STAT 421 Homework 7

Lecture 17 Problem 1:
In this problem, we will be using the data from problem 5.23 in the textbook. Problem 5.23 states that we will consider the data in Problem 5.8 assuming that replicates are blocks. In this experiment, the factors that influence the breaking strength of a synthetic fiber are being studied. Four production machines and three operators are chosen and a factorial experiment is run using fiber from the same production batch. The data from this experiment are shown below.

	Machine			
Operator	**1**	**2**	**3**	**4**
1	109	110	108	110
	110	115	109	108
2	110	110	111	114
	112	111	109	112
3	116	112	114	120
	114	115	119	117

The goal of this problem is to produce the ANOVA table, and state our conclusions (using $\alpha = 0.05$). We will do all of the analysis in R and then draw conclusions, this work in R is shown below.

```r
# Set up the number of each factor:
num_a <- 3
num_b <- 4
num_c <- 2

# Set up an empty matrix for the data:
y_mat <- matrix(nrow=num_a*num_c, ncol=num_b)

# Fill in the matrix with the data from the problem:
y_mat[1,] <- c(109, 110, 108, 110)
y_mat[2,] <- c(110, 115, 109, 108)
y_mat[3,] <- c(110, 110, 111, 114)
y_mat[4,] <- c(112, 111, 109, 112)
y_mat[5,] <- c(116, 112, 114, 120)
y_mat[6,] <- c(114, 115, 119, 117)

# Make the data into a vector (in row order):
y <- as.vector(t(y_mat))

# Set up the factors to match the data:
A <- as.factor(rep(1:num_a, each=num_b*num_c))
```

```r
B <- as.factor(rep(1:num_b, times=num_a*num_c))
C <- as.factor(rep(rep(1:num_c, each=num_b), times=num_a))

# Confirm that the data in R is the same as the problem:
cbind(A, B, C, y)
```

```
##       A B C   y
##  [1,] 1 1 1 109
##  [2,] 1 2 1 110
##  [3,] 1 3 1 108
##  [4,] 1 4 1 110
##  [5,] 1 1 2 110
##  [6,] 1 2 2 115
##  [7,] 1 3 2 109
##  [8,] 1 4 2 108
##  [9,] 2 1 1 110
## [10,] 2 2 1 110
## [11,] 2 3 1 111
## [12,] 2 4 1 114
## [13,] 2 1 2 112
## [14,] 2 2 2 111
## [15,] 2 3 2 109
## [16,] 2 4 2 112
## [17,] 3 1 1 116
## [18,] 3 2 1 112
## [19,] 3 3 1 114
## [20,] 3 4 1 120
## [21,] 3 1 2 114
## [22,] 3 2 2 115
## [23,] 3 3 2 119
## [24,] 3 4 2 117
```

```r
# Run ANOVA in R to get the results:
summary.aov(lm(y~A+B+C+A:B))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## A            2 160.33   80.17  20.291 0.000204 ***
## B            3  12.46    4.15   1.051 0.408659
## C            1   2.04    2.04   0.517 0.487209
## A:B          6  44.67    7.44   1.884 0.171630
## Residuals   11  43.46    3.95
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As computed above, since the p-value of 0.000204 is less than the significance level of 0.05, we reject the null in favor of the alternative and conclude that the factor A (the operator) does effect the breaking strength of a synthetic fiber. Furthermore, since the p-value of 0.408659 is greater than the significance level of 0.05, we fail to reject the null in favor of the alternative and conclude that their is no evidence that the factor B (the machine) effects the breaking strength of a synthetic fiber. Also, since the p-value of 0.171630 is greater than the significance level of 0.05, we fail to reject the null in favor of the alternative and conclude that their is no evidence of an interaction effect between the two treatments A and B. Lastly, since the p-value of 0.487209 is greater than the significance level of 0.05, we fail to reject the null in favor of the alternative and cautiously conclude that the there is no block effect.

Lecture 17 Problem 2a:
In this problem, we will be using the data from example 9.1 in the textbook. In this experiment, a machine is used to fill 5-gallon metal containers with soft drink syrup. The variable of interest is the amount of syrup loss due to frothing. Three factors are thought to influence frothing: the nozzle design (A), the filling speed (B), and the operating pressure (C). Three nozzles, three filling speeds, and three pressures are chosen, and two replicates of a 3^3 factorial experiment are run. The data are shown below.

■ **TABLE 9.1**
Syrup Loss Data for Example 9.1 (units are cubic centimeters −70)

	Nozzle Type (*A*)								
	1			2			3		
	Speed (in RPM) (*B*)								
Pressure (in psi) (*C*)	100	120	140	100	120	140	100	120	140
10	−35	−45	−40	17	−65	20	−39	−55	15
	−25	−60	15	24	−58	4	−35	−67	−30
15	110	−10	80	55	−55	110	90	−28	110
	75	30	54	120	−44	44	113	−26	135
20	4	−40	31	−23	−64	−20	−30	−61	54
	5	−30	36	−5	−62	−31	−55	−52	4

Considering only the first replicate, the goal of this problem is to fit a model that includes at most 2-way interactions, produce the ANOVA table, and state our conclusions (using $\alpha = 0.05$). In the end we want to discover which effects are significant, and which are not. We will do all of the analysis in R and then draw conclusions, this work in R is shown below.

