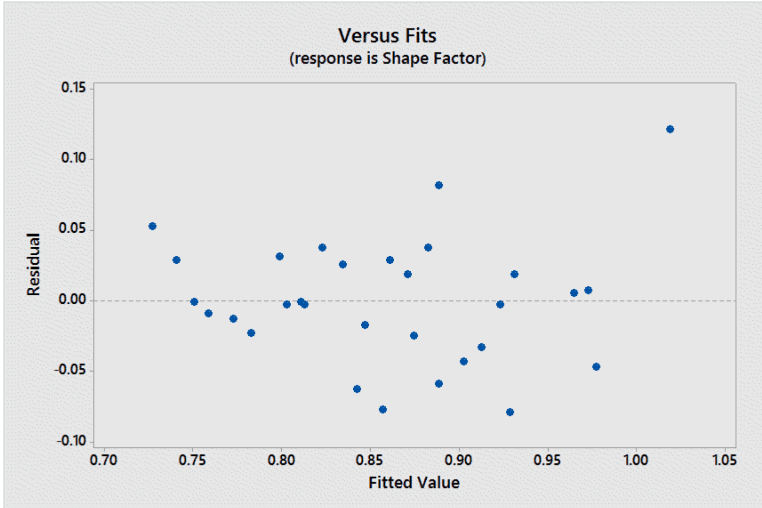
Source	Variance	% of Total	StDev	% of Total
JEV	0.0019417	40.40%	0.0440643	63.56%

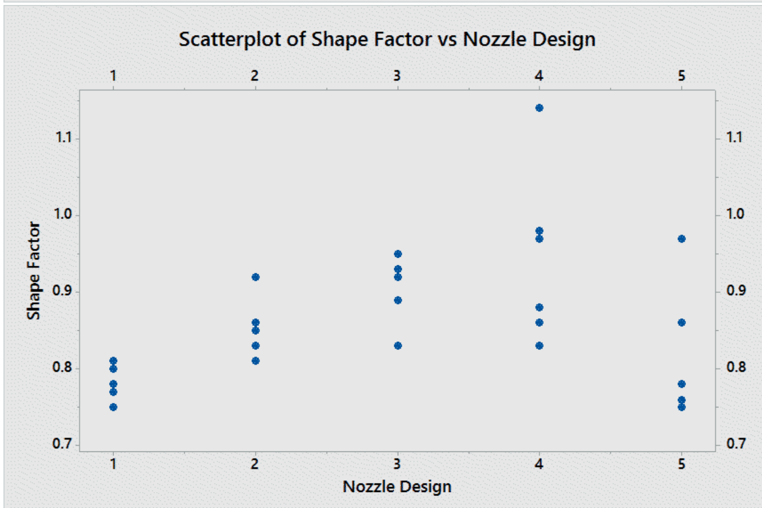
Error	0.002865	59.60%	0.0535257	77.20%
Total	0.0048067		0.0693301	



b.)Analyze the residuals from this experiment.
The normal probability plot has a relatively straight line and does not give us reason to question the normality assumption. The residual plot does not show any clear pattern, so it also doesn't give us reason to question the normal and indepently distributed assumption. It is worth noting that there does seem to be a small outlier on both graphs, but since the data was found to be significant at the 5% level, it does not give us reason to question our assumptions.





