101b_hw05_Daren_Satahsivam

Daren Sathasivam

2024-05-20

Problem 1 - 7.4: Consider the data from the first replicate of Problem 6.11. Construct a design with two blocks of eight observations each with ABCD confounded. Analyze the data.

```r
# 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)

chemical_process_data <- data.frame(Treatment_Combination = rep(treatment_combination,
    each = 2), Replicate = rep(c("I", "II"), times = 16), Yield = c(rbind(replicate_1,
    replicate_2)))

# ABCD Factorials
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)

# Calculate ABCD interaction
chemical_process_data$ABCD <- with(chemical_process_data, A *
    B * C * D)

# Assign blocks
chemical_process_data$Block <- ifelse(chemical_process_data$ABCD ==
    1, "Block 1", "Block 2")

chemical_process_data
```

```
##     Treatment_Combination Replicate Yield  A  B  C  D ABCD   Block
## 1                     (1)         I    90 -1 -1 -1 -1    1 Block 1
## 2                     (1)        II    93 -1 -1 -1 -1    1 Block 1
## 3                       a         I    74  1 -1 -1 -1   -1 Block 2
```

```
## 4                      a         II   78  1 -1 -1 -1   -1 Block 2
## 5                      b          I   81 -1  1 -1 -1   -1 Block 2
## 6                      b         II   85 -1  1 -1 -1   -1 Block 2
## 7                     ab          I   83  1  1 -1 -1    1 Block 1
## 8                     ab         II   80  1  1 -1 -1    1 Block 1
## 9                      c          I   77 -1 -1  1 -1   -1 Block 2
## 10                     c         II   78 -1 -1  1 -1   -1 Block 2
## 11                    ac          I   81  1 -1  1 -1    1 Block 1
## 12                    ac         II   80  1 -1  1 -1    1 Block 1
## 13                    bc          I   88 -1  1  1 -1    1 Block 1
## 14                    bc         II   82 -1  1  1 -1    1 Block 1
## 15                   abc          I   73  1  1  1 -1   -1 Block 2
## 16                   abc         II   70  1  1  1 -1   -1 Block 2
## 17                     d          I   98 -1 -1 -1  1   -1 Block 2
## 18                     d         II   95 -1 -1 -1  1   -1 Block 2
## 19                    ad          I   72  1 -1 -1  1    