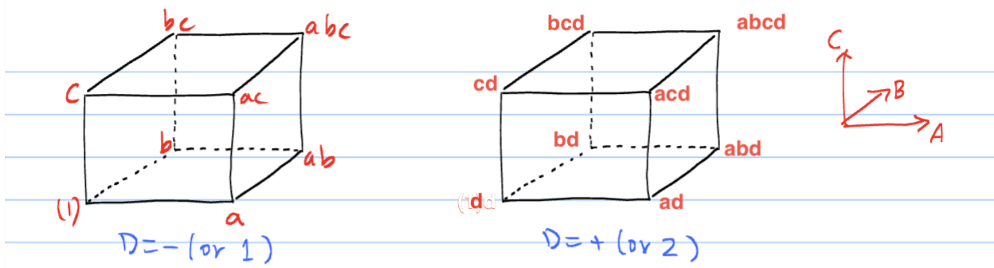
STAT 421 Homework 8

Lecture 19 Problem 1a
In this problem, we will consider a 2^4 experiment with n replications. The goal of this problem is to label the vertices of the $D = +$ (or 2) cube according to Yates. This labeling is done below.



Lecture 19 Problem 1b
In this problem, we will write the AD effect in Yates' notation starting from $AD = \frac{1}{2}[(A|D = +) - (A|D = -)]$. This computation is done below.

$$\begin{aligned}
\text{AD Effect} &= \frac{1}{2}\left[(A|D = +) - (A|D = -)\right] \\
&= \frac{1}{2}\left[[(\text{avg. y}|\text{A=+}) - (\text{avg. y}|\text{A=-})|D = +] - [(\text{avg. y}|\text{A=+}) - (\text{avg. y}|\text{A=-})|D = -]\right] \\
&= \frac{1}{2}\left[(\text{avg. y}|\text{A=+,D=+}) - (\text{avg. y}|\text{A=-,D=+}) - (\text{avg. y}|\text{A=+,D=-}) + (\text{avg. y}|\text{A=-,D=-})\right] \\
&= \frac{1}{2}\left[\bar{y}_{2..2.} - \bar{y}_{1..2.} - \bar{y}_{2..1.} + \bar{y}_{1..1.}\right] \\
&= \frac{1}{2n}\left[y_{2..2.} - y_{1..2.} - y_{2..1.} + y_{1..1.}\right] \\
&= \frac{1}{2n}[\frac{1}{4}(y_{2222.} + y_{2212.} + y_{2122.} + y_{2112.}) - \frac{1}{4}(y_{1222.} + y_{1212.} + y_{1122.} + y_{1112.}) \\
&\quad - \frac{1}{4}(y_{2221.} + y_{2211.} + y_{2121.} + y_{2111.}) + \frac{1}{4}(y_{1221.} + y_{1211.} + y_{1121.} + y_{1111.})] \\
&= \frac{1}{8n}[y_{2222.} + y_{2212.} + y_{2122.} + y_{2112.} - y_{1222.} - y_{1212.} - y_{1122.} - y_{1112.} - y_{2221.} \\
&\quad - y_{2211.} - y_{2121.} - y_{2111.} + y_{1221.} + y_{1211.} + y_{1121.} + y_{1111.}] \\
&= \frac{1}{8n}\left[abcd + abd + acd + ad - bcd - bd - cd - d - abc - ab - ac - a + bc + b + c + (1)\right] \\
&= \frac{1}{8n}\left[(1) - a + b - ab + c - ac + bc - abc - d + ad - bd + abd - cd + acd - bcd + abcd\right]
\end{aligned}$$

Hence, as shown above, the AD effect in Yates' notation is computed as $\frac{1}{8n}[(1) - a + b - ab + c - ac - bc - abc - d + ad - bd + abd - cd + acd - bcd + abcd]$.

Lecture 19 Problem 1c
In this problem, we will generate the $\pm$ table, including only the A, B, C, D, and AD columns. Furthermore, we will label each row with the appropriate Yates "basis." This table is shown below.

	A	B	C	D	AD
(1)	-	-	-	-	+
a	+	-	-	-	-
b	-	+	-	-	+
ab	+	+	-	-	-
c	-	-	+	-	+
ac	+	-	+	-	-
bc	-	+	+	-	+
abc	+	+	+	-	-
d	-	-	-	+	-
ad	+	-	-	+	+
bd	-	+	-	+	-
abd	+	+	-	+	+
cd	-	-	+	+	-
acd	+	-	+	+	+
bcd	-	+	+	+	-
abcd	+	+	+	+	+

As can be seen above, the contrast vector for the AD effect found from the $\pm$ table is equal to the contrast vector for the AD effect that we found in the previous part, as expected.

Lecture 19 Problem 2
In this problem, we will consider the full 2^2 replicated model: $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$. The goal of this problem is to confirm that SS_{AB}, defined as $SS_{AB} = \sum_i^2 \sum_j^2 \sum_k^n (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$, can be written in terms of (contrast AB)^2. In particular, considering only the term in SS_{AB} where $i = 1$ and $j = 1$, we must show that it is equal to $\frac{1}{16n}$(contrast AB)^2. This computation is done below.

$$
\begin{aligned}
SS_{AB} &= \sum_i^2 \sum_j^2 \sum_k^n (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 \\
&= n \sum_i^2 \sum_j^2 \left(\frac{1}{n} y_{ij.} - \frac{1}{2n} y_{i..} - \frac{1}{2n} y_{.j.} + \frac{1}{4n} y_{...} \right)^2
\end{aligned}
$$

This sum of squares has four different squared terms, one for each i and j pair, however we will only consider the term where $i=1$ and $j=1$. We will continue the computation below.