```r
# Set up the number of each factor:
num_a <- 3
num_b <- 3
num_c <- 3

# Set up an empty matrix for the data:
y_mat <- matrix(nrow=num_a, ncol=num_b*num_c)

# Fill in the matrix with the data from the problem:
y_mat[1,] <- c(-35, -45, -40, 17, -65, 20, -39, -55, 15)
y_mat[2,] <- c(110, -10, 80, 55, -55, 110, 90, -28, 110)
y_mat[3,] <- c(4, -40, 31, -23, -64, -20, -30, -61, 54)

# Make the data into a vector (in row order):
y <- as.vector(t(y_mat))

# Set up the factors to match the data:
A <- as.factor(rep(rep(1:num_a, each=num_b), times=num_c))
B <- as.factor(rep(1:num_b, times=num_a*num_c))
C <- as.factor(rep(1:num_c, each=num_a*num_b))

# Confirm that the data in R is the same as the problem:
cbind(A, B, C, y)
```

```
##       A B C   y
##  [1,] 1 1 1 -35
```

```
##  [2,] 1 2 1 -45
##  [3,] 1 3 1 -40
##  [4,] 2 1 1  17
##  [5,] 2 2 1 -65
##  [6,] 2 3 1  20
##  [7,] 3 1 1 -39
##  [8,] 3 2 1 -55
##  [9,] 3 3 1  15
## [10,] 1 1 2 110
## [11,] 1 2 2 -10
## [12,] 1 3 2  80
## [13,] 2 1 2  55
## [14,] 2 2 2 -55
## [15,] 2 3 2 110
## [16,] 3 1 2  90
## [17,] 3 2 2 -28
## [18,] 3 3 2 110
## [19,] 1 1 3   4
## [20,] 1 2 3 -40
## [21,] 1 3 3  31
## [22,] 2 1 3 -23
## [23,] 2 2 3 -64
## [24,] 2 3 3 -20
## [25,] 3 1 3 -30
## [26,] 3 2 3 -61
## [27,] 3 3 3  54
```

```
# Run ANOVA in R to get the results:
summary.aov(lm(y~A*B + A*C + B*C))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## A            2    480     240   0.506 0.620709
## B            2  36474   18237  38.481 7.86e-05 ***
## C            2  31634   15817  33.375 0.000131 ***
## A:B          4   3399     850   1.793 0.223418
## A:C          4   3729     932   1.967 0.192726
## B:C          4   7626    1906   4.023 0.044649 *
## Residuals    8   3791     474
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As calculated above, at the 5% level of significance, there appears to be a B effect (speed effect), C effect (pressure effect), and lastly an interaction between treatment B and treatment C (a speed-pressure interaction). On the other hand, we have no evidence to suggest that there exists an A effect (nozzle effect), an interaction between treatment A and treatment B, and lastly an interaction between treatment A and treatment C.

