CSCI 688
Homework 4

Megan Rose Bryant
Department of Mathematics
William and Mary

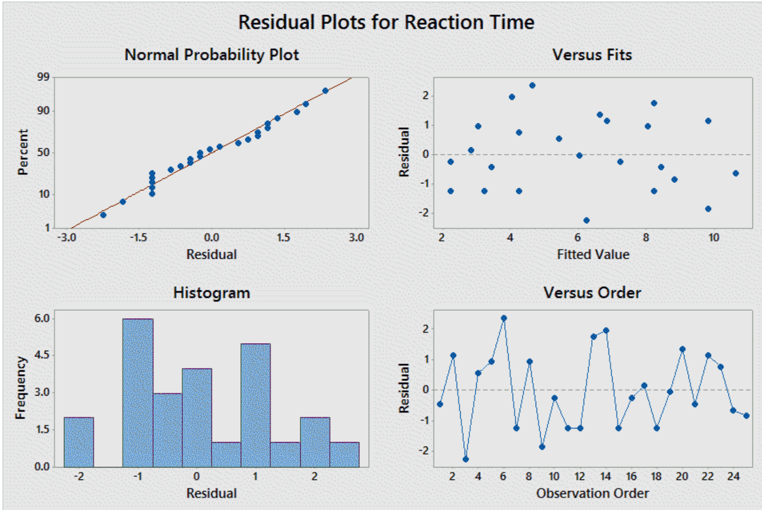
October 6, 2014

4.20 *The effect of five different ingredients (A,B,C,D,E) on the reaction time of a chemical process is being studied. Each batch of new material is only large enough to permit five runs to be made. Furthermore, each run requires approximately $1\frac{1}{2}$ hours, so only five runs can be made in one day. The experimenter decides to run the experiment as a Latin square so that day and batch effects may be systematically controlled. She obtains the data that follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.*

	Day				
Batch	**1**	**2**	**3**	**4**	**5**
1	A = 8	B = 7	D = 1	C = 7	E = 3
2	C = 11	E = 2	A = 7	D = 3	B = 8
3	B = 4	A = 9	C = 10	E =1	D = 5
4	D = 6	C = 8	E = 6	B = 6	A = 10
5	E = 4	D = 2	B = 3	A = 8	C = 8

```
General Linear Model: Reaction Time versus Batch, Day, Ingredient

Method

Factor coding  (-1, 0, +1)


Factor Information

Factor      Type    Levels  Values
Batch       Random       5  1, 2, 3, 4, 5
Day         Random       5  1, 2, 3, 4, 5
Ingredient  Fixed        5  1, 2, 3, 4, 5


Analysis of Variance

Source         DF  Adj SS  Adj MS  F-Value  P-Value
  Batch         4   15.44   3.860     1.23    0.348
  Day           4   12.24   3.060     0.98    0.455
  Ingredient    4  141.44  35.360    11.31    0.000
Error          12   37.52   3.127
Total          24  206.64


Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
1.76824  81.84%     63.69%      21.19%
```

Based on the obtained minitab results, we can conclude that the chemical ingredients have a significant effect the reaction time because it has a p-value of 0. Also, we can conclude that the effects of batch and day are negligible on reaction time (once they are controlled for) because they both have p-values greater than our α of .05. The normal probability plot and the residual graphs below do not give us any reason to question our normality or identical independent assumptions.



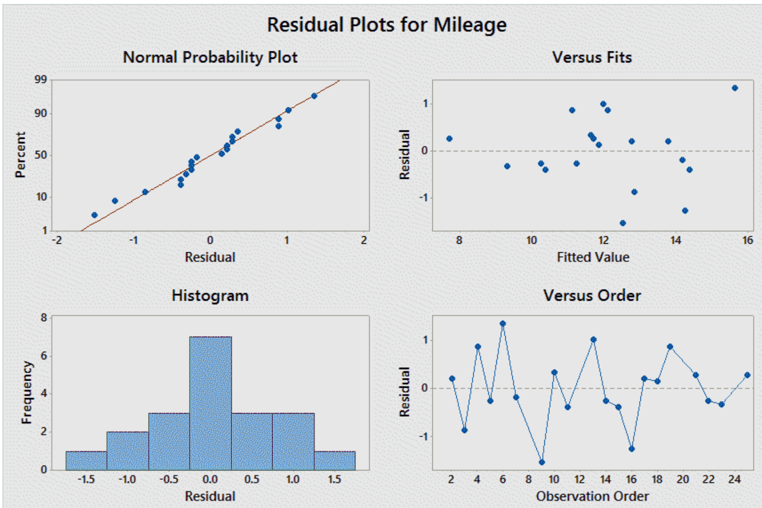
4.33 *An engineer is studying the mileage performance characteristics of five types of gasoline additives. In the road test he wishes to use cars as block; however, because of a time constraint, he must use an incomplete block design. He runs the balanced design with the five blocks that follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.*