$$\begin{aligned}
n\left(\frac{1}{n}y_{11.}-\frac{1}{2n}y_{1..}-\frac{1}{2n}y_{.1.}+\frac{1}{4n}y_{...}\right)^2 &= \frac{n}{n^2}\left(y_{11.}-\frac{1}{2}y_{1..}-\frac{1}{2}y_{.1.}+\frac{1}{4}y_{...}\right)^2\\
&= \frac{1}{n}\left(y_{11.}-\frac{1}{2}(y_{11.}+y_{12.})-\frac{1}{2}(y_{11.}+y_{21.})+\frac{1}{4}(y_{11.}+y_{12.}+y_{21.}+y_{22.})\right)^2\\
&= \frac{1}{n}\left(y_{11.}-\frac{1}{2}y_{11.}-\frac{1}{2}y_{12.}-\frac{1}{2}y_{11.}-\frac{1}{2}y_{21.}+\frac{1}{4}y_{11.}+\frac{1}{4}y_{12.}+\frac{1}{4}y_{21.}+\frac{1}{4}y_{22.}\right)^2\\
&= \frac{1}{n}\left(\frac{1}{4}y_{11.}-\frac{1}{4}y_{12.}-\frac{1}{4}y_{21.}+\frac{1}{4}y_{22.}\right)^2\\
&= \frac{1}{16n}\left(y_{11.}-y_{12.}-y_{21.}+y_{22.}\right)^2\\
&= \frac{1}{16n}\left((1)-b-a+ab\right)^2\\
&= \frac{1}{16n}\left((1)-a-b+ab\right)^2
\end{aligned}$$

As shown in the previous lecture, the contrast for the AB effect is $(1)-a-b+ab$. Hence, we have shown that considering only the term in SS_{AB} where $i=1$ and $j=1$, it is equal to $\frac{1}{16n}(\text{contrast AB})^2$.

Lecture 19 Problem 3a

In this problem, we will consider a 2^1 design with the model $y_{ij}=\mu+\alpha_i+\epsilon_{ij},\ i=1,\ldots,a;\ j=1,\ldots,n$. The goal of this problem is to show that SS_A is proportional to the square of a contrast in (1) and a. Furthermore we want to say what this contrast is. We will start with $SS_A=\sum_i^a\sum_j^n(\bar{y}_{i.}-\bar{y}_{..})^2$, then we will also show this result holds with $\frac{1}{n}\sum_i y_{i.}^2-\frac{1}{an}y_{..}^2$. Considering that this is a 2^1 design, i goes from 1 to 2 thus our formulas become: $SS_A=\sum_i^2\sum_j^n(\bar{y}_{i.}-\bar{y}_{..})^2$ and $\frac{1}{n}\sum_i y_{i.}^2-\frac{1}{2n}y_{..}^2$. We will start with SS_A, this computation is done below.

$$
\begin{aligned}
SS_A &= \sum_i^2 \sum_j^n (\bar{y}_{i.} - \bar{y}_{..})^2 \\
&= n \sum_i^2 (\bar{y}_{i.} - \bar{y}_{..})^2 \\
&= n \sum_i^2 \left(\frac{1}{n} y_{i.} - \frac{1}{2n} y_{..} \right)^2 \\
&= \frac{n}{n^2} \sum_i^2 \left(y_{i.} - \frac{1}{2} y_{..} \right)^2 \\
&= \frac{1}{n} \left[\left(y_{1.} - \frac{1}{2}(y_{1.} + y_{2.}) \right)^2 + \left(y_{2.} - \frac{1}{2}(y_{1.} + y_{2.}) \right)^2 \right] \\
&= \frac{1}{n} \left[\left(y_{1.} - \frac{1}{2} y_{1.} - \frac{1}{2} y_{2.} \right)^2 + \left(y_{2.} - \frac{1}{2} y_{1.} - \frac{1}{2} y_{2.} \right)^2 \right] \\
&= \frac{1}{n} \left[\left(\frac{1}{2} y_{1.} - \frac{1}{2} y_{2.} \right)^2 + \left(\frac{1}{2} y_{2.} - \frac{1}{2} y_{1.} \right)^2 \right] \\
&= \frac{1}{4n} \left[(y_{1.} - y_{2.})^2 + (y_{2.} - y_{1.})^2 \right] \\
&= \frac{2}{4n} (y_{2.} - y_{1.})^2 \\
&= \frac{1}{2n} (a - (1))^2 \\
&= \frac{1}{2n} (-(1) + a)^2
\end{aligned}
$$

Hence as shown above, starting from SS_A, we showed that SS_A is proportional to the square of a contrast in (1) and a. In particular this contrast is $-(1)+a$, which is the A effect. We will now start with $\frac{1}{n}\sum_i y_{i.}^2 - \frac{1}{2n} y_{..}^2$ and see if we get the same result, this computation is done below.

$$
\begin{aligned}
\frac{1}{n} \sum_i y_{i.}^2 - \frac{1}{2n} y_{..}^2 &= \frac{1}{n} \left[\sum_i y_{i.}^2 - \frac{1}{2} y_{..}^2 \right] \\
&= \frac{1}{n} \left[y_{1.}^2 + y_{2.}^2 - \frac{1}{2} (y_{1.} + y_{2.})^2 \right] \\
&= \frac{1}{n} \left[y_{1.}^2 + y_{2.}^2 - \frac{1}{2} (y_{1.} + y_{2.} + 2 y_{1.} y_{2.}) \right] \\
&= \frac{1}{n} \left[y_{1.}^2 + y_{2.}^2 - \frac{1}{2} y_{1.} - \frac{1}{2} y_{2.} - y_{1.} y_{2.} \right] \\
&= \frac{1}{n} \left[\frac{1}{2} y_{1.}^2 + \frac{1}{2} y_{2.}^2 - y_{1.} y_{2.} \right] \\
&= \frac{1}{2n} \left[y_{1.}^2 + y_{2.}^2 - 2 y_{1.} y_{2.} \right] \\
&= \frac{1}{2n} (y_{2.} - y_{1.})^2 \\
&= \frac{1}{2n} (a - (1))^2 \\
&= \frac{1}{2n} (-(1) + a)^2
\end{aligned}
$$

Hence as shown above, starting from $\frac{1}{n}\sum_i y_{i.}^2 - \frac{1}{2n}y_{..}^2$ we showed that SS_A is proportional to the square of a contrast in (1) and a. In particular this contrast is $-(1)+a$, which is the A effect. This is the same contrast as the one we got when we started from SS_A, which makes sense as $\frac{1}{n}\sum_i y_{i.}^2 - \frac{1}{2n}y_{..}^2$ is a different way to compute SS_A.

