101B_hw04

Daren Sathasivam

2024-05-09

Problem 1: 5.20

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

```r
# Data cooking_time <- c(rep(3, 18), rep(4, 18)) pressure
# <- rep(rep(c(400, 500, 650), each = 2), 6) hardwood <-
# rep(rep(c(2, 4, 8), each = 6), 2) strength <- c(196.6,
# 196.0, 197.7, 196.0, 199.8, 199.4, 198.5, 197.2, 196.0,
# 196.9, 198.4, 197.6, 197.5, 196.6, 195.6, 196.2, 197.4,
# 198.1, 198.4, 198.6, 199.6, 200.4, 200.6, 200.9, 197.5,
# 198.1, 198.7, 198.0, 199.6, 199.0, 197.6, 198.4, 197.0,
# 197.8, 198.5, 199.8) paper_data <-
# data.frame(cooking_time, pressure, hardwood, strength)
# paper_data

time <- rep(c(rep(3, 6), rep(4, 6)), 3)
pressure <- rep(c(400, 500, 650), 12)
hardwood <- c(rep(2, 12), rep(4, 12), rep(8, 12))
strength <- c(196.6, 197.7, 199.8, 196, 196, 199.4, 198.4, 199.6,
    200.6, 198.6, 200.4, 200.9, 198.5, 196, 198.4, 197.2, 196.9,
    197.6, 197.5, 198.7, 199.6, 198.1, 198, 199, 197.5, 195.6,
    197.4, 196.6, 196.2, 198.1, 197.6, 197, 198.5, 198.4, 197.8,
    199.8)
# data1 <- data.frame(strength, time, pressure, hardwood)
# data1

# Model: ANOVA Model
model1 <- aov(strength ~ factor(time) * factor(hardwood) * factor(pressure))
summary(model1)
```

```
##                                                   Df Sum Sq Mean Sq F value
## factor(time)                                       1 20.250  20.250  55.395
## factor(hardwood)                                   2  7.764   3.882  10.619
## factor(pressure)                                   2 19.374   9.687  26.499
## factor(time):factor(hardwood)                      2  2.082   1.041   2.847
## factor(time):factor(pressure)                      2  2.195   1.097   3.002
## factor(hardwood):factor(pressure)                  4  6.091   1.523   4.166
## factor(time):factor(hardwood):factor(pressure)    4  1.973   0.493   1.350
## Residuals                                         18  6.580   0.366
##                                                     Pr(>F)
## factor(time)                                      6.75e-07 ***
## factor(hardwood)                                    0.0009 ***
```

```
## factor(pressure)                                 4.33e-06 ***
## factor(time):factor(hardwood)                     0.0843 .
## factor(time):factor(pressure)                     0.0750 .
## factor(hardwood):factor(pressure)                 0.0146 *
## factor(time):factor(hardwood):factor(pressure)    0.2903
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Given the ANOVA table, we can observe that all three factors, hardwood concentration, pressure, and cooking time, all have statistically significant p-values indicating that all affect paper strength. The most significant of the three being the time factor.

b. Prepare appropriate residual plots and comment on the model's adequacy.

```
par(mfrow = c(2, 2))
plot(model1)
```
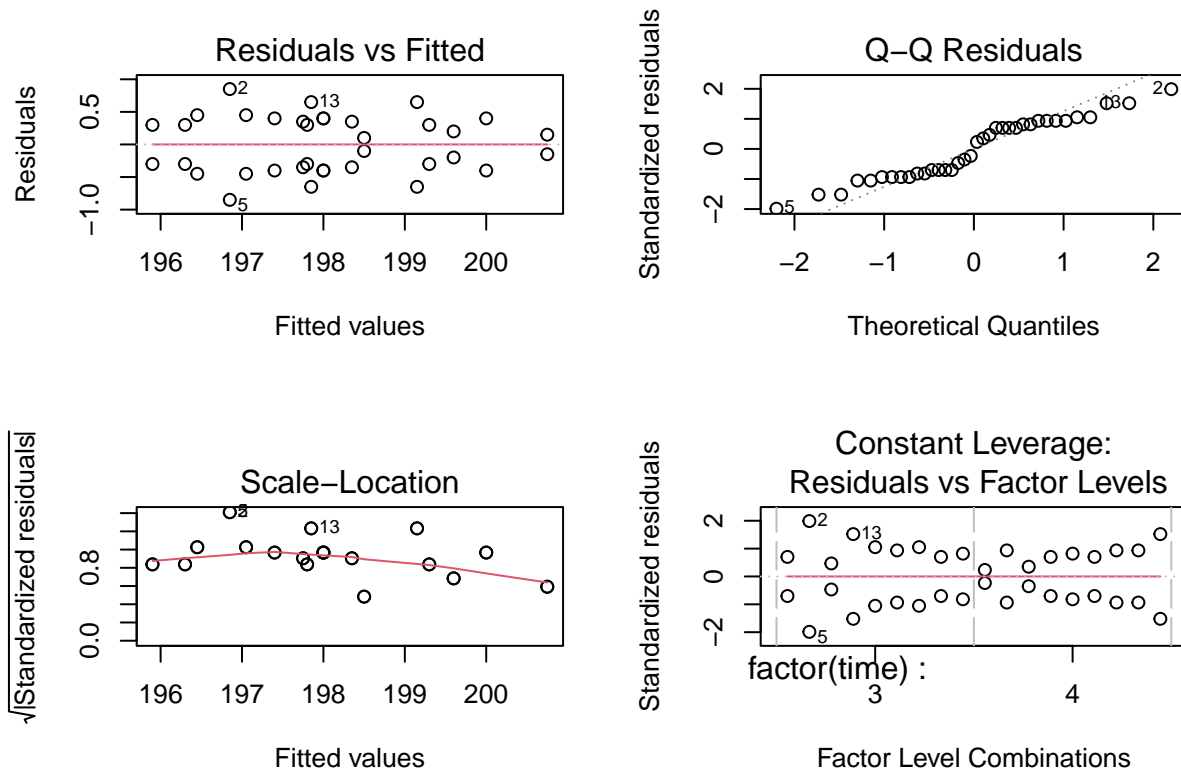