	Day				
Additive	**1**	**2**	**3**	**4**	**5**
1		17	14	13	12
2	14	14		13	10
3	12		13	12	9
4	13	11	11	12	
5	11	12	10		8

```
General Linear Model: Mileage versus Car, Additive

Method

Factor coding  (-1, 0, +1)
Rows unused    5


Factor Information

Factor    Type    Levels  Values
Car       Random       5  1, 2, 3, 4, 5
Additive  Fixed        5  1, 2, 3, 4, 5


Analysis of Variance

Source      DF  Adj SS  Adj MS  F-Value  P-Value
  Car        4   35.23  8.8083     9.67    0.001
  Additive   4   35.73  8.9333     9.81    0.001
Error       11   10.02  0.9106
Total       19   76.95
```

```
Model Summary


       S    R-sq  R-sq(adj)  R-sq(pred)
0.954257  86.98%     77.52%      56.97%
```

Based on the obtained Minitab output, we can conclude that both the car and the gasoline additive have a statistically significant affect on the mileage as both have p-values of 0.01. The normal probability plot and the residual graphs included below do not give us any reason to question the normality or identically independently distributed assumptions. Though it is woth noting that there seems to be some mild grouping in the NPP.



4.42 *Verify that a BIBD with the parameters $a = 8, r = 8, k = 4,$ and $b = 16$ does not exist.*
We know that for a BIBD with those parameters to exist, there must exist an integer λ such that

$$\lambda = \frac{r(k-1)}{a-1} = \frac{8(4-1)}{8-1} = \frac{24}{7} \notin \mathbb{Z}$$

Therefore, a BIBD does not exist with those parameters.

5.3 *The yield of a chemical process is being studied. The two most important variables are thought to be the pressure and the temperature. Three levels of each factor are selected and a factorial experiment with two replicates is performed. The yield data are as follows.*

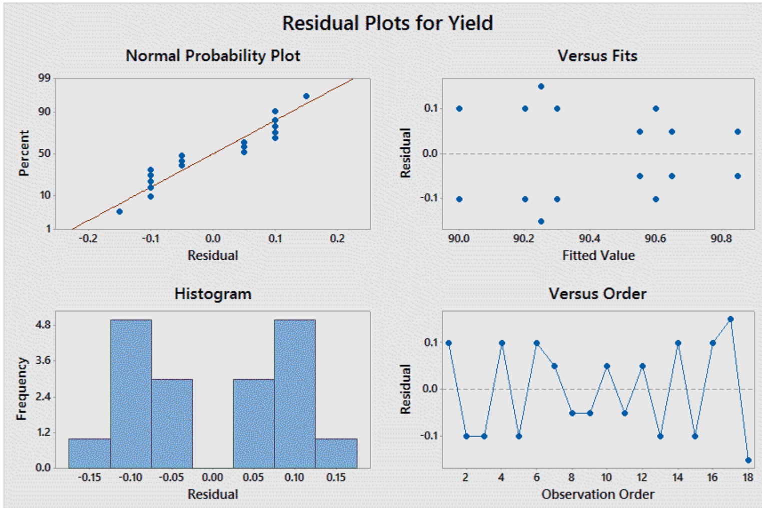
	Pressure (psig)		
Temperature °C	**200**	**215**	**230**
150	90.4	90.7	90.2
	90.2	90.6	90.4
160	90.1	90.5	89.9
	90.3	90.6	90.1
170	90.5	90.8	90.4
	90.7	90.9	90.1