1 Block 1
## 20                    ad         II   76  1 -1 -1  1    1 Block 1
## 21                    bd          I   87 -1  1 -1  1    1 Block 1
## 22                    bd         II   83 -1  1 -1  1    1 Block 1
## 23                   abd          I   85  1  1 -1  1   -1 Block 2
## 24                   abd         II   86  1  1 -1  1   -1 Block 2
## 25                    cd          I   99 -1 -1  1  1    1 Block 1
## 26                    cd         II   90 -1 -1  1  1    1 Block 1
## 27                   acd          I   79  1 -1  1  1   -1 Block 2
## 28                   acd         II   75  1 -1  1  1   -1 Block 2
## 29                   bcd          I   87 -1  1  1  1   -1 Block 2
## 30                   bcd         II   84 -1  1  1  1   -1 Block 2
## 31                  abcd          I   80  1  1  1  1    1 Block 1
## 32                  abcd         II   80  1  1  1  1    1 Block 1
```
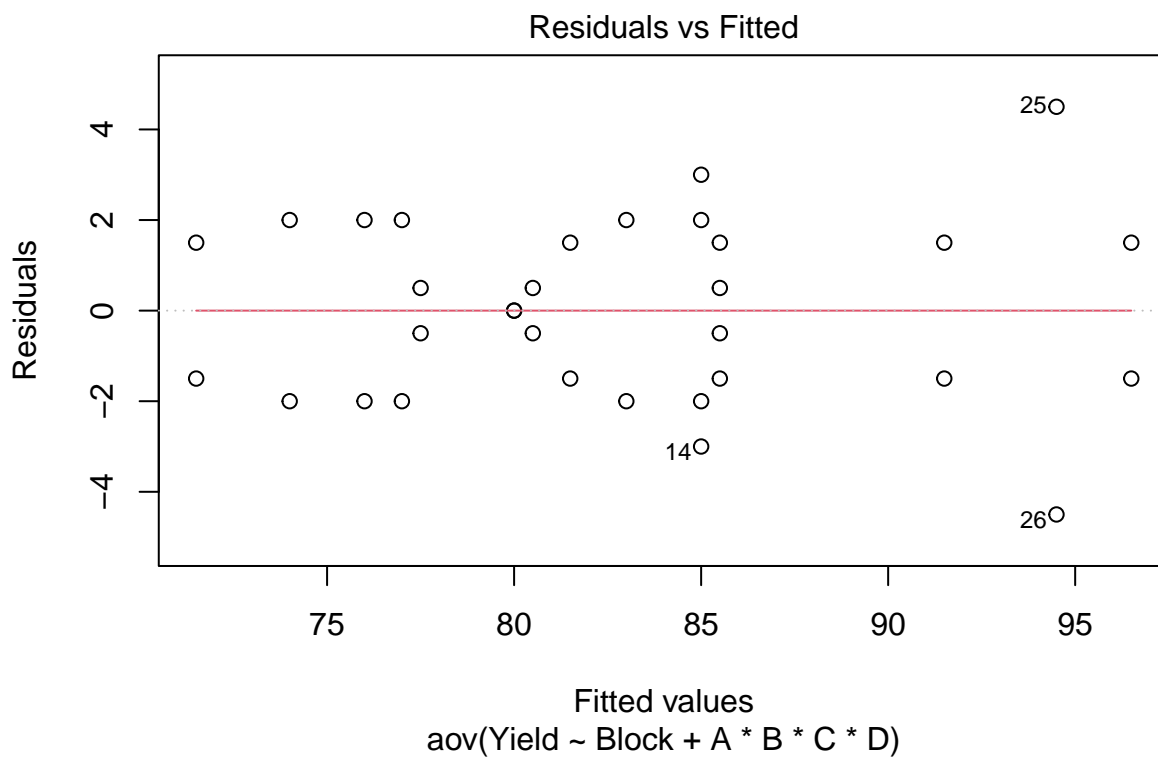
```r
# Model
model1 <- aov(Yield ~ Block + A * B * C * D, data = chemical_process_data)
summary(model1)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Block        1   47.5    47.5   6.208 0.024077 *