```
# plot(model1$fitted.values, model1$residuals)
# qqnorm(model1$residuals); qqline(model1$residuals)
```

- The residuals vs fitted plot displays a relatively horizontal line with no observable pattern, indicating that the model does not violate the homscedasticity assumption. The normal Q-Q plot displays a many points along the line, indicating that the model does not violate the assumption of normality. The Scale-Location plot displays a relatively horizontal line with evenly scattered points, suggesting that the model does not violate the homoscedasticity assumption. Lastly, the residuals vs leverage plot does not observe many outliers in the data and remains relatively consistent.

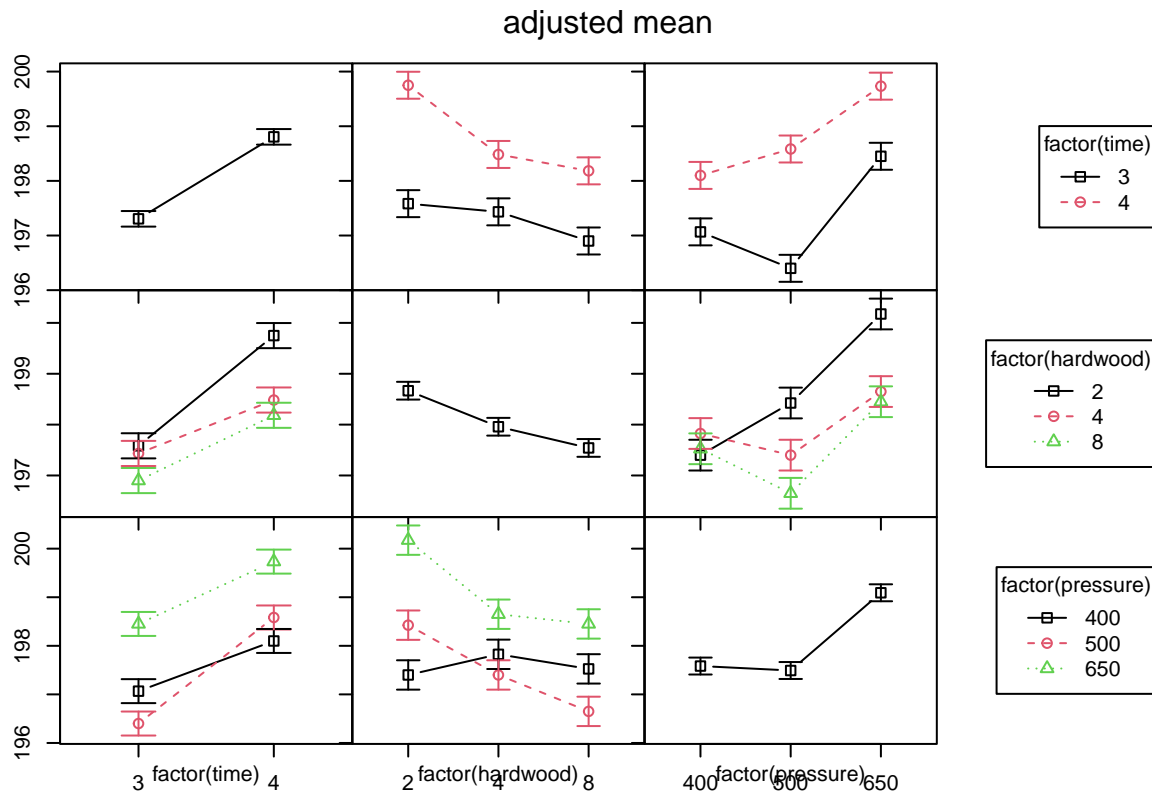## c. Under what set of conditions would you operate this process? Why?

```
library(phia)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
mip1 <- interactionMeans(model1)
plot(mip1)
```


adjusted mean

- In operating this process, the ideal conditions would involve selecting the combination of the three factors, hardwood concentration, pressure, and cooking time, that maximizes the paper strength while remaining practical in terms of time and cost. While statistically, the highest level of cooking time (4.0 hours) and the lowest level of hardwood concentration (2%) yielded the strongest paper, practical considerations suggest optimizing for cost and production speed. Thus, a recommended setting might be a cooking time of 3 hours and a hardwood concentration of 2% at a pressure of 650 psi. This approach maximizes output as more paper can be manufactured with less time being spent cooking and less of the hardwood concentration ingredient being utilized.

Problem 2: 5.37

Reconsider the three-factor factorial experiment in Problem 5.20. Suppose that this experiment had been conducted in two blocks, with each replicate a block. Assume that the observations in the data table are given in order, that is, the first observation in each cell comes from the first replicate, and so on. Reanalyze the data as a factorial experiment in blocks and estimate the variance component for blocks. Does it appear that blocking was useful in this experiment?

```
# Create blocks
blocks <- rep(c(rep(1, 3), rep(2, 3)), 6)

# Model:
model1_blocks <- aov(strength ~ factor(time) * factor(hardwood) *
    factor(pressure) * factor(blocks))

summary(model1_blocks)
```

```
##                                                      Df Sum Sq Mean Sq
## factor(time)                                          1 20.250  20.250
## factor(hardwood)                                      2  7.764   3.882
## factor(pressure)                                      2 19.374   9.687
## factor(blocks)                                        1  0.000   0.000
## factor(time):factor(hardwood)                         2  2.082   1.041
## factor(time):factor(pressure)                         2  2.195   1.097
## factor(hardwood):factor(pressure)                     4  6.091   1.523
## factor(time):factor(blocks)                           1  1.361   1.361
## factor(hardwood):factor(blocks)                       2  1.372   0.686
## factor(pressure):factor(blocks)                       2  0.182   0.091
## factor(time):factor(hardwood):factor(pressure)        4  1.973   0.493
## factor(time):factor(hardwood):factor(blocks)          2  0.514   0.257
## factor(time):factor(pressure):factor(blocks)          2  0.541   0.270
## factor(hardwood):factor(pressure):factor(blocks)      4  0.807   0.202
## factor(time):factor(hardwood):factor(pressure):factor(blocks)  4  1.804   0.451
```

- Given that the introduction of blocks did not greatly alter the ANOVA table, it appears that blocking was not useful in reducing unexplained variability in this particular experiment. This may indicate that either the blocks were not efficient in capturing the sources of variation that differ between the replicates or that the experiment is well controlled in respects to the different factors.

Problem 3: 6.11

a. Estimate the factor effects

```
# Data:
treatment_combination <- c("(1)", "a", "b", "ab", "c", "ac",
    "bc", "abc", "d", "ad", "bd", "abd", "cd", "acd", "bcd",
    "abcd")
replicate_1 <- c(90, 74, 81, 83, 77, 81, 88, 73, 98, 72, 87,
    85, 99, 79, 87, 80)
replicate_2 <- c(93, 78, 85, 80, 78, 80, 82, 70, 95, 76, 83,
    86, 90, 75, 84, 80)
treatment_combination <- rep(treatment_combination, each = 2)
```

4

```
replicate_id <- rep(c("I", "II"), each = 1)
Yield <- c(rbind(replicate_1, replicate_2))

chemical_process_data <- data.frame(Treatment_Combination = treatment_combination,
    Replicate = replicate_id, Yield = Yield)
chemical_process_data$A <- ifelse(grepl("a", chemical_process_data$Treatment_Combination),
    1, -1)
chemical_process_data$B <- ifelse(grepl("b", chemical_process_data$Treatment_Combination),
    1, -1)
chemical_process_data$C <- ifelse(grepl("c", chemical_process_data$Treatment_Combination),
    1, -1)
chemical_process_data$D <- ifelse(grepl("d", chemical_process_data$Treatment_Combination),
    1, -1)

# chemical_process_data


# Model:
model2 <- lm(Yield ~ A * B * C * D, data = chemical_process_data)
# summary(model2)

# Effects:
effects1 <- 2 * coef(model2)
effects1
```

```
## (Intercept)           A           B           C           D         A:B
##    165.5625     -9.0625     -1.3125     -2.6875      3.9375      4.0625
##         A:C         B:C         A:D         B:D         C:D       A:B:C
##      0.6875     -0.5625     -2.1875     -0.1875      1.6875     -5.1875
##       A:B:D       A:C:D       B:C:D     A:B:C:D
##      4.6875     -0.9375     -0.9375      2.4375
```

b. Prepare an analysis of variance table and determine which factors are important in explaining yield.

```
anova2 <- anova(model2)
anova2
```

```
## Analysis of Variance Table
##
## Response: Yield
##           Df Sum Sq Mean Sq F value      Pr(>F)
## A          1 657.03  657.03 85.8163 7.875e-08 ***
## B          1  13.78   13.78  1.8000 0.1984451
## C          1  57.78   57.78  7.5469 0.0143171 *
## D          1 124.03  124.03 16.2000 0.0009794 ***
## A:B        1 132.03  132.03 17.2449 0.0007491 ***
## A:C        1   3.78    3.78  0.4939 0.4923019
## B:C        1   2.53    2.53  0.3306 0.5732962
## A:D        1  38.28   38.28  5.0000 0.0399447 *
## B:D        1   0.28    0.28  0.0367 0.8504174
## C:D        1  22.78   22.78  2.9755 0.1037933
## A:B:C      1 215.28  215.28 28.1184 7.146e-05 ***
## A:B:D      1 175.78  175.78 22.9592 0.0001997 ***
```

```
## A:C:D      1   7.03    7.03  0.9184 0.3521621
## B:C:D      1   7.03    7.03  0.9184 0.3521621
## A:B:C:D    1  47.53   47.53  6.2082 0.0240766 *
## Residuals 16 122.50    7.66
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- From the ANOVA table, we can observe that factor A and D are the most significant individual contributors in explaining yield. Additionally, the combinations AB, ABC, and ABD can be observed as significant interactions which have a statistically significant influence on yield.

c. Write down a regression model for predicting yield, assuming that all four factors were varied over the range from -1 to +1(in coded units).