a.) *Analyze the data and draw conclusions. Use $\alpha - = 0.05$.*

```
General Linear Model: Yield versus Temp, Pressure

Method

Factor coding  (-1, 0, +1)


Factor Information

Factor    Type   Levels  Values
Temp      Fixed       3  1, 2, 3
Pressure  Fixed       3  1, 2, 3


Analysis of Variance

Source           DF   Adj SS   Adj MS  F-Value  P-Value
  Temp            2  0.30111  0.15056     8.47    0.009
  Pressure        2  0.76778  0.38389    21.59    0.000
  Temp*Pressure   4  0.06889  0.01722     0.97    0.470
Error             9  0.16000  0.01778
Total            17  1.29778


Model Summary

       S    R-sq  R-sq(adj)  R-sq(pred)
0.133333  87.67%     76.71%      50.68%
```

Based on the obtained Minitab output, we would conclude that both pressure and temperature are significant, since they have p-values that are less than our α of 0.05. However, we would also conclude that the interaction between pressure and temperature is not significant since the p-value is 0.47, which is larger than our α.

b.) *Prepare appropriate residual plots and comment on the model's adequacy.*
The normal probability and residual plots do not give us significant cause to question the normality or identically distributed assumptions, though it is important to note that there seems to be more of a variation in the third batches of both temperature and pressure.



c.) *Under what conditions would you operate this process?*

Comparisons for Yield

Tukey Pairwise Comparisons: Response = Yield, Term = Temp

Grouping Information Using the Tukey Method and 95% Confidence

Temp	N	Mean	Grouping	
3	6	90.5667	A	
1	6	90.4167	A	B
2	6	90.2500		B

Means that do not share a letter are significantly different.

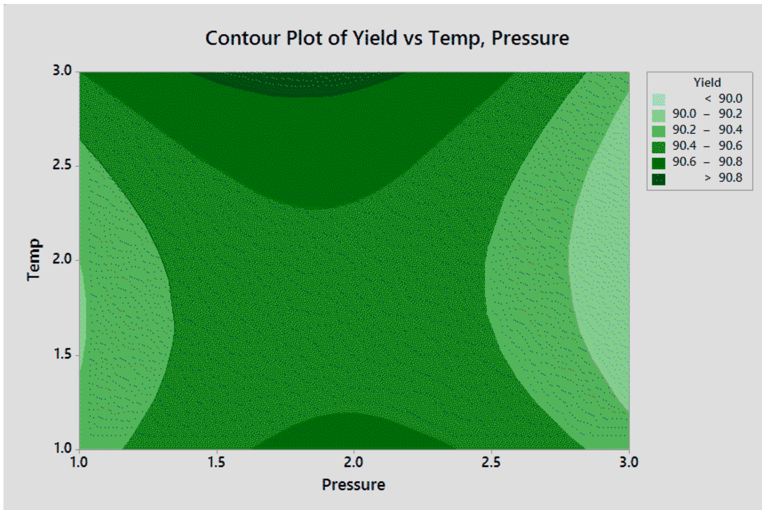
Tukey Pairwise Comparisons: Response = Yield, Term = Pressure

Grouping Information Using the Tukey Method and 95% Confidence

Pressure	N	Mean	Grouping	
2	6	90.6833	A	
1	6	90.3667		B
3	6	90.1833		B

Means that do not share a letter are significantly different.

We see from our Tukey Pairwise Comparisons for Yield that the process should be operated using temperature 3: 170° and pressure 2: 215 psig for the highest yield. It is important to note that Tukey did not find a significant difference between temperature 1 and temperature 3 or temperature 1 and temperature 2, however we will select temperature 3 both because it has the highest mean yield and because it is statistically significantly different from temperature 2. This selection is also supported by the contour plot included below.