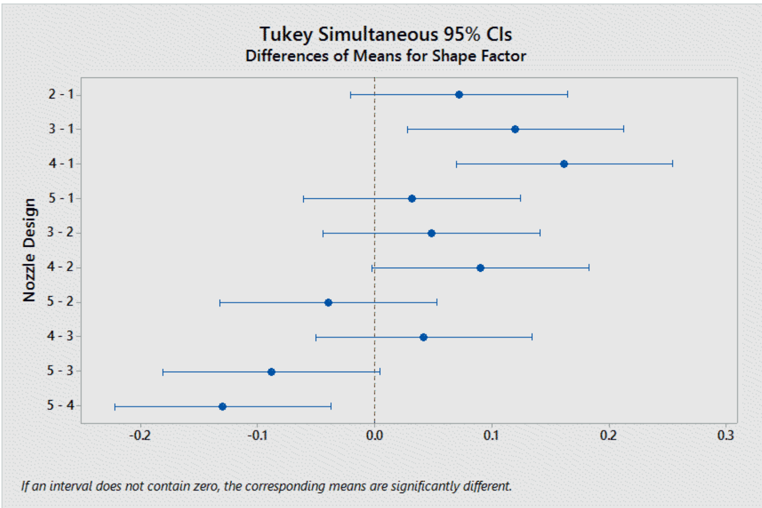


c.) Which nozzle designs are different with respect to shape factor? Draw a graph of the average shape factor for each nozzle type and compare this to a scaled t distribution. Compare the conclusions that you draw from this plot to those from Duncan's multiple range test.
Since we are omitting the graph and the Duncan's test, we need to select appropriate alternative post hoc tests for comparisons. We will use Tukey, Fisher, and Bonferroni.

```
Comparisons for Shape Factor

Tukey Pairwise Comparisons: Response = Shape Factor, Term = Nozzle Design

Grouping Information Using the Tukey Method and 95% Confidence

Nozzle
Design  N      Mean  Grouping
4       6  0.943333  A
3       6  0.901667  A    B
2       6  0.853333  A    B  C
5       6  0.813333       B  C
1       6  0.781667          C
```

Means that do not share a letter are significantly different.

Tukey Simultaneous 95% CIs



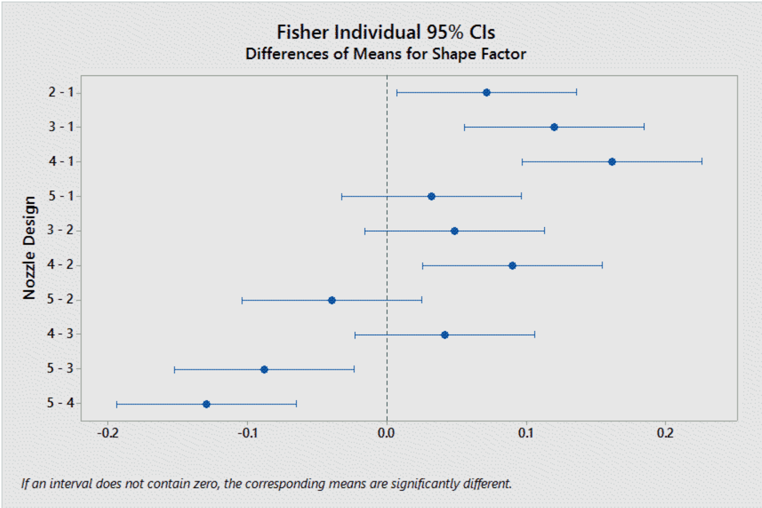
Fisher Pairwise Comparisons: Response = Shape Factor, Term = Nozzle Design

Grouping Information Using Fisher LSD Method and 95% Confidence

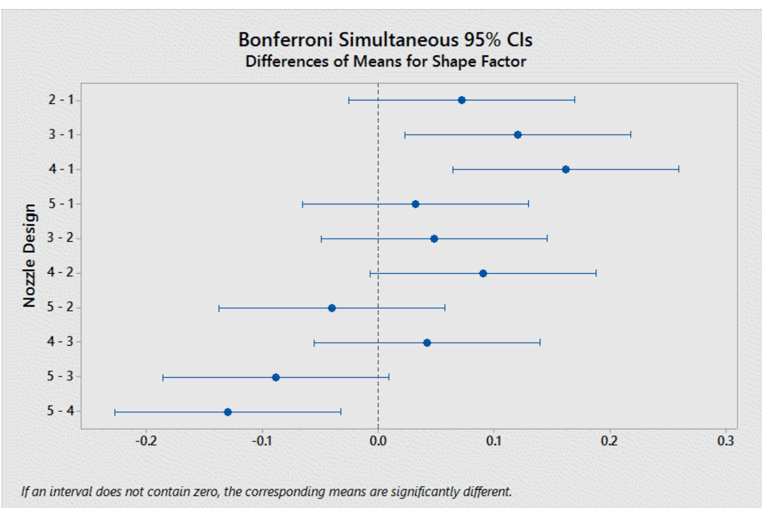
Nozzle Design	N	Mean	Grouping			
4	6	0.943333	A			
3	6	0.901667	A	B		
2	6	0.853333		B	C	
5	6	0.813333			C	D
1	6	0.781667				D

Means that do not share a letter are significantly different.