Lecture 19 Problem 3b

In this problem, we will show/explain if SS_E can be written in terms of a contrast squared. Similar to our approach from last problem, we will start with the definition of SS_E and see if we can get it in in terms of a contrast squared. This calculation is done below.

$$\begin{aligned}
SS_E &= \sum_i^2 \sum_j^n \hat{\epsilon}_{ij}^2 \\
&= \sum_i^2 \sum_j^n (y_{ij} - \hat{y}_{ij})^2 \\
&= \sum_i^2 \sum_j^n (y_{ij} - \bar{y}_{i.})^2 \\
&= \sum_i^2 \sum_j^n \left(y_{ij} - \frac{1}{n} y_{i.} \right)^2 \\
&= \sum_i^2 \sum_j^n \left(y_{ij}^2 + \frac{1}{n^2} y_{i.}^2 - \frac{2}{n} y_{ij} y_{i.} \right) \\
&= \sum_i^2 \sum_j^n y_{ij}^2 + \frac{1}{n^2} \sum_i^2 \sum_j^n y_{i.}^2 - \frac{2}{n} \sum_i^2 \sum_j^n y_{ij} y_{i.} \\
&= \sum_i^2 \sum_j^n y_{ij}^2 + \frac{n}{n^2} \sum_i^2 y_{i.}^2 - \frac{2}{n} \sum_i^2 y_{i.} \sum_j^n y_{ij} \\
&= \sum_i^2 \sum_j^n y_{ij}^2 + \frac{1}{n} \sum_i^2 y_{i.}^2 - \frac{2}{n} \sum_i^2 y_{i.} y_{i.} \\
&= \sum_i^2 \sum_j^n y_{ij}^2 + \frac{1}{n} \sum_i^2 y_{i.}^2 - \frac{2}{n} \sum_i^2 y_{i.}^2 \\
&= \sum_i^2 \sum_j^n y_{ij}^2 - \frac{1}{n} \sum_i^2 y_{i.}^2 \\
&= \sum_j^n (y_{1j}^2 + y_{2j}^2) - \frac{1}{n}(y_{1.}^2 + y_{2.}^2)
\end{aligned}$$

Since we are dealing with a 2^1 design, the only possible contrasts are $(1) - a$ and $-(1) + a$, given that there isn't anymore non-trivial simplification that we can do to the above expression, and the fact that it is currently not able to be written in contrast squared form, we conclude that SS_E can not be written in terms of a contrast squared.

Lecture 20 Problem 1a:

In this problem, we will consider the data from problem 6.15 in the textbook. In this experiment, a nickel–titanium alloy is used to make components for jet turbine aircraft engines. It turns out that cracking is a potentially serious problem in the final part because it can lead to nonrecoverable failure. Therefore, a test is run at the parts producer to determine the effect of four factors on cracks. The four factors are pouring temperature (A), titanium content (B), heat treatment method (C), and amount of grain refiner used (D). Two replicates of a 2^4 design are run, and the length of crack (in $mm \times 10^{-2}$) induced in a sample coupon subjected to a standard test is measured. The data for this experiment are shown below.

■ **TABLE P6.2**
The Experiment for problem 6.15

A	*B*	*C*	*D*	Treatment Combination	Replicate I	Replicate II
−	−	−	−	(1)	7.037	6.376
+	−	−	−	*a*	14.707	15.219
−	+	−	−	*b*	11.635	12.089
+	+	−	−	*ab*	17.273	17.815
−	−	+	−	*c*	10.403	10.151
+	−	+	−	*ac*	4.368	4.098
−	+	+	−	*bc*	9.360	9.253
+	+	+	−	*abc*	13.440	12.923
−	−	−	+	*d*	8.561	8.951
+	−	−	+	*ad*	16.867	17.052
−	+	−	+	*bd*	13.876	13.658
+	+	−	+	*abd*	19.824	19.639
−	−	+	+	*cd*	11.846	12.337
+	−	+	+	*acd*	6.125	5.904
−	+	+	+	*bcd*	11.190	10.935
+	+	+	+	*abcd*	15.653	15.053

The goal of this problem is to develop a full model, and generate the ANOVA table. Furthermore, we will find out and present which parameters/effects are significant. These computations are done in R below.

```
# Set up important numbers for calculations:
num_rep <- 2
num_sums <- 16
n <- 2
k <- 4

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

# Fill in the matrix with the data from the problem:
y_mat[1,] <- c(7.037, 14.707, 11.635, 17.273, 10.403, 4.368, 9.360, 13.440,
               8.561, 16.867, 13.876, 19.824, 11.846, 6.125, 11.190, 15.653)
y_mat[2,] <- c(6.376, 15.219, 12.089, 17.815, 10.151, 4.098, 9.253, 12.923,
               8.951, 17.052, 13.658, 19.639, 12.337, 5.904, 10.935, 15.053)

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

# Set up the factors to match the data:
design <- gen.factorial(2,5,varNames=c("A","B","C","D","R"), factors="all")
attach(design)
```