- $Yield = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 C + \beta_4 D + \beta_{12} AB + \beta_{13} AC + \beta_{14} AD + \beta_{23} BC + \beta_{24} BD + \beta_{34} CD + \beta_{123} ABC + \beta_{124} ABD + \beta_{134} ACD + \beta_{234} BCD + \beta_{1234} ABCD + \epsilon$

```
regmodel <- lm(Yield ~ A * B * C * D, data = chemical_process_data)
summary(regmodel)
```

```
##
## Call:
## lm.default(formula = Yield ~ A * B * C * D, data = chemical_process_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -4.5   -1.5    0.0    1.5    4.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 82.78125    0.48914 169.238  < 2e-16 ***
## A           -4.53125    0.48914  -9.264 7.87e-08 ***
## B           -0.65625    0.48914  -1.342 0.198445
## C           -1.34375    0.48914  -2.747 0.014317 *
## D            1.96875    0.48914   4.025 0.000979 ***
## A:B          2.03125    0.48914   4.153 0.000749 ***
## A:C          0.34375    0.48914   0.703 0.492302
## B:C         -0.28125    0.48914  -0.575 0.573296
## A:D         -1.09375    0.48914  -2.236 0.039945 *
## B:D         -0.09375    0.48914  -0.192 0.850417
## C:D          0.84375    0.48914   1.725 0.103793
## A:B:C       -2.59375    0.48914  -5.303 7.15e-05 ***
## A:B:D        2.34375    0.48914   4.792 0.000200 ***
## A:C:D       -0.46875    0.48914  -0.958 0.352162
## B:C:D       -0.46875    0.48914  -0.958 0.352162
## A:B:C:D      1.21875    0.48914   2.492 0.024077 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.767 on 16 degrees of freedom
## Multiple R-squared:  0.9247, Adjusted R-squared:  0.8542
## F-statistic:  13.1 on 15 and 16 DF,  p-value: 2.919e-06
```

d. Plot the residuals versus the predicted yield and on a normal probability scale. Does the residual analysis appear satisfactory?

```
par(mfrow = c(1, 2))

# Residuals vs. Predicted Yield
plot(model2$fitted.values, model2$residuals)
abline(h = 0, color = "red", lty = 2)
```

```
## Warning in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...): "color" is
## not a graphical parameter
```

```
# Normal Probability Plot of Residuals
qqnorm(model2$residuals)
qqline(model2$residuals)
```
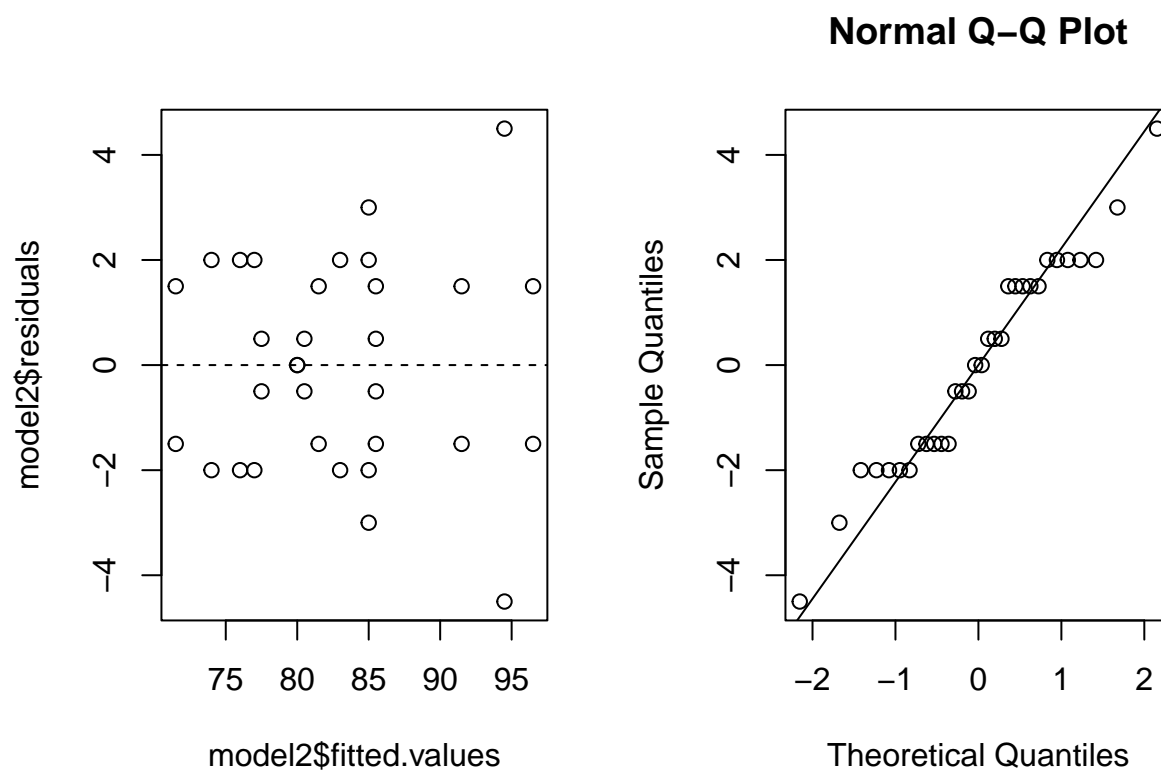

- The residuals vs predicted plot displays a random scatter around the horizontal line. This indicates that the model has constant variance(homoscedasticity). The normal Q-Q plot shows displays the residuals following the line but has slightly heavier tails, indicating that the residuals are more than likely to be normally distributed but have some weight on either end of the distribution.

e. Two three-factor interactions, ABC and ABD, apparently have large effects. Draw a cube plot in the factors A, B, and C with the average yields shown at each corner. Repeat using the factors A, B,and D. Do these two plots aid in data interpretation? Where would you recommend that the process be run with respect to the four variables?

```
# from rdrr.io install.packages('FrF2')
library(FrF2)
```

7

```
A <- chemical_process_data$A
B <- chemical_process_data$B
C <- chemical_process_data$C
D <- chemical_process_data$D

cubePlot(Yield, A, B, C)
```