Lecture 17 Problem 2b:
In this part of the problem, we will consider only the portion of the data that would follow if we had followed a Latin Square Design, of the form

$$\begin{pmatrix} 3 & 1 & 2 \\ 2 & 3 & 1 \\ 1 & 2 & 3 \end{pmatrix}$$

Where the factor A is on the columns and factor B is on the rows (each with 3 levels). Our goal in this problem is to develop an additive model, produce the ANOVA table, and state our conclusions (using $\alpha = 0.05$). In the end we want to discover which effects are significant, and which are not. We will do all of the analysis in R and then draw conclusions, this work in R is shown below.

```
# Set up the number of each factor:
p <- 3

# Set up an empty matrix for the data:
y_mat <- matrix(nrow=p, ncol=p)

# Fill in the matrix with the data from the problem:
y_mat[1,] <- c(4, 17, 90)
y_mat[2,] <- c(-10, -64, -55)
y_mat[3,] <- c(-40, 110, 54)

# Make the data into a vector (in row order):
y <- as.vector(t(y_mat))

# Set up the factors to match the data:
A <- as.factor(rep(1:p, times=p))
B <- as.factor(rep(1:p, each=p))
C <- as.factor(c(3, 1, 2, 2, 3, 1, 1, 2, 3))

# Confirm that the data in R is the same as the problem:
cbind(A, B, C, y)
```

```
##       A B C   y
##  [1,] 1 1 3   4
##  [2,] 2 1 1  17
##  [3,] 3 1 2  90
##  [4,] 1 2 2 -10
##  [5,] 2 2 3 -64
##  [6,] 3 2 1 -55
##  [7,] 1 3 1 -40
##  [8,] 2 3 2 110
##  [9,] 3 3 3  54
```

```
# Run ANOVA in R to get the results:
summary.aov(lm(y~A+B+C))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## A            2   3420    1710   2.809 0.2625
## B            2  13531    6765  11.113 0.0826 .
## C            2  12825    6412  10.533 0.0867 .
## Residuals    2   1218     609
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As calculated above, at the 5% level of significance, none of the p-values are below this threshold and hence we have no evidence to suggest that there exists an A effect (nozzle effect), a B effect (speed effect), or a C effect (pressure effect). However, like the model above, the smallest p-values/most significant factors are owned by the factors B and C (both significant under a significance-level of 0.1).

Lecture 17 Problem 2c:
In this part of the problem, for the LSD above, we will consider a model that includes $A + B + C$ but also a single 2-way interaction term. Furthermore, we will develop the model and produce the ANOVA table. From there we will state what comments we can make about this table. Since the two factors that were the most significant in the LSD above were B and C, it makes sense to include the BC interaction term as our single two-way interaction term in our new model. The ANOVA calculations are done in R below.