5.4 *An engineer suspects that the surface finish of a meat part is influenced by the feed rate and the depth of cut. He selects three feed rates and four depths of cut. He then conducts a factorial experiment and obtains the folowing data:*

Feed Rate	**Depth of Cut (in)**			
(in/min)	**0.15**	**0.18**	**0.20**	**0.25**
	74	79	82	99
0.20	64	68	88	104
	60	73	92	96
	92	98	99	104
0.25	86	104	108	110
	88	88	95	99
	99	104	108	114
0.30	98	99	110	111
	102	95	99	107

a.) *Analyze the data and draw concusions. Use* $\alpha = 0.05$.

```
General Linear Model: Surface Finish versus Feed Rate, Depth of Cut

Method

Factor coding  (-1, 0, +1)


Factor Information

Factor         Type   Levels  Values
Feed Rate      Fixed       3  1, 2, 3
Depth of Cut   Fixed       4  1, 2, 3, 4


Analysis of Variance
```

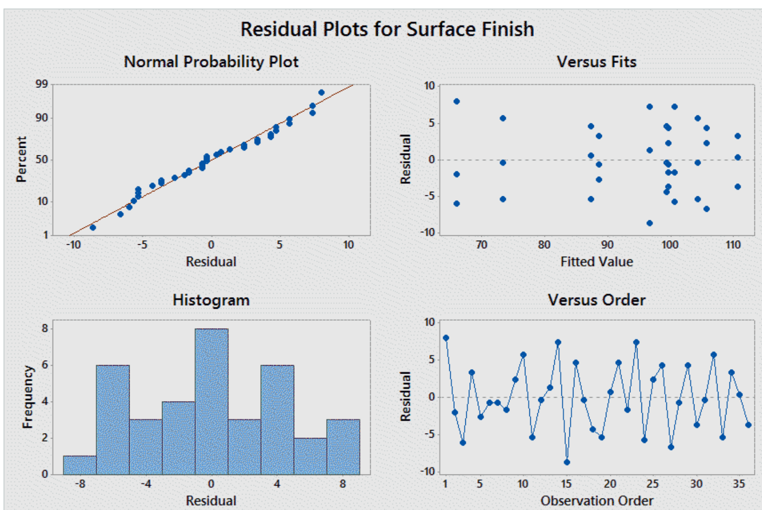
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Feed Rate	2	3160.5	1580.25	55.02	0.000
Depth of Cut	3	2125.1	708.37	24.66	0.000
Feed Rate*Depth of Cut	6	557.1	92.84	3.23	0.018
Error	24	689.3	28.72		
Total	35	6532.0			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5.35931	89.45%	84.61%	76.26%

Based on the obtained Minitab output, we can conclude that both Feed Rate and Depth of Cut are significant since they have p-values of less than our α of 0.05. We can also declare the interaction between Feed Rate and Depth of Cut to be significant as it has a p-value of 0.018, which is also less than α.

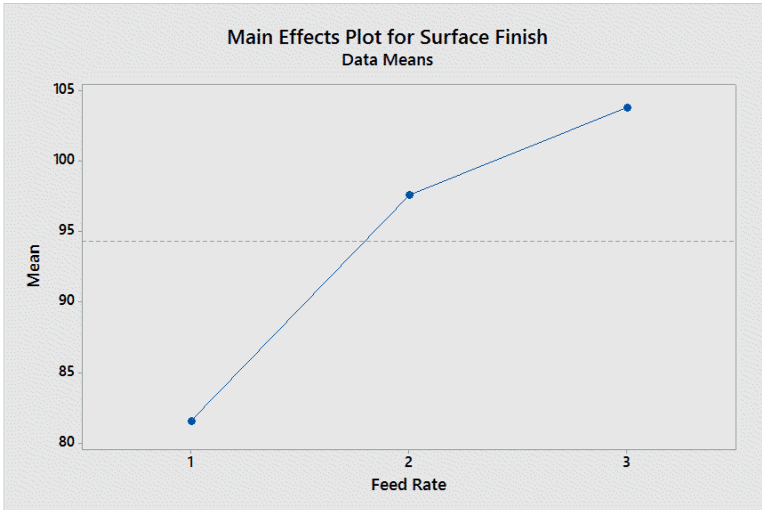
b.) *Prepare appropriate residuals plots and comment on the model's adequacy.*
The normal probability and residual plots do not give us any reason to doubt our normality and independent identical distribution assumptions.



c.) *Obtain point estimates of the mean surface finish at each feed rate.*

Term	Fitted Mean	SE Mean
Feed Rate		
1	81.58	1.55
2	97.58	1.55
3	103.83	1.55
Depth of Cut		
1	84.78	1.79
2	89.78	1.79
3	97.89	1.79
4	104.89	1.79

We are only interested in the point estimates of the mean surface finish at each feed rate, so we will use a one-factor plot.