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

```
##       A B C D R      y
##  [1,] 1 1 1 1 1  7.037
##  [2,] 2 1 1 1 1 14.707
##  [3,] 1 2 1 1 1 11.635
##  [4,] 2 2 1 1 1 17.273
##  [5,] 1 1 2 1 1 10.403
##  [6,] 2 1 2 1 1  4.368
##  [7,] 1 2 2 1 1  9.360
##  [8,] 2 2 2 1 1 13.440
##  [9,] 1 1 1 2 1  8.561
## [10,] 2 1 1 2 1 16.867
## [11,] 1 2 1 2 1 13.876
## [12,] 2 2 1 2 1 19.824
## [13,] 1 1 2 2 1 11.846
## [14,] 2 1 2 2 1  6.125
## [15,] 1 2 2 2 1 11.190
## [16,] 2 2 2 2 1 15.653
## [17,] 1 1 1 1 2  6.376
## [18,] 2 1 1 1 2 15.219
## [19,] 1 2 1 1 2 12.089
## [20,] 2 2 1 1 2 17.815
## [21,] 1 1 2 1 2 10.151
## [22,] 2 1 2 1 2  4.098
## [23,] 1 2 2 1 2  9.253
## [24,] 2 2 2 1 2 12.923
## [25,] 1 1 1 2 2  8.951
## [26,] 2 1 1 2 2 17.052
## [27,] 1 2 1 2 2 13.658
## [28,] 2 2 1 2 2 19.639
## [29,] 1 1 2 2 2 12.337
## [30,] 2 1 2 2 2  5.904
## [31,] 1 2 2 2 2 10.935
## [32,] 2 2 2 2 2 15.053
```

```r
# Save model 1 into a variable and save the MSE for later:
model_1 <- lm(y~A*B*C*D)
MSE <- summary.aov(model_1)[[1]][16,3]

# Run ANOVA in R to get the results:
summary.aov(model_1)
```

```
##             Df Sum Sq Mean Sq  F value   Pr(>F)
## A            1  72.91   72.91  898.339 1.74e-15 ***
## B            1 126.46  126.46 1558.172  < 2e-16 ***
## C            1 103.46  103.46 1274.822  < 2e-16 ***
## D            1  30.66   30.66  377.802 1.49e-12 ***
## A:B          1  29.93   29.93  368.739 1.79e-12 ***
## A:C          1 128.50  128.50 1583.256  < 2e-16 ***
## B:C          1   0.07    0.07    0.908    0.355
## A:D          1   0.05    0.05    0.577    0.459
```

```
## B:D          1   0.02    0.02    0.220    0.645
## C:D          1   0.05    0.05    0.583    0.456
## A:B:C        1  78.75   78.75  970.325 9.49e-16 ***
## A:B:D        1   0.08    0.08    0.947    0.345
## A:C:D        1   0.00    0.00    0.036    0.852
## B:C:D        1   0.01    0.01    0.125    0.728
## A:B:C:D      1   0.00    0.00    0.020    0.890
## Residuals   16   1.30    0.08
## ---
## 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 an A effect (pairing temperature effect), B effect (titanium content effect), C effect (heat treatment method effect), D effect (grain refiner effect), AB interaction effect, AC interaction effect, and an ABC three-way interaction effect. On the other hand, we have no evidence to suggest that any other effect exists.

Lecture 20 Problem 1b:
In this problem, we will develop a full model and generate estimates of the parameters, as well as the corresponding results from t-tests. Furthermore, we will mention which of these parameters are significant. These computations are done in R below.

```r
# Define Helmert's contrast for use in model 2:
contr <- as.character("contr.helmert")

# Save model 2 into a variable:
model_2 <- lm(formula=y~A*B*C*D, contrasts = list(A=contr, B=contr, C=contr, D=contr))

# Display the results of the linear model:
summary(model_2)
```

```
##
## Call:
## lm(formula = y ~ A * B * C * D, contrasts = list(A = contr, B = contr,
##     C = contr, D = contr))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.3305 -0.1500  0.0000  0.1500  0.3305
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.988062   0.050361 238.042  < 2e-16 ***
## A1           1.509438   0.050361  29.972 1.74e-15 ***
## B1           1.987937   0.050361  39.474  < 2e-16 ***
## C1          -1.798125   0.050361 -35.705  < 2e-16 ***
## D1           0.978875   0.050361  19.437 1.49e-12 ***
## A1:B1        0.967062   0.050361  19.203 1.79e-12 ***
## A1:C1       -2.003875   0.050361 -39.790  < 2e-16 ***
## B1:C1        0.048000   0.050361   0.953    0.355
## A1:D1        0.038250   0.050361   0.760    0.459
## B1:D1        0.023625   0.050361   0.469    0.645
## C1:D1       -0.038438   0.050361  -0.763    0.456
## A1:B1:C1     1.568750   0.050361  31.150 9.49e-16 ***
## A1:B1:D1     0.049000   0.050361   0.973    0.345
```

```
## A1:C1:D1     0.009563   0.050361   0.190    0.852
## B1:C1:D1     0.017813   0.050361   0.354    0.728
## A1:B1:C1:D1  0.007062   0.050361   0.140    0.890
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2849 on 16 degrees of freedom
## Multiple R-squared:  0.9977, Adjusted R-squared:  0.9956
## F-statistic:   469 on 15 and 16 DF,  p-value: < 2.2e-16
```

As calculated above, at the 5% level of significance, there appears to be an A effect (pairing temperature effect), B effect (titanium content effect), C effect (heat treatment method effect), D effect (grain refiner effect), AB interaction effect, AC interaction effect, and an ABC three-way interaction effect. On the other hand, we have no evidence to suggest that any other effect exists. These match the results from the previous part. As mentioned in class, the estimated effects are actually two times the parameter estimates shown in the above output, thus we must adjust them accordingly before reporting. This adjustment is done below.