Cube plot for Yield
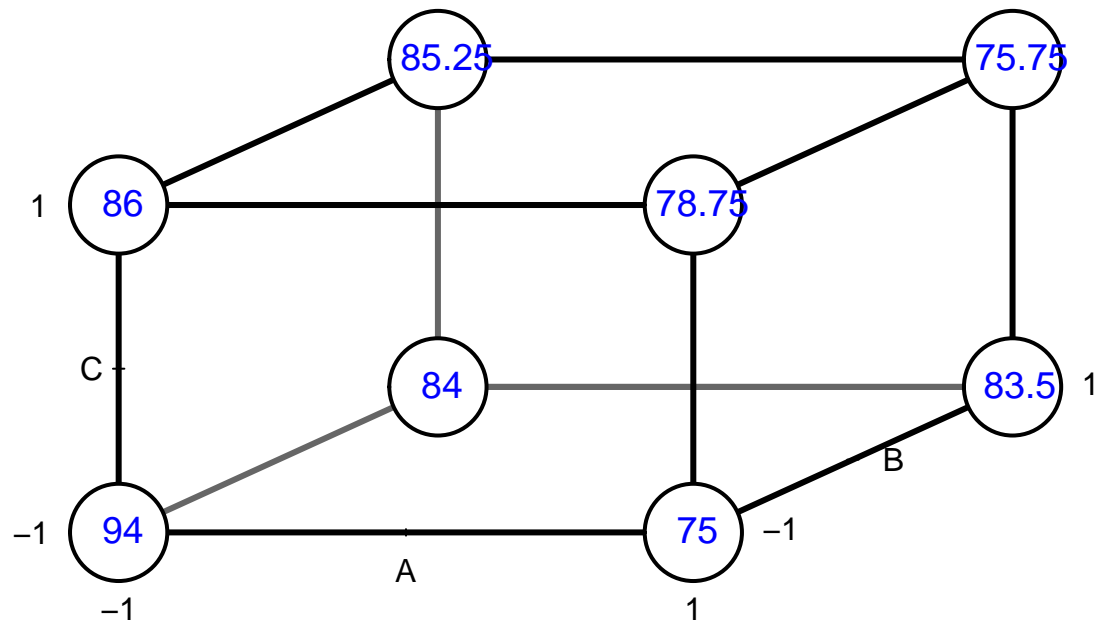

```
cubePlot(Yield, A, B, D)
```

Cube plot for Yield
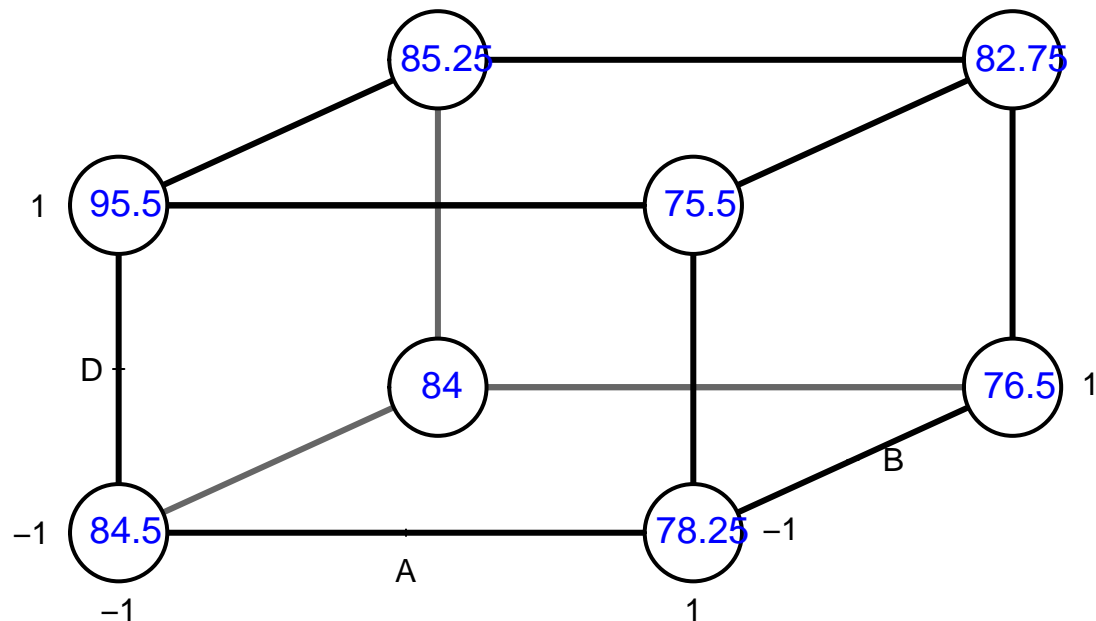
- These two cube plots aid in data interpretation as it displays the interaction between the different

factors. It is helpful as identifying which combinations of factors leading to better or worse outcomes is easily observed. Through these interaction cube plots, it is observed that the combination of ABD is the optimal factor combination to run the experiment as the value is 82.75 and the value for ABC is 75.75.

Problem 4: 6.22

a. Analyze the data from this experiment. Which factors significantly affect the customer response rate?

```r
# Data
treatment_combination <- rep(c("(1)", "a", "b", "ab", "c", "ac",
    "bc", "abc"), each = 2)
replicate_id <- rep(c("I", "II"), each = 1)
orders <- c(50, 54, 44, 42, 46, 48, 42, 43, 49, 46, 48, 45, 47,
    48, 56, 54)
mail_data <- data.frame(Treatment_Combination = treatment_combination,
    Replicate = replicate_id, Orders = orders)

mail_data$A <- ifelse(grepl("a", mail_data$Treatment_Combination),
    1, -1)
mail_data$B <- ifelse(grepl("b", mail_data$Treatment_Combination),
    1, -1)
mail_data$C <- ifelse(grepl("c", mail_data$Treatment_Combination),
    1, -1)
mail_data
```

```
##    Treatment_Combination Replicate Orders  A  B  C
## 1                   (1)         I     50 -1 -1 -1
## 2                   (1)        II     54 -1 -1 -1
## 3                     a         I     44  1 -1 -1
## 4                     a        II     42  1 -1 -1
## 5                     b         I     46 -1  1 -1
## 6                     b        II     48 -1  1 -1
## 7                    ab         I     42  1  1 -1
## 8                    ab        II     43  1  1 -1
## 9                     c         I     49 -1 -1  1
## 10                    c        II     46 -1 -1  1
## 11                   ac         I     48  1 -1  1
## 12                   ac        II     45  1 -1  1
## 13                   bc         I     47 -1  1  1
## 14                   bc        II     48 -1  1  1
## 15                  abc         I     56  1  1  1
## 16                  abc        II     54  1  1  1
```

```r
# Model
model3 <- aov(Orders ~ factor(A) * factor(B) * factor(C), data = mail_data)
summary(model3)
```

```
##                      Df Sum Sq Mean Sq F value   Pr(>F)
## factor(A)             1  12.25   12.25   4.083 0.077971 .
## factor(B)             1   2.25    2.25   0.750 0.411694
## factor(C)             1  36.00   36.00  12.000 0.008516 **
## factor(A):factor(B)   1  42.25   42.25  14.083 0.005602 **
## factor(A):factor(C)   1 100.00  100.00  33.333 0.000418 ***
```

```
## factor(B):factor(C)           1  49.00   49.00  16.333 0.003728 **
## factor(A):factor(B):factor(C) 1   4.00    4.00   1.333 0.281537
## Residuals                      8  24.00    3.00
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Effects
effects2 <- coef(model3)
effects2
```

```
##                     (Intercept)                       factor(A)1
##                            52.0                             -9.0
##                       factor(B)1                       factor(C)1
##                            -5.0                             -4.5
##           factor(A)1:factor(B)1            factor(A)1:factor(C)1
##                             4.5                              8.0
##           factor(B)1:factor(C)1 factor(A)1:factor(B)1:factor(C)1
##                             5.0                              4.0
```

- From the model summary, it can be observed that the most significant factors are C(Offered Price) and the combination of factors A(Type of Mail Used) and C(Offered Price) being a significant factor combination in customer response rate. However, the p-value of A(Mail Type) is greater than 0.05, which can indicate that it is not significant enough to reject the null hypothesis.