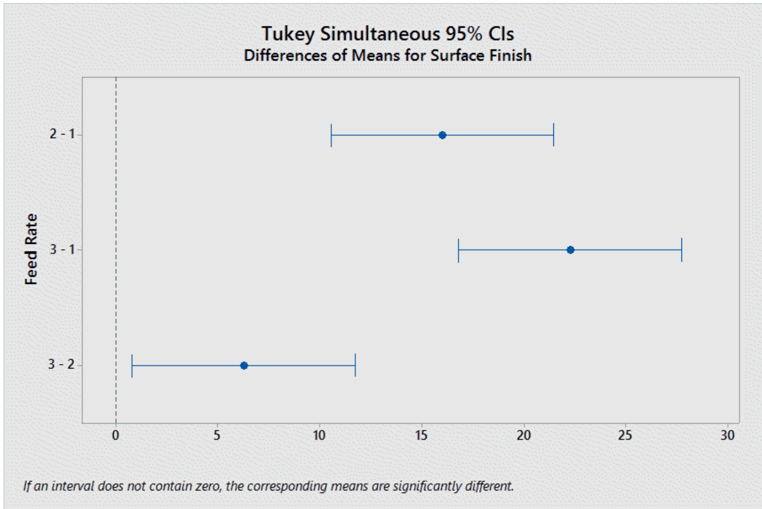
d.) *Find the P-values for the tests in part (a).*
As obtained in part *a.)*, the p-values are as follows:

Factor	P-Value
Feed Rate	0.000
Depth of Cut	0.000
Feed Rate * Depth of Cut	0.018

5.5 *For the data in problem 5.4 compute a 95 percent confidence interval estimate of the mean difference in response for feed rates of 0.20 and 0.25 in/min.*

```
Tukey Pairwise Comparisons: Response = Surface Finish, Term = Feed Rate

Grouping Information Using the Tukey Method and 95% Confidence

Feed
Rate   N     Mean  Grouping
3     12  103.833  A
2     12   97.583      B
1     12   81.583         C

Means that do not share a letter are significantly different.


Tukey Simultaneous Tests for Differences of Means

Difference
of Feed      Difference      SE of   Simultaneous              Adjusted
Rate Levels    of Means  Difference      95% CI      T-Value    P-Value
2 - 1             16.00        2.19  (10.54, 21.46)     7.31      0.000
3 - 1             22.25        2.19  (16.79, 27.71)    10.17      0.000
```

```
3 - 2              6.25        2.19  ( 0.79, 11.71)     2.86     0.023

Individual confidence level = 98.02%
```



We are only interested in a 95% confidence interval for the mean difference between levels 1 and 2 of Feed Rate. Therefore, our desired confidence interval is $(10.54, 21.46)$ e.g. $16 + / - 5.46$.

5.20 *In Problem 5.3, suppose that we wish to reject the null hypothesis with a high probability if the difference in the true mean yield at any two pressures is as great as 0.5. If a reasonable prior estimate of the standard deviation of yield is 0.1, how many should be run?*
We know that

$$a = 3b = 3D = 0.5$$

Therefore, we can calculate the desired sample size by finding the minimum value of ϕ^2.

$$\phi^2 = \frac{naD^2}{2b\sigma^2} = \frac{n * 3 * (0.5)^2}{2 * (3) * (0.1)^2} = 12.5n$$

For a sample size of $n = 2$, we see that we have a $\phi = 5$. Using the given formulas and the charts available in Appendix V (the Operating Charecteristic Curves...) we can obtain the following:

n	ϕ^2	ϕ	ν_1	ν_2	β
2	25	5	2	9	¡ .05

For $\nu_1 = 2$ and $\nu_2 = 9$, we have a $\beta < 0.05$ on the OCC graph. Therefore, a sample size of 2 is sufficient to detect a difference of 0.5 between pressures.