```r
# Correctly adjust the parameter estimates:
eff <- 2*model_2$coef

# Remove the grand mean:
eff <- eff[2:length(eff)]

# Display the adjusted parameter estimates:
as.matrix(eff, col=1)
```

```
##                   [,1]
## A1            3.018875
## B1            3.975875
## C1           -3.596250
## D1            1.957750
## A1:B1         1.934125
## A1:C1        -4.007750
## B1:C1         0.096000
## A1:D1         0.076500
## B1:D1         0.047250
## C1:D1        -0.076875
## A1:B1:C1      3.137500
## A1:B1:D1      0.098000
## A1:C1:D1      0.019125
## B1:C1:D1      0.035625
## A1:B1:C1:D1   0.014125
```

As shown above the values of the estimated A, B, C, and D effects are 3.018875, 3.975875, -3.59625, and 1.95775, respectively. Furthermore, the values of the estimated AB, AC, BC, AD, BD, and CD interaction effects are 1.934125, -4.00775, 0.096, 0.0765, 0.04725, -0.076875, respectively. Also, the values of the estimated ABC, ABD, ACD, and BCD three-way interaction effects are 3.1375, 0.098, 0.019125, and 0.035625, respectively. Lastly, the value of the estimated ABCD four-way interaction effect is 0.014125.

Lecture 20 Problem 1c:
In this problem, we will use the method described in the problem description to compute all of the contrasts. This is computed in R and displayed below.

```r
# Set up the factors to match the data, omit factors="all" to get -1 and +1:
design <- gen.factorial(2,5,varNames=c("A","B","C","D","R"))
attach(design)

# Compute the estimated effects and store them in a vector (in the same order as
# the above effects):
contrasts <- c(sum(A*y), sum(B*y), sum(C*y), sum(D*y), sum(A*B*y), sum(A*C*y),
               sum(B*C*y), sum(A*D*y), sum(B*D*y), sum(C*D*y), sum(A*B*C*y),
               sum(A*B*D*y), sum(A*C*D*y), sum(B*C*D*y), sum(A*B*C*D*y))

# Display the estimated effects:
contrasts
```

```
##  [1]  48.302  63.614 -57.540  31.324  30.946 -64.124   1.536   1.224   0.756
## [10]  -1.230  50.200   1.568   0.306   0.570   0.226
```

As shown above the values of the A, B, C, and D contrasts are 48.302, 63.614, -57.540, and 31.324, respectively. Furthermore, the values of the AB, AC, BC, AD, BD, and CD contrasts are 30.946, -64.124, 1.536, 1.224, 0.756, and -1.230, respectively. Also, the values of the ABC, ABD, ACD, and BCD contrasts are 50.200, 1.568, 0.306, and 0.570, respectively. Lastly, the value of the ABCD contrast is 0.226.

Lecture 20 Problem 1d:
In this problem, we will use the above contrasts to compute the effects, t-ratios, and perform a t-test to and obtain a p-value for each of them. In each t-test we will use $N - p$ degrees of freedom, where $N = n2^k$ is the total amount of runs, and p is the number of parameters in the model. Furthermore, we will see if the effects, the t-ratios, and the p-values equal to the "by R" results in part b. All of these effects and t-tests are computed in R and displayed below.

```r
# Set up the number of parameters in the model:
p <- 16

# Compute the all of the effects:
effects <- contrasts / (n*2^(k-1))

# Obtain the t-ratios for all of the effects:
t_ratios <- effects / sqrt(MSE / (n*2^(k-2)))

# Obtain a p-value for all of the effects:
p_values <- 2*pt(abs(t_ratios), n*2^k-p, lower.tail=F)

# Display the all of the effects, t-ratios, and p-values:
cbind(effects, t_ratios, p_values)
```

```
##         effects    t_ratios     p_values
##  [1,]  3.018875  29.9723026 1.740225e-15
##  [2,]  3.975875  39.4736876 2.247046e-17
##  [3,] -3.596250 -35.7046559 1.100354e-16
##  [4,]  1.957750  19.4371332 1.485334e-12
##  [5,]  1.934125  19.2025770 1.790240e-12
##  [6,] -4.007750 -39.7901522 1.979949e-17
##  [7,]  0.096000   0.9531170 3.547091e-01
##  [8,]  0.076500   0.7595151 4.585895e-01
##  [9,]  0.047250   0.4691123 6.453180e-01
```

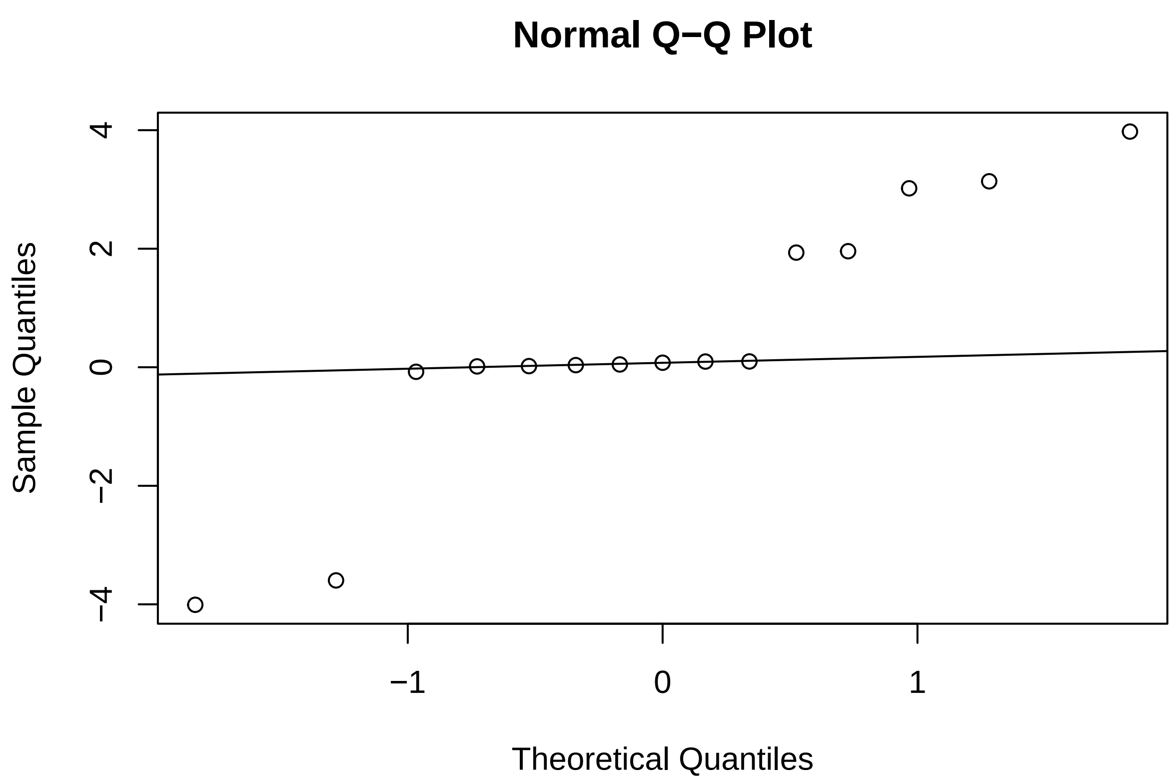
```
## [10,] -0.076875  -0.7632382 4.564284e-01
## [11,]  3.137500  31.1500474 9.485296e-16
## [12,]  0.098000   0.9729736 3.450474e-01
## [13,]  0.019125   0.1898788 8.517922e-01
## [14,]  0.035625   0.3536958 7.281845e-01
## [15,]  0.014125   0.1402373 8.902228e-01
```

As can be seen from the above output, the effects, the t-ratios, and the p-values are all equal to the "by R" results we found in part b.

Lecture 20 Problem 1e:
In this problem, the existence of replication allowed us to estimate σ_ϵ^2 (with MS_E), which in turn allowed us to perform tests on our estimated effects. If we suppose that we didn't have MS_E, then the goal of this problem is to use Daniel's idea to identify the significant effects. Furthermore, we will identify if they are the same effects as found via the F-test in part a and the t-tests in part b. Daniel's method uses a QQplot of the effects to assess significance, this qqplot is computed in R and displayed below.