b. Analyze the residuals from this experiment. Are there any indications of model inadequacy?

```r
par(mfrow = c(2, 2))
plot(model3)
```
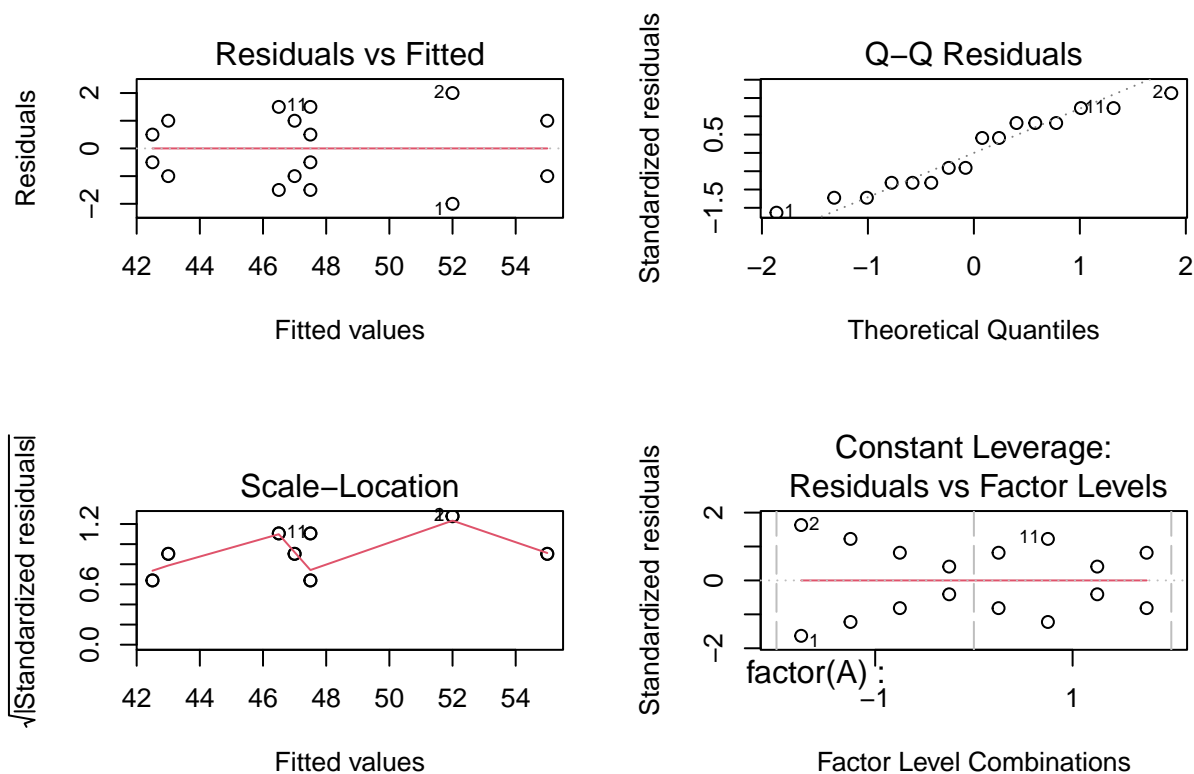

Based on the residuals of the model, we can observe that the Normal Q-Q plot holds the normality assumption

as the residual points follow the line. Additionally, the Residuals vs. Fitted Plot displays a relatively random scatter of points, indicating that the residuals maintain homoscedasticity. Therefore the residuals do not indicate model inadequacy.

c. What would you recommend to the company? (Answer based on the interaction plot)

```
mip2 <- interactionMeans(model3)
plot(mip2)
```


adjusted mean
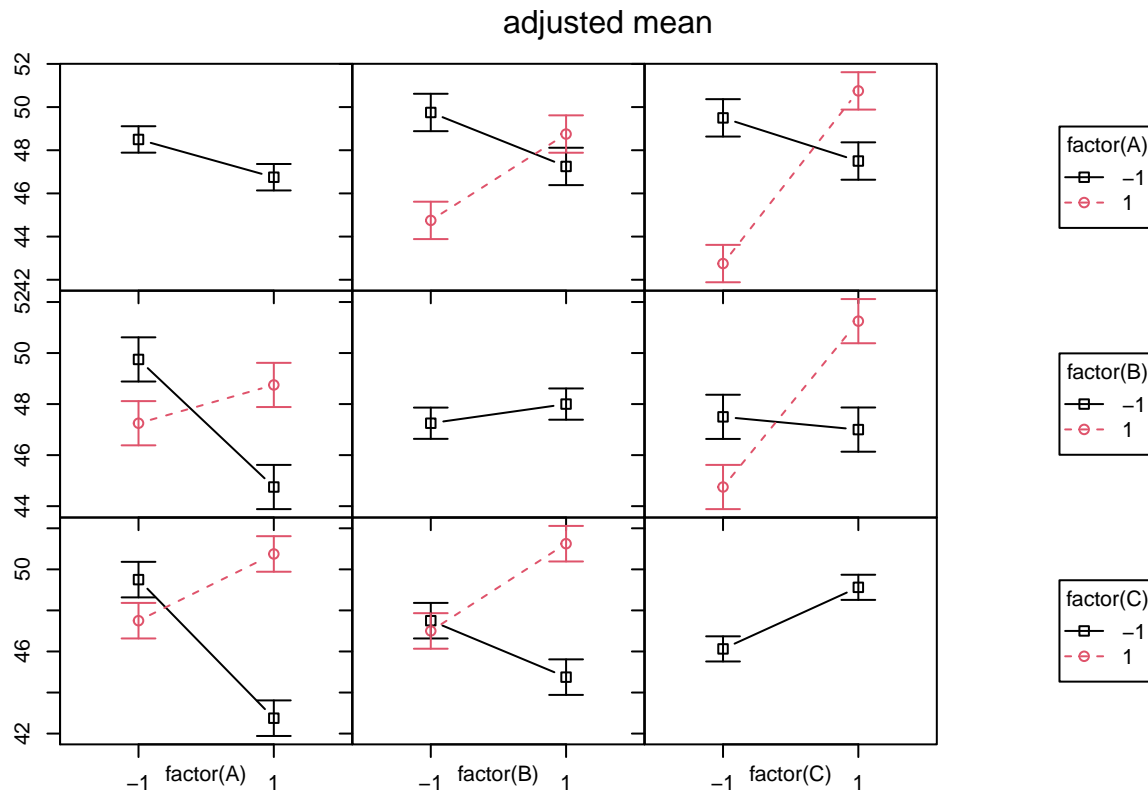
- I would recommend the company to prioritize their pricing strategy as the factor C(Offered Price) had the most significant impact on the response rate. Additionally, they may also work on their mailing methods as the combination of A(Mail Type) and C(Offered Price) also has shown significant effect on the response rate. Overall, I would highly recommend focusing on the pricing strategy of their products and maybe utilize online advertisements rather than using physical mail as it may be more costly.