```
# Run ANOVA in R to get the results:
summary.aov(lm(y~A+B+C+B:C))
```

```
##             Df Sum Sq Mean Sq
## A            2   3420    1710
## B            2  13531    6765
## C            2  12825    6412
## B:C          2   1218     609
```

As can be seen from our LSD model with a two-way interaction term, the sum of square and mean square of our error/residuals is missing (i.e. SS_E and MS_E are not included). Since the error of the residuals is missing, we cannot compute observed F statistics, p-values, and thus we can't test any of our main or interaction effects. The reason why SS_E and MS_E are missing can be seen by looking at the Df column of the ANOVA table. Since estimation of these effects takes up all of our degrees of freedom, this model simply does not have enough degrees of freedom to estimate and test all of them. Thus when we give all of our degrees of freedom to the parameters we want to estimate, there are no degrees of freedom left for the residuals, and hence no testing can be done.

Lecture 18 Problem 1:
For the full model involving two factors A, B, if we consider the A effect, the B effect, and the interaction effect, all written in Yates' notation, we showed in class that the product of the contrast vector for A, and for B, is equal to the contrast vector of the interaction effect. In this problem, we will show that the product of the main effect for A, and for B, is not equal to the interaction effect.

As shown in class, for a 2^2 design, the A main effect is written in Yates' notation as $\frac{1}{2n}[-(1) + a - b + ab]$, the B main effect is written in Yates' notation as $\frac{1}{2n}[-(1) - a + b + ab]$, and lastly the AB interaction effect is written in Yates' notation as $\frac{1}{2n}[(1) - a - b + ab]$. We will now show that the product of the A main effect with the B main effect is not equal to the AB interaction effect. This calculation is shown below.

$$\begin{aligned}
\text{A effect} \cdot \text{B effect} &= \frac{1}{2n}[-(1) + a - b + ab] \cdot \frac{1}{2n}[-(1) - a + b + ab] \\
&= \frac{1}{4n^2}[(1)^2 + (1) \cdot a - (1) \cdot b - (1) \cdot ab - a \cdot (1) - a^2 + a \cdot b + a \cdot ab + b \cdot (1) + b \cdot a - b^2 \\
&\qquad - b \cdot ab - ab \cdot (1) - ab \cdot a + ab \cdot b + (ab)^2] \\
&= \frac{1}{4n^2}\left[(1)^2 - 2(1)(ab) - a^2 + 2(a)(b) - b^2 + (ab)^2\right] \\
&= \frac{1}{4n^2}\left[(1)^2 - 2(1)(ab) + (ab)^2 - a^2 + 2(a)(b) - b^2\right] \\
&= \frac{1}{4n^2}\left[(ab - (1))^2 - (a - b)^2\right]
\end{aligned}$$

As seen above, we get that the AB interaction effect is written in Yates' notation as $\frac{1}{4n^2}\left[(ab - (1))^2 - (a - b)^2\right]$, however, we have already shown that the AB interaction effect is actually written in Yates' notation as $\frac{1}{2n}[(1) - a - b + ab]$. Hence we can see that these two terms are not equal to each other, and thus the product of the main effect for A, and for B, is not equal to the interaction effect.

Lecture 18 Problem 2a:
As shown in class, the two ways of computing effects are: the differences of 2 averages over y,

and the difference of conditional effects. For example, in a 2^3 design, the A effect is computed as (avg. y|A=+) $-$ (avg. y|A=-) which is the same as computing $\frac{1}{2}\left[(A|B=+)+(A|B=-)\right]$. Furthermore, the AB effect is computed as (avg. y|one diagonal) $-$ (avg. y|the other diagonal) which is the same as computing $\frac{1}{2}\left[(A|B=+)-(A|B=-)\right]$.

In this problem, using the fact that the ABC effect can be defined, similarly, in terms of effects as $\frac{1}{2}\left[(AB|C=+)-(AB|C=-)\right]$, we will write the ABC effect in Yates' notation. This calculation is shown below.

$$\begin{aligned}
\text{ABC Effect} &= \frac{1}{2}\left[(AB|C=+)-(AB|C=-)\right] \\