## A            1  657.0   657.0  85.816 7.87e-08 ***
## B            1   13.8    13.8   1.800 0.198445
## C            1   57.8    57.8   7.547 0.014317 *
## D            1  124.0   124.0  16.200 0.000979 ***
## A:B          1  132.0   132.0  17.245 0.000749 ***
## A:C          1    3.8     3.8   0.494 0.492302
## B:C          1    2.5     2.5   0.331 0.573296
## A:D          1   38.3    38.3   5.000 0.039945 *
## B:D          1    0.3     0.3   0.037 0.850417
## C:D          1   22.8    22.8   2.976 0.103793
## A:B:C        1  215.3   215.3  28.118 7.15e-05 ***
## A:B:D        1  175.8   175.8  22.959 0.000200 ***
## A:C:D        1    7.0     7.0   0.918 0.352162
## B:C:D        1    7.0     7.0   0.918 0.352162
## Residuals   16  122.5     7.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
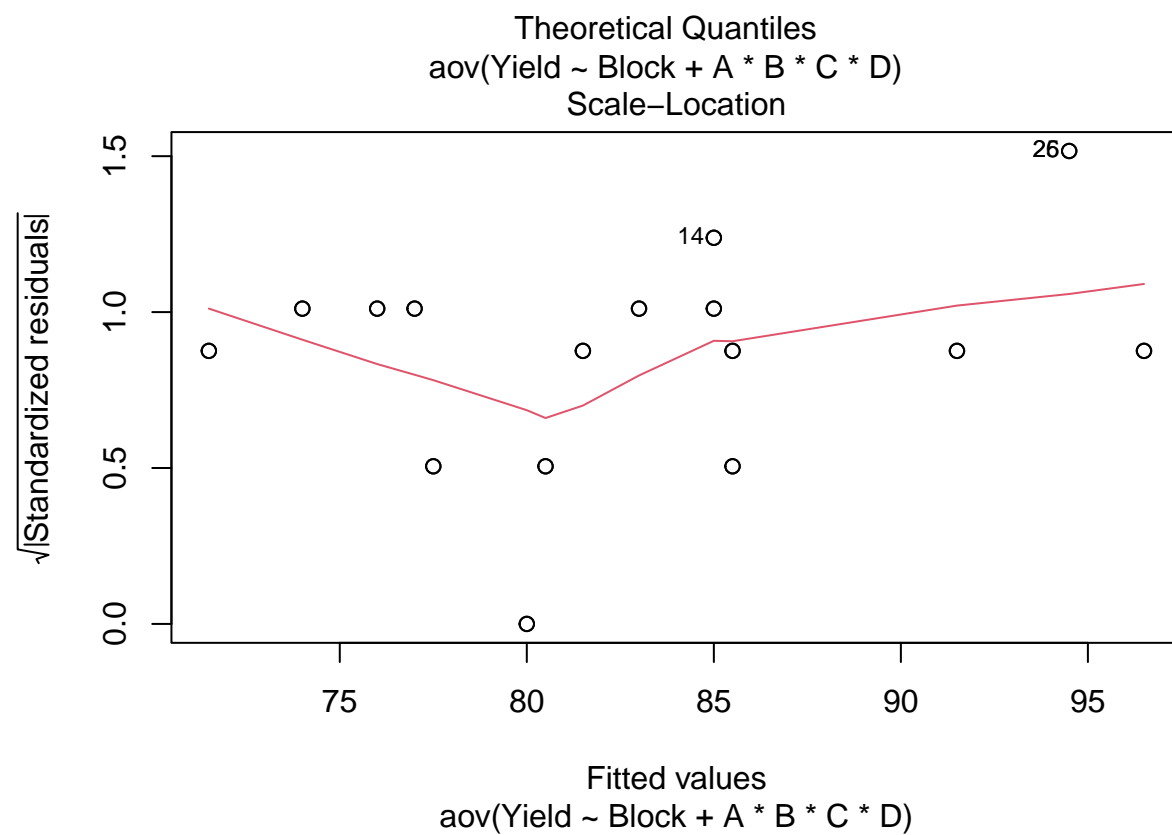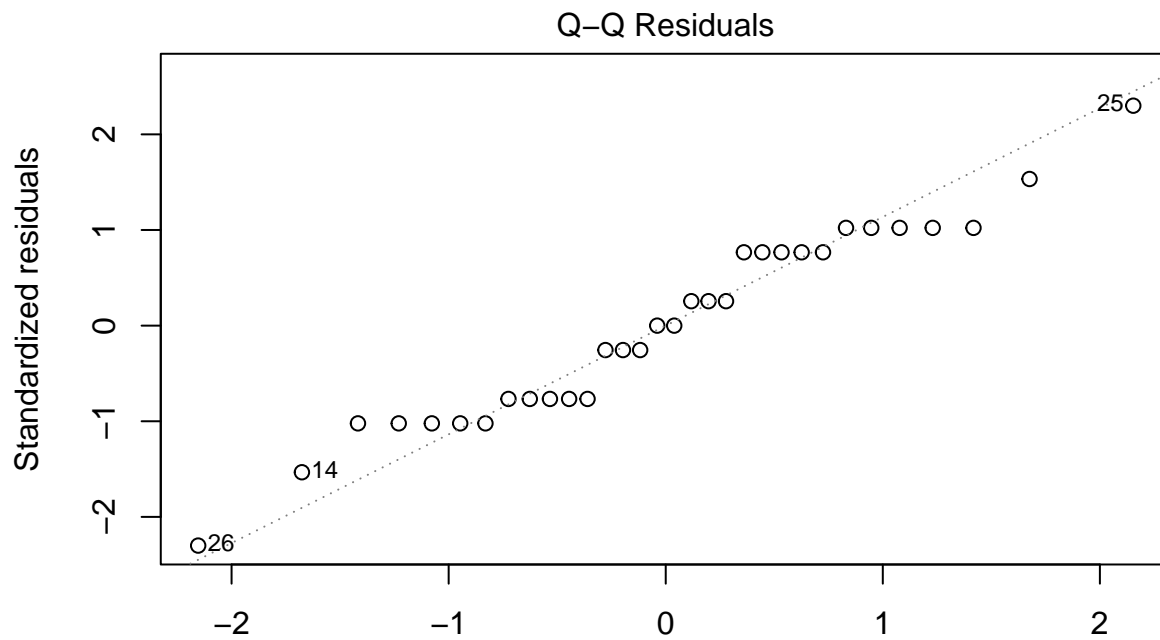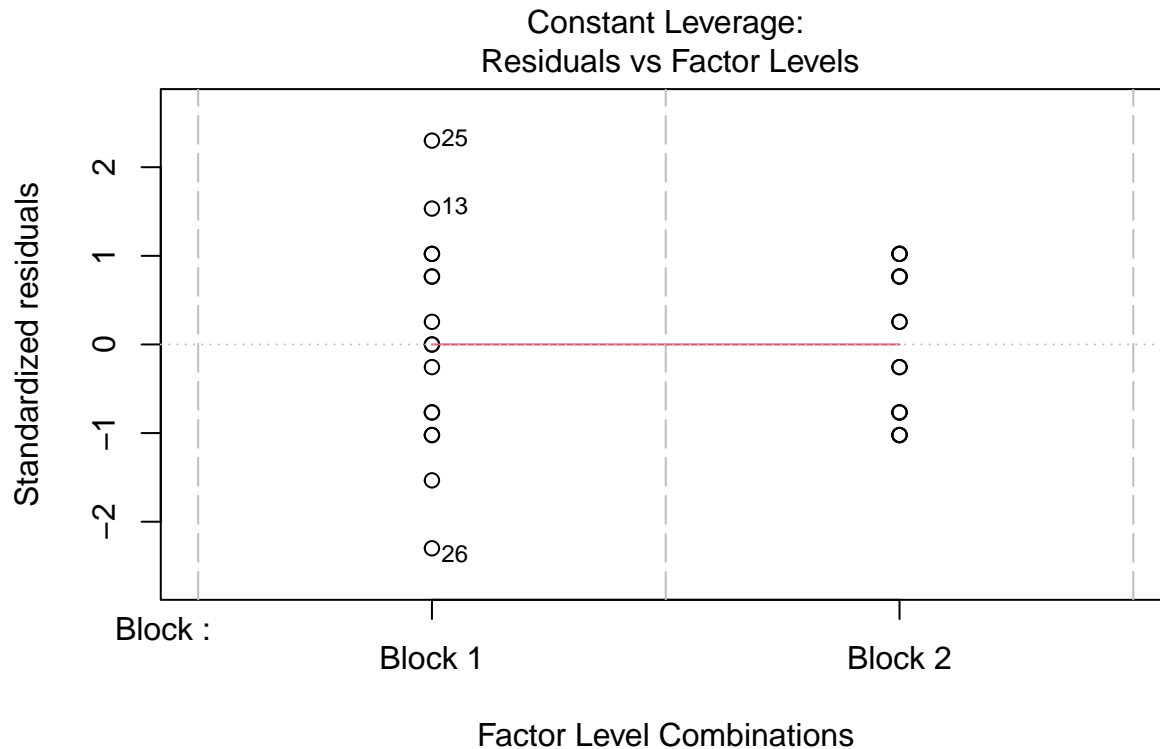
```r
# Effects
effects1 <- 2 * coef(lm(Yield ~ Block + A * B * C * D, data = chemical_process_data))[-1]
effects1
```