Problem 5: 6.35

a. Analyze the data from this experiment. Identify the significant factors and interactions.

```
# Data A <- rep(c(-1, 1), each = 1, times = 16) B <-
# rep(c(-1, 1), each = 2, times = 8) C <- rep(c(-1, 1),
# each = 4, times = 4) D <- rep(c(-1, 1), each = 8, times =
# 2) E <- rep(c(-1, 1), each = 16, times = 1)

y <- c(8.11, 5.56, 5.77, 5.82, 9.17, 7.8, 3.23, 5.69, 8.82, 14.23,
```

11

```r
    9.2, 8.94, 8.68, 11.49, 6.25, 9.12, 7.93, 5, 7.47, 12, 9.86,
    3.65, 6.4, 11.61, 12.43, 17.55, 8.87, 25.38, 13.06, 18.85,
    11.78, 26.05)
treatment_combination <- c("(1)", "a", "b", "ab", "c", "ac",
    "bc", "abc", "d", "ad", "bd", "abd", "cd", "acd", "bcd",
    "abcd", "e", "ae", "be", "abe", "ce", "ace", "bce", "abce",
    "de", "ade", "bde", "abde", "cde", "acde", "bcde", "abcde")

experiment_data <- data.frame(Treatment_Combination = treatment_combination,
    y = y)

experiment_data$A <- ifelse(grepl("a", experiment_data$Treatment_Combination),
    1, -1)
experiment_data$B <- ifelse(grepl("b", experiment_data$Treatment_Combination),
    1, -1)
experiment_data$C <- ifelse(grepl("c", experiment_data$Treatment_Combination),
    1, -1)
experiment_data$D <- ifelse(grepl("d", experiment_data$Treatment_Combination),
    1, -1)
experiment_data$E <- ifelse(grepl("e", experiment_data$Treatment_Combination),
    1, -1)

experiment_data
```

```
##    Treatment_Combination     y  A  B  C  D  E
## 1                    (1)  8.11 -1 -1 -1 -1 -1
## 2                      a  5.56  1 -1 -1 -1 -1
## 3                      b  5.77 -1  1 -1 -1 -1
## 4                     ab  5.82  1  1 -1 -1 -1
## 5                      c  9.17 -1 -1  1 -1 -1
## 6                     ac  7.80  1 -1  1 -1 -1
## 7                     bc  3.23 -1  1  1 -1 -1
## 8                    abc  5.69  1  1  1 -1 -1
## 9                      d  8.82 -1 -1 -1  1 -1
## 10                    ad 14.23  1 -1 -1  1 -1
## 11                    bd  9.20 -1  1 -1  1 -1
## 12                   abd  8.94  1  1 -1  1 -1
## 13                    cd  8.68 -1 -1  1  1 -1
## 14                   acd 11.49  1 -1  1  1 -1
## 15                   bcd  6.25 -1  1  1  1 -1
## 16                  abcd  9.12  1  1  1  1 -1
## 17                     e  7.93 -1 -1 -1 -1  1
## 18                    ae  5.00  1 -1 -1 -1  1
## 19                    be  7.47 -1  1 -1 -1  1
## 20                   abe 12.00  1  1 -1 -1  1
## 21                    ce  9.86 -1 -1  1 -1  1
## 22                   ace  3.65  1 -1  1 -1  1
## 23                   bce  6.40 -1  1  1 -1  1
## 24                  abce 11.61  1  1  1 -1  1
## 25                    de 12.43 -1 -1 -1  1  1
## 26                   ade 17.55  1 -1 -1  1  1
## 27                   bde  8.87 -1  1 -1  1  1
## 28                  abde 25.38  1  1 -1  1  1
## 29                   cde 13.06 -1 -1  1  1  1
```

```
## 30                acde 18.85  1 -1  1  1  1
## 31                bcde 11.78 -1  1  1  1  1
## 32               abcde 26.05  1  1  1  1  1
```

```r
# Model(not A*B*C*D*E):
model4 <- aov(y ~ factor(A) * factor(B) * factor(C) * factor(D) *
    factor(E), data = experiment_data)
summary(model4)
```

```
##                                              Df Sum Sq Mean Sq
## factor(A)                                     1  83.56   83.56
## factor(B)                                     1   0.06    0.06
## factor(C)                                     1   0.00    0.00
## factor(D)                                     1 285.78  285.78
## factor(E)                                     1 153.17  153.17
## factor(A):factor(B)                           1  48.93   48.93
## factor(A):factor(C)                           1   0.00    0.00
## factor(B):factor(C)                           1   1.22    1.22
## factor(A):factor(D)                           1  88.88   88.88
## factor(B):factor(D)                           1   0.01    0.01
## factor(C):factor(D)                           1   0.00    0.00
## factor(A):factor(E)                           1  33.76   33.76
## factor(B):factor(E)                           1  52.71   52.71
## factor(C):factor(E)                           1   2.91    2.91
## factor(D):factor(E)                           1  61.80   61.80
## factor(A):factor(B):factor(C)                 1   2.01    2.01
## factor(A):factor(B):factor(D)                 1   3.82    3.82
## factor(A):factor(C):factor(D)                 1   0.13    0.13
## factor(B):factor(C):factor(D)                 1   2.98    2.98
## factor(A):factor(B):factor(E)                 1  44.96   44.96
## factor(A):factor(C):factor(E)                 1   2.15    2.15
## factor(B):factor(C):factor(E)                 1   0.94    0.94
## factor(A):factor(D):factor(E)                 1  26.01   26.01
## factor(B):factor(D):factor(E)                 1   0.05    0.05
## factor(C):factor(D):factor(E)                 1   5.02    5.02
## factor(A):factor(B):factor(C):factor(D)       1   0.18    0.18
## factor(A):factor(B):factor(C):factor(E)       1   1.09    1.09
## factor(A):factor(B):factor(D):factor(E)       1   5.31    5.31
## factor(A):factor(C):factor(D):factor(E)       1   0.52    0.52
## factor(B):factor(C):factor(D):factor(E)       1   0.18    0.18
## factor(A):factor(B):factor(C):factor(D):factor(E) 1   4.04    4.04
```

```r
a <- data.frame(summary(model4)[[1]])[, 2, drop = FALSE]
a$PercntSS <- round(a$Sum.Sq/sum(a$Sum.Sq), 3)
a
```

```
##                              Sum.Sq PercntSS
## factor(A)              8.356013e+01    0.092
## factor(B)              6.037813e-02    0.000
## factor(C)              4.753125e-03    0.000
## factor(D)              2.857843e+02    0.313
## factor(E)              1.531688e+02    0.168
## factor(A):factor(B)    4.893078e+01    0.054
## factor(A):factor(C)    7.812500e-05    0.000
## factor(B):factor(C)    1.220703e+00    0.001
## factor(A):factor(D)    8.887778e+01    0.097
```

```
## factor(B):factor(D)                                    5.778125e-03    0.000
## factor(C):factor(D)                                    3.781250e-04    0.000
## factor(A):factor(E)                                    3.376365e+01    0.037
## factor(B):factor(E)                                    5.271078e+01    0.058
## factor(C):factor(E)                                    2.910078e+00    0.003
## factor(D):factor(E)                                    6.179940e+01    0.068
## factor(A):factor(B):factor(C)                          2.005003e+00    0.002
## factor(A):factor(B):factor(D)                          3.815703e+00    0.004
## factor(A):factor(C):factor(D)                          1.287781e-01    0.000
## factor(B):factor(C):factor(D)                          2.982903e+00    0.003
## factor(A):factor(B):factor(E)                          4.495890e+01    0.049
## factor(A):factor(C):factor(E)                          2.147628e+00    0.002
## factor(B):factor(C):factor(E)                          9.350281e-01    0.001
## factor(A):factor(D):factor(E)                          2.601008e+01    0.029
## factor(B):factor(D):factor(E)                          5.040312e-02    0.000
## factor(C):factor(D):factor(E)                          5.016528e+00    0.005
## factor(A):factor(B):factor(C):factor(D)                1.755281e-01    0.000
## factor(A):factor(B):factor(C):factor(E)                1.091503e+00    0.001
## factor(A):factor(B):factor(D):factor(E)                5.305653e+00    0.006
## factor(A):factor(C):factor(D):factor(E)                5.227531e-01    0.001
## factor(B):factor(C):factor(D):factor(E)                1.785031e-01    0.000
## factor(A):factor(B):factor(C):factor(D):factor(E) 4.039903e+00    0.004
```

```r
# Effects:
effects3 <- coef(model4)
effects3
```

```
##                                         (Intercept)
##                                                8.11
##                                           factor(A)1
##                                               -2.55
##                                           factor(B)1
##                                               -2.34
##                                           factor(C)1
##                                                1.06
##                                           factor(D)1
##                                                0.71
##                                           factor(E)1
##                                               -0.18
##                                 factor(A)1:factor(B)1
##                                                2.60
##                                 factor(A)1:factor(C)1
##                                                1.18
##                                 factor(B)1:factor(C)1
##                                               -3.60
##                                 factor(A)1:factor(D)1
##                                                7.96
##                                 factor(B)1:factor(D)1
##                                                2.72
##                                 factor(C)1:factor(D)1
##                                               -1.20
##                                 factor(A)1:factor(E)1
##                                               -0.38
##                                 factor(B)1:factor(E)1
##                                                1.88
```

```
##                                  factor(C)1:factor(E)1
##                                                   0.87
##                                  factor(D)1:factor(E)1
##                                                   3.79
##                      factor(A)1:factor(B)1:factor(C)1
##                                                   1.23
##                      factor(A)1:factor(B)1:factor(D)1
##                                                  -8.27
##                      factor(A)1:factor(C)1:factor(D)1
##                                                  -3.78
##                      factor(B)1:factor(C)1:factor(D)1
##                                                   0.79
##                      factor(A)1:factor(B)1:factor(E)1
##                                                   4.86
##                      factor(A)1:factor(C)1:factor(E)1
##                                                  -4.46
##                      factor(B)1:factor(C)1:factor(E)1
##                                                   0.60
##                      factor(A)1:factor(D)1:factor(E)1
##                                                   0.09
##                      factor(B)1:factor(D)1:factor(E)1
##                                                  -5.82
##                      factor(C)1:factor(D)1:factor(E)1
##                                                  -0.10
##          factor(A)1:factor(B)1:factor(C)1:factor(D)1
##                                                   4.50
##          factor(A)1:factor(B)1:factor(C)1:factor(E)1
##                                                   2.73
##          factor(A)1:factor(B)1:factor(D)1:factor(E)1
##                                                  12.20
##          factor(A)1:factor(C)1:factor(D)1:factor(E)1
##                                                   7.73
##          factor(B)1:factor(C)1:factor(D)1:factor(E)1
##                                                   4.49
## factor(A)1:factor(B)1:factor(C)1:factor(D)1:factor(E)1
##                                                 -11.37
```