&= \frac{1}{2}\left[\frac{1}{2}\left[(A|B=+)-(A|B=-)|C=+\right]-\frac{1}{2}\left[(A|B=+)-(A|B=-)|C=-\right]\right] \\
&= \frac{1}{4}\left[(A|B=+,C=+)-(A|B=-,C=+)-(A|B=+,C=-)+(A|B=-,C=-)\right] \\
&= \frac{1}{4n}\left[(abc-bc)-(ac-c)-(ab-b)+(a-(1))\right] \\
&= \frac{1}{4n}\left[abc-bc-ac+c-ab+b+a-(1)\right] \\
&= \frac{1}{4n}\left[-(1)+a+b-ab+c-ac-bc+abc\right]
\end{aligned}$$

Hence, as shown above, the ABC effect in Yates' notation is computed as $\frac{1}{4n}\left[-(1)+a+b-ab+c-ac-bc+abc\right]$.

Lecture 18 Problem 2b:
In this problem, using the fact that the CBA effect can be defined, similarly, in terms of effects as $\frac{1}{2}\left[(CB|A=+)-(CB|A=-)\right]$, we will write the CBA effect in Yates' notation. This calculation is shown below.

$$\begin{aligned}
\text{CBA Effect} &= \frac{1}{2}\left[(CB|A=+)-(CB|A=-)\right] \\
&= \frac{1}{2}\left[\frac{1}{2}\left[(C|B=+)-(C|B=-)|A=+\right]-\frac{1}{2}\left[(C|B=+)-(C|B=-)|A=-\right]\right] \\
&= \frac{1}{4}\left[(C|B=+,A=+)-(C|B=-,A=+)-(C|B=+,A=-)+(C|B=-,A=-)\right] \\
&= \frac{1}{4n}\left[(abc-ab)-(ac-a)-(bc-b)+(c-(1))\right] \\
&= \frac{1}{4n}\left[abc-ab-ac+a-bc+b+c-(1)\right] \\
&= \frac{1}{4n}\left[-(1)+a+b-ab+c-ac-bc+abc\right]
\end{aligned}$$

Hence, as shown above, the CBA effect in Yates' notation is computed as $\frac{1}{4n}\left[-(1)+a+b-ab+c-ac-bc+abc\right]$. This is the same as the ABC effect in Yates' notation, as expected.

Lecture 18 Problem 3a:
In class, for a 2^3 design, we were able to write the A effect, AB effect, and ABC effect in Yates' notation. In this problem, using the two methods for computing effects, we will write the BC effect in Yates' notation.

This calculation is shown below.

$$
\begin{aligned}
\text{BC Effect} &= \frac{1}{2}\left[(B|C=+)-(B|C=-)\right] \\
&= \frac{1}{2}\left[[(\text{avg. y|B=+})-(\text{avg. y|B=-})|C=+]-[(\text{avg. y|B=+})-(\text{avg. y|B=-})|C=-]\right] \\
&= \frac{1}{2}\left[(\text{avg. y|B=+,C=+})-(\text{avg. y|B=-,C=+})-(\text{avg. y|B=+,C=-})+(\text{avg. y|B=-,C=-})\right] \\
&= \frac{1}{2}\left[\bar{y}_{.22.}-\bar{y}_{.12.}-\bar{y}_{.21.}+\bar{y}_{.11.}\right] \\
&= \frac{1}{2n}\left[y_{.22.}-y_{.12.}-y_{.21.}+y_{.11.}\right] \\
&= \frac{1}{2n}\left[\frac{1}{2}(y_{122.}+y_{222.})-\frac{1}{2}(y_{112.}+y_{212.})-\frac{1}{2}(y_{121.}+y_{221.})+\frac{1}{2}(y_{111.}+y_{211.})\right] \\
&= \frac{1}{4n}\left[y_{122.}+y_{222.}-y_{112.}-y_{212.}-y_{121.}-y_{221.}+y_{111.}+y_{211.}\right] \\
&= \frac{1}{4n}\left[bc+abc-c-ac-b-ab+(1)+a\right] \\
&= \frac{1}{4n}\left[(1)+a-b-ab-c-ac+bc+abc\right]
\end{aligned}
$$

Hence, as shown above, the BC effect in Yates' notation is computes as $\frac{1}{4n}\left[(1)+a-b-ab-c-ac+bc+abc\right]$.

Lecture 18 Problem 3b:

In this problem, we will multiply the contrast vector for the BC effect by the contrast vector for the ABC effect and identify the corresponding effect. To start off, as computed in the previous problem, the contrast vector for the BC effect is written as $\vec{C}_{BC} = (1,1,-1,-1,-1,-1,1,1)$. Furthermore, as show in class, the ABC effect is written in Yates' notation as $\frac{1}{4n}\left[abc+c-bc-ac-ab-(1)+b+a\right]$. Hence the contrast vector for the ABC effect is written as $\vec{C}_{ABC} = (-1,1,1,-1,1,-1,-1,1)$. Multiplying these two vectors together, we obtain the following contrast vector $\vec{C} = \vec{C}_{BC} * \vec{C}_{BC} = (1,1,-1,-1,-1,-1,1,1) * (-1,1,1,-1,1,-1,-1,1) = (-1,1,-1,1,-1,1,-1,1)$. Lastly, as shown in class, the A effect is written in Yates' notation as $\frac{1}{4n}\left[a+ab+ac+abc-(1)-b-c-bc\right]$. Hence the contrast vector for the A effect is written as $\vec{C}_A = (-1,1,-1,1,-1,1,-1,1)$. Therefore, after multiplying the contrast vector for the BC effect by the contrast vector for the ABC effect, we end up with the contrast vector for the A effect.