```
## BlockBlock 2            A            B            C            D          A:B
##      -4.8750      -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           NA
```

```r
# Plot
plot(model1)
```

Q–Q Residuals

Theoretical Quantiles
aov(Yield ~ Block + A * B * C * D)

Scale–Location

Fitted values
aov(Yield ~ Block + A * B * C * D)

Constant Leverage:
Residuals vs Factor Levels

- Using the data and constructing a design with two blocks of eight observations each with ABCD confounded, the ANOVA results show that the main effects of A, D, and C have a significant effect on the yield. The strongest being A, then D, then C. These three have interaction effects that are seen to be significant such as some two-factor interactions consisting of AB and AD and three factor interactions consisting of ABC and ABD. Thus, we can observe that the block effect is quite significant and validates the design's confounding of the ABCD interaction. The effects of the model also support the findings from the ANOVA.

Problem 2 - 7.14

The design in Problem 6.38 is a 2^3 factorial replicated twice. Suppose that each replicate was a block. Analyze all of the responses from this blocked design. Are the results comparable to those from Problem 6.38? Is the block effect large?

```r
# Data
Apatite <- rep(c(1, 1, 1, 1, 1, 1, 1, 1, -1, -1, -1, -1, -1,
    -1, -1, -1), 2)
pH <- rep(c(1, 1, 1, 1, -1, -1, -1, -1), 4)
Pb <- rep(c(1, 1, -1, -1), 8)
Block <- c(rep(-1, 16), rep(1, 16))

pH_final <- c(5.22, 5.12, 6.84, 6.61, 3.35, 3.34, 5.77, 6.25,
    5.29, 5.06, 5.93, 6.02, 3.39, 3.34, 4.5, 4.74, 3.49, 3.46,
    5.84, 5.9, 2.7, 2.74, 3.36, 3.24, 3.22, 3.22, 5.53, 5.43,
    2.82, 2.79, 3.28, 3.28)
Pb_final <- c(1.82, 1.81, 0.01, 0, 1.11, 1.04, 0, 0.01, 2.11,
    2.18, 0.03, 0.05, 1.7, 1.69, 0.05, 0.05, 0.11, 0.12, 0, 0,
    0.8, 0.76, 0.03, 0.05, 1.03, 1.05, 0, 0, 1.34, 1.26, 0.06,
    0.07)
```

```
ph_pb_data <- data.frame(Block = Block, Apatite = Apatite, pH = pH,
    Pb = Pb, Pb_final = Pb_final, pH_final = pH_final)
ph_pb_data
```

```
##    Block Apatite pH Pb Pb_final pH_final
## 1     -1       1  1  1     1.82     5.22
## 2     -1       1  1  1     1.81     5.12
## 3     -1       1  1 -1     0.01     6.84
## 4     -1       1  1 -1     0.00     6.61
## 5     -1       1 -1  1     1.11     3.35
## 6     -1       1 -1  1     1.04     3.34
## 7     -1       1 -1 -1     0.00     5.77
## 8     -1       1 -1 -1     0.01     6.25
## 9     -1      -1  1  1     2.11     5.29
## 10    -1      -1  1  1     2.18     5.06
## 11    -1      -1  1 -1     0.03     5.93
## 12    -1      -1  1 -1     0.05     6.02
## 13    -1      -1 -1  1     1.70     3.39
## 14    -1      -1 -1  1     1.69     3.34
## 15    -1      -1 -1 -1     0.05     4.50
## 16    -1      -1 -1 -1     0.05     4.74
## 17     1       1  1  1     0.11     3.49
## 18     1       1  1  1     0.12     3.46
## 19     1       1  1 -1     0.00     5.84
## 20     1       1  1 -1     0.00     5.90
## 21     1       1 -1  1     0.80     2.70
## 22     1       1 -1  1     0.76     2.74
## 23     1       1 -1 -1     0.03     3.36
## 24     1       1 -1 -1     0.05     3.24
## 25     1      -1  1  1     1.03     3.22
## 26     1      -1  1  1     1.05     3.22
## 27     1      -1  1 -1     0.00     5.53
## 28     1      -1  1 -1     0.00     5.43
## 29     1      -1 -1  1     1.34     2.82
## 30     1      -1 -1  1     1.26     2.79
## 31     1      -1 -1 -1     0.06     3.28
## 32     1      -1 -1 -1     0.07     3.28
```
```r
# Model
model2 <- aov(pH_final ~ Block + pH * Apatite * Pb)
model2_2 <- aov(Pb_final ~ Block + pH * Apatite * Pb)

cat("\n pH ANOVA:\n")
```

```
##
##  pH ANOVA:
```
```r
summary(model2)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Block        1 13.094  13.094  64.858 3.83e-08 ***
## pH           1 16.951  16.951  83.958 3.87e-09 ***
## Apatite      1  0.908   0.908   4.497   0.0450 *
## Pb           1 17.955  17.955  88.933 2.28e-09 ***
## pH:Apatite   1  0.001   0.001   0.004   0.9473
```