- We can observe that the factors A, D, and E have a statistically significant affect on the response of this experiment. Their interactions between one another such as AD and DE show a statisticalyl significant affect on the response of this experiment also.

b. Analyze the residuals from this experiment. Are there any indications of model inadequacy or violations of the assumptions? (SKIP)

c. One of the factors from this experiment does not seem to be important. If you drop this factor, what type of design remains? Analyze the data using the full factorial model for only the four active factors. Compare your results with those obtained in part (a).

```r
model4_reduced <- aov(y ~ factor(A) * factor(B) * factor(D) *
    factor(E), data = experiment_data)
summary(model4_reduced)
```

```
##                                 Df Sum Sq Mean Sq F value   Pr(>F)
```

```
## factor(A)                                   1  83.56   83.56  57.233 1.14e-06 ***
## factor(B)                                   1   0.06    0.06   0.041 0.841418
## factor(D)                                   1 285.78  285.78 195.742 2.16e-10 ***
## factor(E)                                   1 153.17  153.17 104.910 1.97e-08 ***
## factor(A):factor(B)                         1  48.93   48.93  33.514 2.77e-05 ***
## factor(A):factor(D)                         1  88.88   88.88  60.875 7.66e-07 ***
## factor(B):factor(D)                         1   0.01    0.01   0.004 0.950618
## factor(A):factor(E)                         1  33.76   33.76  23.126 0.000193 ***
## factor(B):factor(E)                         1  52.71   52.71  36.103 1.82e-05 ***
## factor(D):factor(E)                         1  61.80   61.80  42.328 7.24e-06 ***
## factor(A):factor(B):factor(D)               1   3.82    3.82   2.613 0.125501
## factor(A):factor(B):factor(E)               1  44.96   44.96  30.794 4.40e-05 ***
## factor(A):factor(D):factor(E)               1  26.01   26.01  17.815 0.000650 ***
## factor(B):factor(D):factor(E)               1   0.05    0.05   0.035 0.854935
## factor(A):factor(B):factor(D):factor(E)  1   5.31    5.31   3.634 0.074735 .
## Residuals                                  16  23.36    1.46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- By reducing the model by removing factor C, it can be observed that the most significant factors are A, D, and E. The type of design goes from 2^5 to 2^4 factorial design. Additionally, it remains that the interaction factors of AD and DE remain the most statistically significant amongst the interaction factors.

d. Find settings of the active factors that maximize the predicted response. (Answer based on the cube plot)

```
effects3_reduced <- coef(model4_reduced)
effects3_reduced
```

```
##                             (Intercept)
##                                   8.640
##                               factor(A)1
##                                  -1.960
##                               factor(B)1
##                                  -4.140
##                               factor(D)1
##                                   0.110
##                               factor(E)1
##                                   0.255
##                    factor(A)1:factor(B)1
##                                   3.215
##                    factor(A)1:factor(D)1
##                                   6.070
##                    factor(B)1:factor(D)1
##                                   3.115
##                    factor(A)1:factor(E)1
##                                  -2.610
##                    factor(B)1:factor(E)1
##                                   2.180
##                    factor(D)1:factor(E)1
##                                   3.740
##          factor(A)1:factor(B)1:factor(D)1
##                                  -6.020
```

```
##           factor(A)1:factor(B)1:factor(E)1
##                                     6.225
##           factor(A)1:factor(D)1:factor(E)1
##                                     3.955
##           factor(B)1:factor(D)1:factor(E)1
##                                    -3.575
## factor(A)1:factor(B)1:factor(D)1:factor(E)1
##                                     6.515
```
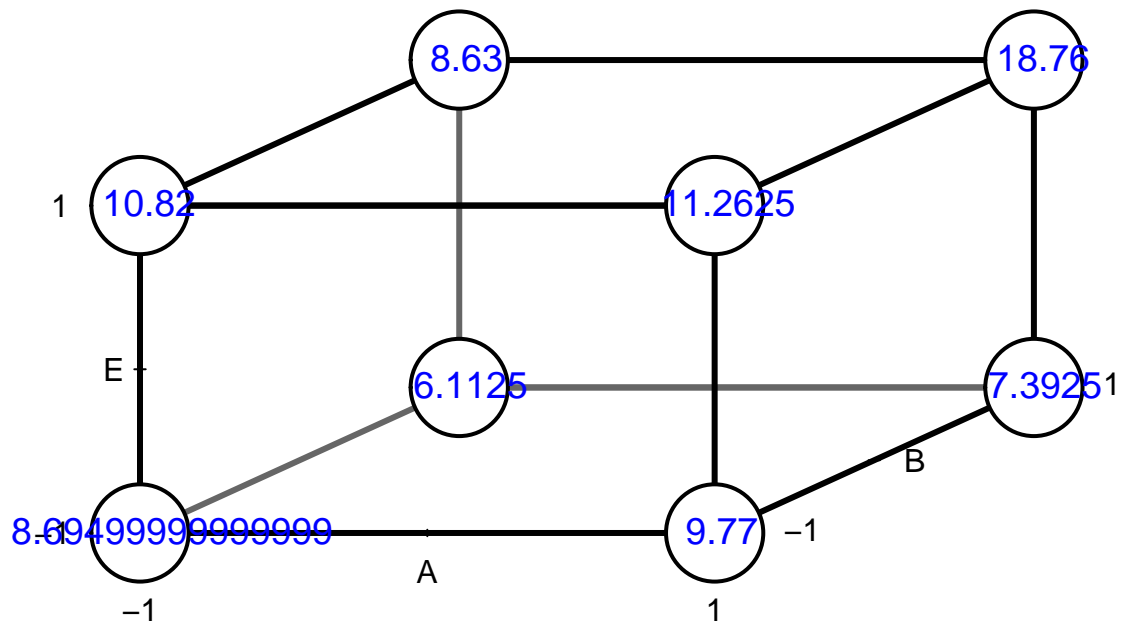
```
A <- experiment_data$A
B <- experiment_data$B
D <- experiment_data$D
E <- experiment_data$E