```
# Create a qqplot of the effects:
qqnorm(effects)

# Create a straight line through the "straight line/middle portion" of the data:
abline(0.075, 0.1)
```



As seen from the above QQplot, the two smallest and five largest in magnitude effects are all considered significant due to their non-normality. These seven effects are the AC, C, AB, D, A, ABC, and B effects, respectively. Thus we can see that Daniel's method gives us the same results as found via the F-test in part a and the t-tests in part b.

Lecture 21 Problem 1:
In lecture 21, for the 2^3 design with four incomplete blocks, three specific block effects were picked. These effects are written below.

$$\left[\frac{\bar{B}_1+\bar{B}_2}{2}-\frac{\bar{B}_3+\bar{B}_4}{2}\right]=\frac{1}{2}\left[\frac{(1)+abc}{2}+\frac{a+bc}{2}-\frac{b+ac}{2}-\frac{c+ab}{2}\right]$$
$$\left[\frac{\bar{B}_1+\bar{B}_3}{2}-\frac{\bar{B}_2+\bar{B}_4}{2}\right]=\frac{1}{2}\left[\frac{(1)+abc}{2}+\frac{b+ac}{2}-\frac{a+bc}{2}-\frac{c+ab}{2}\right]$$
$$\left[\frac{\bar{B}_1+\bar{B}_4}{2}-\frac{\bar{B}_2+\bar{B}_3}{2}\right]=\frac{1}{2}\left[\frac{(1)+abc}{2}+\frac{c+ab}{2}-\frac{a+bc}{2}-\frac{b+ac}{2}\right]$$

Now suppose we use the block effect formula $\left(B_i-\bar{B}\right)$, where B_i denotes the average of the y values when the block factor is at the ith level, and $\bar{B}$ denotes the average of the B_i's. The goal of this problem is to show, for the specific blocking in this example, which treatment effect is confounded with the B_4-B_3 block effect. This treatment effect can be found by considering sums and/or differences of the effects shown in the lecture/above. As shown in lecture, the blocking that led to the above effects is displayed below.

$$\begin{aligned}&\text{Block 1}:[(1),\ abc]\\&\text{Block 2}:[a,\ bc]\\&\text{Block 3}:[b,\ ac]\\&\text{Block 4}:[c,\ ab]\end{aligned}$$

Now that we have the specific blocking and effects from the lecture example, we must find the B_4-B_3 effect written in Yates' notation. This computation is shown below.

$$\begin{aligned}B_4-B_3\text{ effect}&=(B_4-\bar{B})-(B_3-\bar{B})\\&=B_4-\bar{B}-B_3+\bar{B}\\&=B_4-B_3\\&=\frac{1}{2}(c+ab)-\frac{1}{2}(b+ac)\\&=\frac{1}{2}(ab+c-ac-b)\end{aligned}$$

Thus, we must add/subtract the above block effects until we find one that leads us to the contrast $\frac{1}{2}(ab+c-ac-b)$. If we look at the third effect from the lecture example, it has ab, c, ac, and b all with the correct signs. Furthermore, if we look at the second effect from the lecture example, it has ab, c, ac, and b all with the signs flipped, thus if we subtract the second effect from the third effect we should get the intended result. We will show this in two ways. The first of which involves only the left side of the block effect, and the second of which involves these block effects in Yates' notation. The first method is shown below.

$$\begin{aligned}\left[\frac{\bar{B}_1+\bar{B}_4}{2}-\frac{\bar{B}_2+\bar{B}_3}{2}\right]-\left[\frac{\bar{B}_1+\bar{B}_3}{2}-\frac{\bar{B}_2+\bar{B}_4}{2}\right]&=\frac{1}{2}\left[\bar{B}_1+\bar{B}_4-\bar{B}_2-\bar{B}_3\right]-\frac{1}{2}\left[\bar{B}_1+\bar{B}_3-\bar{B}_2-\bar{B}_4\right]\\&=\frac{1}{2}\left[\bar{B}_1+\bar{B}_4-\bar{B}_2-\bar{B}_3-\bar{B}_1-\bar{B}_3+\bar{B}_2+\bar{B}_4\right]\\&=\frac{1}{2}\left[2\bar{B}_4-2\bar{B}_4\right]\\&=\bar{B}_4-\bar{B}_3\end{aligned}$$

As shown above, if we subtract the second effect from the third effect, we get the intended result; $\bar{B}_4-\bar{B}_3$. However, we will more rigorously show this result holds by considering this same proof but in Yates' notation.

This second method is shown below.

$$
\begin{aligned}
\text{effect 3 - effect 2} &= \frac{1}{2}\left[\frac{(1)+abc}{2}+\frac{c+ab}{2}-\frac{a+bc}{2}-\frac{b+ac}{2}\right]-\frac{1}{2}\left[\frac{(1)+abc}{2}+\frac{b+ac}{2}-\frac{a+bc}{2}-\frac{c+ab}{2}\right] \\
&= \frac{1}{4}\left[(1)+abc+c+ab-a-bc-b-ac\right]-\frac{1}{4}\left[(1)+abc+b+ac-a-bc-c-ab\right] \\
&= \frac{1}{4}\left[(1)+abc+c+ab-a-bc-b-ac-(1)-abc-b-ac+a+bc+c+ab\right] \\
&= \frac{1}{4}\left[2c+2ab-2b-2ac\right] \\
&= \frac{1}{2}\left[ab+c-ac-b\right]
\end{aligned}
$$

Hence, as shown above, if we subtract the second effect from the third effect, we get the intended result; $\frac{1}{2}(ab+c-ac-b)$. Thus, the treatment effect that is confounded with the $B_4 - B_3$ block effect is $AB - AC$.

Lecture 21 Problem 2a:
In this problem, we will consider a 2^4 design involving four incomplete blocks. The $+/-$ table for a 2^4 design is shown below.

TABLE 6.11
Contrast Constants for the 2^4 Design

	A	*B*	*AB*	*C*	*AC*	*BC*	*ABC*	*D*	*AD*	*BD*	*ABD*	*CD*	*ACD*	*BCD*	*ABCD*
(1)	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+
a	+	−	−	−	−	+	+	−	−	+	+	+	+	−	−
b	−	+	−	−	+	−	+	−	+	−	+	+	−	+	−