```
## pH:Pb          1  0.518   0.518   2.564   0.1230
## Apatite:Pb     1  0.723   0.723   3.581   0.0711 .
## pH:Apatite:Pb  1  0.049   0.049   0.242   0.6275
## Residuals     23  4.644   0.202
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cat("\n Pb_mM ANOVA:\n")
```

```
##
##   Pb_mM ANOVA:
```

```
summary(model2_2)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## Block           1  1.523   1.523  12.311  0.00189 **
## pH              1  0.003   0.003   0.023  0.88145
## Apatite         1  0.781   0.781   6.317  0.01941 *
## Pb              1 11.907  11.907  96.281 1.09e-09 ***
## pH:Apatite      1  0.001   0.001   0.006  0.93659
## pH:Pb           1  0.018   0.018   0.146  0.70594
## Apatite:Pb      1  0.656   0.656   5.300  0.03071 *
## pH:Apatite:Pb   1  0.003   0.003   0.023  0.88145
## Residuals      23  2.844   0.124
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Fishbone Analysis: Observing the ANOVA for lead concentration, it can be seen that the block effect is significant. Additionally, all interaction factors are highly significant except for Apatite * pH * Pb, suggesting that these factors have a strong affect on lead removal efficiency. Observing the ANOVA for pH, it can be observed that the block effect is signifcant once again. However, the interaction factors that consist of Apatite and pH or Pb are less significant, indicating that there is less complexity in how these factors affect the resulting pH.

- Hydroxyapatite Analysis: Observing the ANOVA for lead concentration, it can be observed that the main effects and their interactions are highly significant in lead removal efficiency. Observing the ANOVA for pH, it can be observed that the main effects of pH and Pb along with their interactions are highly significant. However, the interaction factor of all three factors is not significant, indicating that the combined effect does not alter the pH beyond the two-factor interactions.

- Results in comparison to problem 6.38 are likely similar in terms of identifying which factors are the most influential in the model. However, the specific values may differ depending on how 6.38 does not view the different replicates as blocks for the model.

- The block effect is significant which suggests that there are notable differences between the replicates/blocks. This may be due to experimental setup and error which may result in incorrect analysis of the experiment. Having a large block effect might mitigate the variability across replicates/blocks and provide a more accurate analysis of the experimental results.

Problem 3 - 8.31: Consider the following experiment:

```
run <- 1:8
treatment_combination <- c("d", "ae", "b", "abde", "cde", "ac",
    "bce", "abcd")
exp_data <- data.frame(Run = run, Treatment_Combination = treatment_combination)
exp_data
```

```
##   Run Treatment_Combination
## 1   1                      d
## 2   2                     ae
## 3   3                      b
## 4   4                   abde
## 5   5                    cde
## 6   6                     ac
## 7   7                    bce
## 8   8                   abcd
```

Part (a): How many factors did this experiment investigate?

- The experiment investigated 5 factors(a, b, c, d, and e).

Part (b): How many factors are in the basic design?

- The basic design includes 3 factors because this is an 8-run experiment. Thus $2^3 = 2^{5-2}$ fractional factorial design so there are three factors in this basic design with two factors that are confounded.

Part (c): Assume that the factors in the experiment are represented by the initial letters of the alphabet (i.e. A, B, etc), what are the design generators for the factors beyond the basic design?

- The design generators for the factors beyond the basic design in this experiment are $D = ABC$(D confounded with the ABC interaction) and another is $E = AB$(E confounded with the AB interaction).

Part (d): Is this design a principal fraction?