Fisher Individual 95% CIs




```
Bonferroni Pairwise Comparisons: Response = Shape Factor, Term = Nozzle Design

Grouping Information Using the Bonferroni Method and 95% Confidence

Nozzle
Design  N      Mean  Grouping
4       6  0.943333  A
3       6  0.901667  A   B
2       6  0.853333  A   B  C
5       6  0.813333      B  C
1       6  0.781667         C

Means that do not share a letter are significantly different.


Bonferroni Simultaneous 95% CIs
```



A comparison of the minitab results reveals that while Tukey and Bonferroni come to the same conclusions regarding which means are statistically significantly different, Fisher finds the relationships to be somewhat

more nuinanced. That is to say that the Fisher test concludes that Nozzle Designs 2 and 3 and also 2 and 5 are significantly different from each other, whereas Tukey and Bonferroni do not.
Since Fisher's test does not correct for multiple comparison's and Tukey's is best suited for samples with different sizes, we would agree with and report the Bonferroni results.

4.11

Consider the ratio control algorithm experiment described in Section 3.8. The experiment was actually conducted as a randomized block design, where six time periods were selected as the blocks, and all four ratio control algorithms were tested in each time period. The average cell voltage and the standard deviation of voltage (shown in parentheses) for each cell are as follows:

Ratio Control Algorithm	Time Period		
	1	2	3
1	4.93(0.05)	4.86(0.04)	4.75(0.05)
2	4.85(0.04)	4.91(0.02)	4.79(0.03)
3	4.83(0.09)	4.88(0.13)	4.90(0.11)
4	4.89(0.03)	4.77(0.04)	4.94 (0.05)

a.)Analyze the average cell voltage data. (Use $\alpha = 0.05$). Does the choice of ration control algorithm affect the average cell voltage:
Based on the minitab output below which gives us a p-value of 0.901, we cannot reject the null hypothesis in favor of the alternative and therefore cannot conclude that the choice of ratio control algorithm has an affect on the average cell voltage.

```
General Linear Model: Avg Cell Voltage versus RCA, Time Period

Method

Factor coding  (-1, 0, +1)


Analysis of Variance

Source         DF    Seq SS  Contribution    Adj SS    Adj MS  F-Value  P-Value
  RCA           3  0.002746         2.97%  0.002746  0.000915     0.19    0.901
  Time Period   5  0.017437        18.88%  0.017437  0.003487     0.72    0.615
Error          15  0.072179        78.15%  0.072179  0.004812
Total          23  0.092363       100.00%


Model Summary

        S    R-sq  R-sq(adj)     PRESS  R-sq(pred)
0.0693682  21.85%      0.00%  0.184779       0.00%


Regression Equation

Avg Cell Voltage = 4.8438 + 0.0163 RCA_1 - 0.0104 RCA_2 + 0.0029 RCA_3 - 0.0088 RCA_4
                   + 0.0312 Time Period_1 + 0.0112 Time Period_2 + 0.0013 Time Period_3
```

+ 0.0088 Time Period_4 - 0.0562 Time Period_5 + 0.0037 Time Period_6

Equation treats random terms as though they are fixed.

Expected Mean Squares, using Adjusted SS

	Source	Expected Mean Square for Each Term
1	RCA	(3) + Q[1]
2	Time Period	(3) + 4.0000 (2)
3	Error	(3)

Error Terms for Tests, using Adjusted SS

	Source	Error DF	Error MS	Synthesis of Error MS
1	RCA	15.00	0.0048	(3)
2	Time Period	15.00	0.0048	(3)

Means

Term	Fitted Mean
RCA	
1	4.86000
2	4.83333
3	4.84667
4	4.83500
Time Period	
1	4.87500
2	4.85500
3	4.84500
4	4.85250
5	4.78750
6	4.84750

Variance Components, using Adjusted SS

Source	Variance	% of Total	StDev	% of Total
Time Period	-0.000331111*	0.00%	0.0000000	0.00%
Error	0.00481194	100.00%	0.0693682	100.00%
Total	0.00481194		0.0693682	

* Value is negative, and is estimated by zero.

b.)Perform an appropriate analysis on the standard deviation of voltage. (Recall that this is called "pot noise.") Does the choice of ratio control algorithm affect the pot noise?