ab	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+
c	−	−	+	+	−	−	+	−	+	+	−	−	+	+	−
ac	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+
bc	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+
abc	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−
d	−	−	+	−	+	+	−	+	−	−	+	−	+	+	−
ad	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+
bd	−	+	−	−	+	−	+	+	−	+	−	−	+	−	+
abd	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−
cd	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+
acd	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−
bcd	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−
abcd	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

The goal of this problem is to write the treatment combinations that must be in each of the four blocks (in Yates' notation) if we want to assure that the ABC effect and the ACD effect are confounded with the block effect. In order to do this, we must first split by ABC using the + table above, this is done below.

$$
\begin{aligned}
\text{ABC=-} &: [(1),\ ab,\ ac,\ bc,\ d,\ abd,\ acd,\ bcd] \\
\text{ABC=+} &: [a,\ b,\ c,\ abc,\ ad,\ bd,\ cd,\ abcd]
\end{aligned}
$$

Now, if we split by ACD and consider the different $+/-$ combinations with ABC, we obtain the following four blocks shown below.

$$
\begin{aligned}
\text{Block 1} &: [(1),\ ac,\ abd,\ bcd] \\
\text{Block 2} &: [ab,\ bc,\ d,\ acd] \\
\text{Block 3} &: [b,\ abc,\ ad,\ cd] \\
\text{Block 4} &: [a,\ c,\ bd,\ abcd]
\end{aligned}
$$

These four blocks include the treatment combinations that ensure the ABC effect and the ACD effect are confounded with the block effect.

Lecture 21 Problem 2b:
In this problem, we will write the treatment combinations that must be in each of the four blocks (in Yates' notation) if we want to assure that the ABC effect and the BD effect are confounded with the block effect. In order to do this, we must first split by ABC using the $+/-$ table above, this is done below.

$$\text{ABC=-} : [(1),\ ab,\ ac,\ bc,\ d,\ abd,\ acd,\ bcd]$$
$$\text{ABC=+} : [a,\ b,\ c,\ abc,\ ad,\ bd,\ cd,\ abcd]$$

Now, if we split by BD and consider the different $+$ combinations with ABC, we obtain the following four blocks shown below.

$$\text{Block 1} : [ab,\ bc,\ d,\ acd]$$
$$\text{Block 2} : [(1),\ ac,\ abd,\ bcd]$$
$$\text{Block 3} : [b,\ abc,\ ad,\ cd]$$
$$\text{Block 4} : [a,\ c,\ bd,\ abcd]$$

These four blocks include the treatment combinations that ensure the ABC effect and the BD effect are confounded with the block effect. Notice that these four blocks are exactly the same, this is because blocking to ensure the ABC and ACD effects are confounded with the block effect, implies that the BD effect is also confounded with the block effect.

Lecture 21 Problem 3a:
In this problem. we will consider the data from example 7.1. However example 7.1 considers the experiment from section 6.2, so we will first describe that experiment and then make a few changes. In this experiment, we consider an investigation into the effect of the concentration of the reactant and the amount of the catalyst on the conversion (yield) in a chemical process. The objective of the experiment is to determine if adjustments to either of these two factors would increase the yield. Let the reactant concentration be factor A and let the two levels of interest be 15 and 25 percent. The catalyst is factor B, with the high level denoting the use of 2 pounds of the catalyst and the low level denoting the use of only 1 pound. The experiment is replicated three times, so there are 12 runs. The order in which the runs are made is random, so this is a completely randomized experiment. The data from this experiment are shown below.

Factor			Replicate			
A	**B**	**Treatment Combination**	**I**	**II**	**III**	**Total**
−	−	*A* low, *B* low	28	25	27	80
+	−	*A* high, *B* low	36	32	32	100
−	+	*A* low, *B* high	18	19	23	60
+	+	*A* high, *B* high	31	30	29	90

Now, suppose that only four experimental trials can be made from a single batch of raw material. Therefore, three batches of raw material will be required to run all three replicates of this design. In this design, each batch of raw material corresponds to a block. The data from this experiment are shown below.

■ **TABLE 7.1**
Chemical Process Experiment in Three Blocks

	Block 1	Block 2	Block 3
	(1) = 28	(1) = 25	(1) = 27
	a = 36	a = 32	a = 32
	b = 18	b = 19	b = 23
	ab = 31	ab = 30	ab = 29
Block totals:	B_1 = 113	B_2 = 106	B_3 = 111

In this sub-part, for the 2^2 design with three complete blocks, we will develop an additive model for the RCBD with replication being blocked. We will report the ANOVA table and all of the effects, where the effects are simply the coefficients returned by $lm()$ with its default contrast. These computations are done in R below.