cubePlot(y, A, B, E)
```
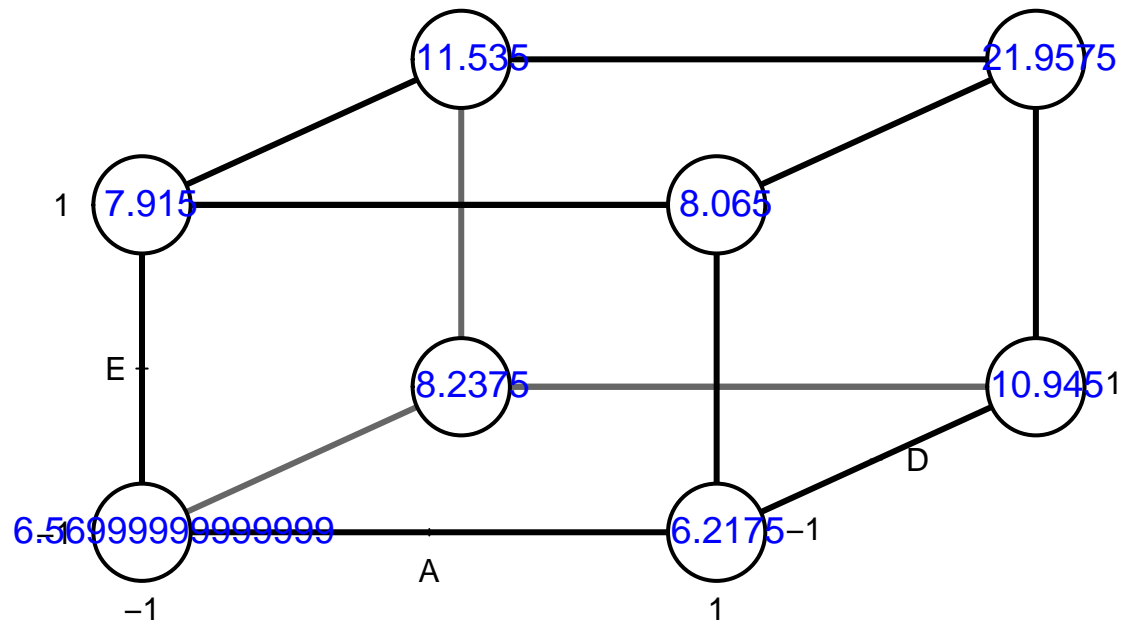

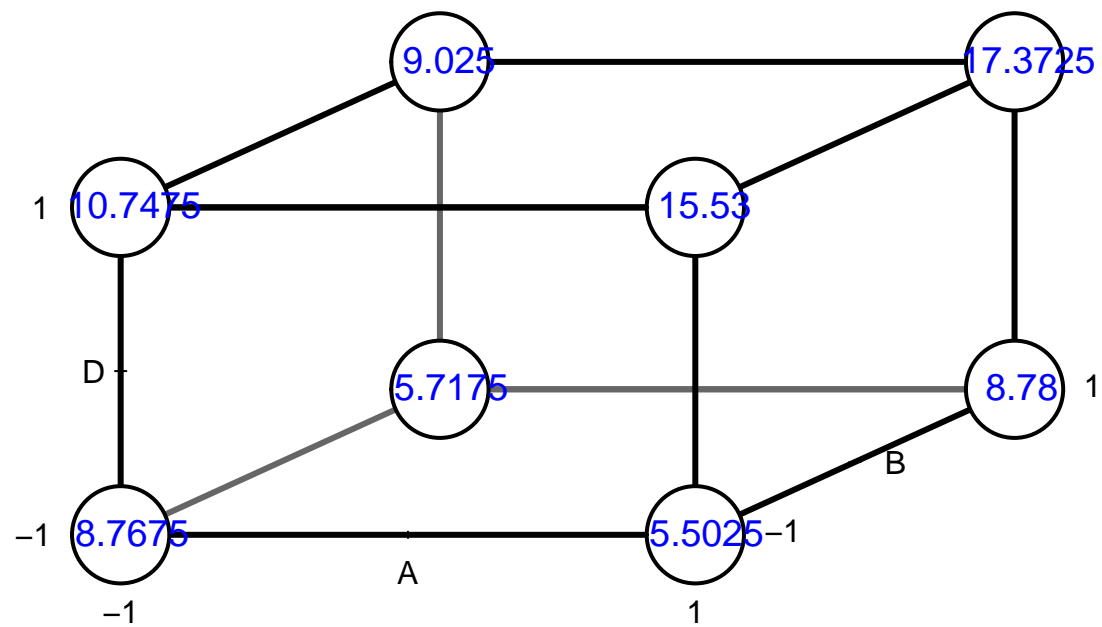## Cube plot for y

```
cubePlot(y, A, D, E)
```
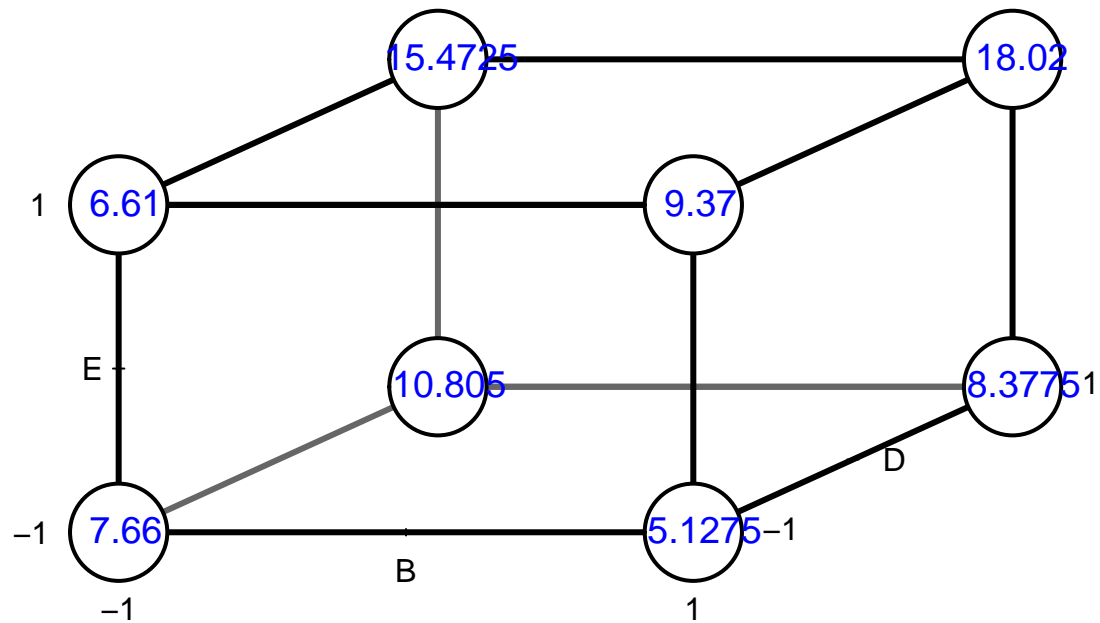
Cube plot for y


```
cubePlot(y, A, B, D)
```

Cube plot for y


```
cubePlot(y, B, D, E)
```

Cube plot for y

- It can be observed from the four different cube plots that the combination of interaction factors of **ADE** results in the largest predicted response out of the different interaction factor combinations.