Based on the minitab results below which report a p-value of 0.00, we can conclude that the choice of ratio control algorithm has a strong affect on the pot noise (the standard deviation of the voltage).

```
General Linear Model: St Dev versus RCA, Time Period

Method

Factor coding  (-1, 0, +1)


Analysis of Variance

Source         DF    Seq SS  Contribution    Adj SS    Adj MS  F-Value  P-Value
  RCA           3  0.026012        83.12%  0.026013  0.008671    50.76    0.000
  Time Period   5  0.002721         8.69%  0.002721  0.000544     3.19    0.037
Error          15  0.002563         8.19%  0.002563  0.000171
Total          23  0.031296       100.00%


Model Summary

        S    R-sq  R-sq(adj)    PRESS  R-sq(pred)
0.0130703  91.81%     87.45%  0.00656      79.04%


Regression Equation

St Dev = 0.05708 - 0.01042 RCA_1 - 0.02542 RCA_2 + 0.05625 RCA_3 - 0.02042 RCA_4
         - 0.00458 Time Period_1 + 0.00042 Time Period_2 + 0.00292 Time Period_3
         + 0.02042 Time Period_4 - 0.01458 Time Period_5 - 0.00458 Time Period_6

Equation treats random terms as though they are fixed.


Expected Mean Squares, using Adjusted SS

                Expected Mean Square
   Source       for Each Term
1  RCA          (3) + Q[1]
2  Time Period  (3) + 4.0000 (2)
3  Error        (3)


Error Terms for Tests, using Adjusted SS

                                    Synthesis
   Source       Error DF  Error MS  of Error MS
1  RCA             15.00    0.0002  (3)
2  Time Period     15.00    0.0002  (3)


Means
```

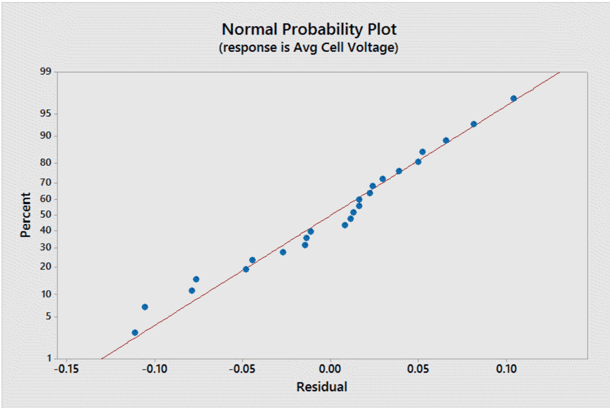
Term	Fitted Mean
RCA	
1	0.046667
2	0.031667
3	0.113333
4	0.036667
Time Period	
1	0.052500
2	0.057500
3	0.060000
4	0.077500
5	0.042500
6	0.052500

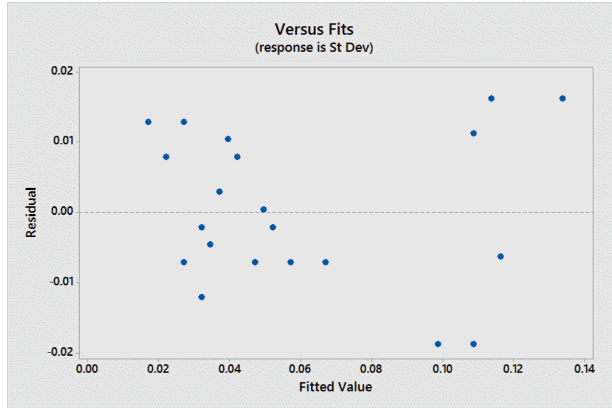
Variance Components, using Adjusted SS

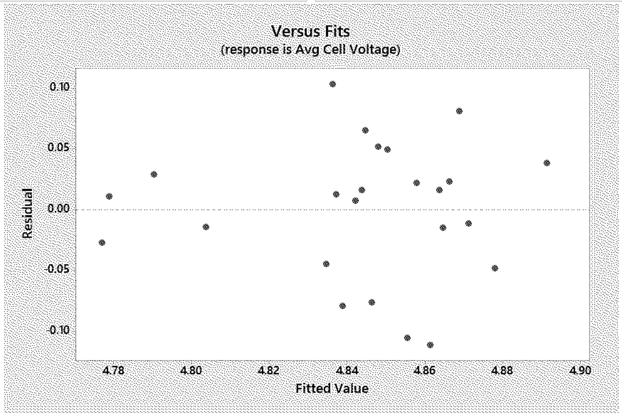
Source	Variance	% of Total	StDev	% of Total
Time Period	0.0000933	35.33%	0.0096609	59.44%
Error	0.0001708	64.67%	0.0130703	80.42%
Total	0.0002642		0.0162532	

c.) Conduct any residual analysis that seem appropriate.

Based on the included graphs below, we do not have any cause for concern on which to base a challenge against our assumptions. The normal probability plot gives some pause, but the residuals are generally straight. The residuals in the scatterplots do not show and specific patterns and there are no extreme outliers.







d.) Which ratio control algorthm would you select if your objective is to reduce both the average cell voltage and the pot noise?"

To reduce both average cell voltage and pot noise, I would select ratio control algorithm number 2 which has the lowest treatment mean for both variables.