```r
# Set up important numbers for calculations:
num_blocks <- 3
num_runs <- 4

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

# Fill in the matrix with the data from the problem:
y_mat[1,] <- c(28, 36, 18, 31)
y_mat[2,] <- c(25, 32, 19, 30)
y_mat[3,] <- c(27, 32, 23, 29)

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

# Set up the factors to match the data:
design <- gen.factorial(c(2, 2, 3),3,varNames=c("A","B","C"), factors="all")
attach(design)

# 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 28
##  [2,] 2 1 1 36
##  [3,] 1 2 1 18
##  [4,] 2 2 1 31
##  [5,] 1 1 2 25
##  [6,] 2 1 2 32
##  [7,] 1 2 2 19
##  [8,] 2 2 2 30
##  [9,] 1 1 3 27
## [10,] 2 1 3 32
## [11,] 1 2 3 23
## [12,] 2 2 3 29
```

```r
# Develop an additive model:
model_3 <- lm(y~A+B+C)

# Run ANOVA in R to get the results:
summary.aov(model_3)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## A            1 208.33  208.33  43.970 0.000296 ***
## B            1  75.00   75.00  15.829 0.005334 ** 
## C            2   6.50    3.25   0.686 0.534523    
## Residuals    7  33.17    4.74                     
## ---
## 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 an A effect (concentration effect), as well as a B effect (catalyst effect), due to the low p-values of 0.000296 and 0.005334, respectively. On the other hand, due to a p-value of 0.534523 we can cautiously conclude that there is no block effect.

```r
# Display the results of the linear model to get the effects/coefficients:
summary(model_3)
```

```
## 
## Call:
## lm(formula = y ~ A + B + C)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -3.5833 -0.9167  0.5417  1.1667  1.9167 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   26.583      1.405  18.920 2.87e-07 ***
## A2             8.333      1.257   6.631 0.000296 ***
## B2            -5.000      1.257  -3.979 0.005334 ** 
## C2            -1.750      1.539  -1.137 0.292967    
## C3            -0.500      1.539  -0.325 0.754787    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.177 on 7 degrees of freedom
## Multiple R-squared:  0.8973, Adjusted R-squared:  0.8386 
## F-statistic: 15.29 on 4 and 7 DF,  p-value: 0.001437
```

As shown above, the estimated value of the A effect is 8.333, the estimated value of the B effect is -5, and the estimated value of the block effects are -1.75 and -0.5. It is important to note that since there are three blocks in this design, which implies two degrees of freedom, there are two block effects.

Lecture 21 Problem 3b:
In this problem, we will focus on a 2^2 design with two incomplete blocks. In particular we will keep only replicates one and two, and we will suppose that block 1 contains the tuns (1) and ab, and block 2 contains the runs a and b (in Yates notation). The goal of this problem is to develop an additive model for the data as if it came from an RCBD, as well as report the ANOVA table. These computations are done in R below.

```r
# Set up important numbers for calculations:
num_blocks <- 2
num_runs <- 4

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

# Fill in the matrix with the data from the problem:
y_mat[1,] <- c(28, 36, 18, 31)
y_mat[2,] <- c(25, 32, 19, 30)

# 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(c(1, 2), times=num_runs))
B <- as.factor(rep(rep(c(1,2), each=num_blocks), times=num_blocks))
C <- as.factor(c(1, 2, 2, 1, 1, 2, 2, 1))

# 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 28
## [2,] 2 1 2 36
## [3,] 1 2 2 18
## [4,] 2 2 1 31
## [5,] 1 1 1 25
## [6,] 2 1 2 32
## [7,] 1 2 2 19
## [8,] 2 2 1 30
```

```r
# Develop an additive model:
model_4 <- lm(y~A+B+C)

# Run ANOVA in R to get the results:
summary.aov(model_4)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## A            1 190.12  190.12   56.33 0.00169 **
## B            1  66.12   66.12   19.59 0.01145 *
## C            1  10.13   10.13    3.00 0.15830
## Residuals    4  13.50    3.37
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpreting the results as if this were an RCBD, at the 5% level of significance, there appears to be an A effect (concentration effect), as well as a B effect (catalyst effect), due to the low p-values of 0.00169 and 0.01145, respectively. On the other hand, due to a p-value of 0.15830 we can cautiously conclude that there is no block effect. However, it is important to note that, due to the presence of incomplete blocks, there is actually confounding in the effects, thus treating this problem like an RCBD is not the correct thing to do.