- With the design generators $D = ABC$ and $E = AB$, this design can be considered a principal fraction under certain conditions. Firstly, D should be aliased with a three-factor interaction. Secondly, E should be confounded with a two-factor interaction. This design is considered a principal fraction because it effectively uses confounding to simplify the experimental setup while also maintain the ability to accurately estimate the most crucial effects in the model.

Part (e): What is the complete defining relation?

- Data Generators: $D = ABC$ and $E = AB$
    - $D * D = ABC * ABC = I$ so $D * D = I$
    - $AB * AB = I$ so $E * E = I$
    - then $DE = ABC * AB = ABCB = AC$(since $B * B = I$)
- Complete defining relation: $I = ABCD = ABE$

Part (f): What is the resolution of this design?

- Given that D is a three-factor interaction and E is a two-factor interaction, the design primarily aliases the main effects with these two types of interactions. Thus, this experiment is a Resolution III design because the main effects of $E = AB$ are aliased with the two-factor interactions. This resolution allows an efficient exploration of the main effects but requires caution when interpreting the interactions between the A and B factors.

Problem 4:

An article in Industrial and Engineering Chemistry("More on Planning Experiments to Increase Research Efficiency", 1970, pp. 60-65) uses a 25-2 design to investigate the effect of A = condensation temperature, B = amount of material 1, C = solvent volume, D = condensation time, and E = amount of material 2 on yield. The results obtained are as follows: e = 23.2 ad = 16.9 cd = 23.9 bde = 16.8 ab = 15.5 bc = 16.2 ace = 23.4 abcde = 18.1

Part (a): Verify that the design generators used were I = ACE and I = BDE.

- I = ACE; I = BDE
- ACE x ACE = I; BDE x BDE = I
- For I = ACE:
  - $(ACE)^2 = I$
  - Aliases:
    * ACE x A = CE x A x A = CE x I = CE
    * ACE x C = AE x C x C = AE x I = AE
    * ACE x E = AC x E x E = AC x I = AC
- For I = BDE:
  - $(BDE)^2 = I$
  - Aliases:
    * BDE x B = DE x B x B = DE x I = DE
    * BDE x D = BE x D x D = BE x I = BE
    * BDE x E = BD x E x E = BD x I = BD
- Therefore the design generators used were I = ACE and I = BDE
- If ACE and BDE are truly equal to I, measurements under these conditions should ideally be near the baseline values but may differ due to experimental variability. If treatment combinations of ace and bde are similar and result in a higher p-value under the ANOVA for the model, then it supports the hypothesis that these are valid generators.

Part (b): Write down the complete defining relation and the aliases for this design.

- I = ACE = BDE
- For ACE:
  - ACE x ACE = I
  - ACE x BDE = ABCDE
- For BDE:
  - BDE x BDE = I
  - BDE x ACE = ABCDE
- For individual factors:
  - ACE x A = CE
  - ACE x B = ABCDE(from AB x CE)
  - ACE x C = AE
  - ACE x D = ABCDE(from AD x CE)
  - ACE x E = AC
  - BDE x A = ABCDE(from AB x DE)
  - BDE x B = DE
  - BDE x C = ABCDE(from BC x DE)
  - BDE x D = BE
  - BDE x E = BD
- Therefore the defining relation is I = ACE = BDE = ABCDE
- Aliases:

- A is aliased with CE (ACE = I), BDE
- B with DE, ACE
- C with AE, BDE
- D with BE, ACE
- E with AC (ACE = I), BD (BDE = I)
- Any two-factor that is aliased with the remaining three factors to create ABCDE i.e. AB aliased with CDE -> ABCDE
- Any three-factor that is aliased with the remaining two factors to create ABCDE i.e. ABC with DE -> ABCDE

Part (c): Estimate the main effects.

```
A <- c(1, -1, 1, 1, -1, 1, -1, -1)
B <- c(1, 1, -1, 1, -1, -1, -1, 1)
C <- c(-1, 1, 1, 1, -1, -1, 1, -1)
E <- A * C
D <- B * E
response <- c(15.5, 16.2, 23.4, 18.1, 23.2, 16.9, 23.8, 16.8)

model3 <- lm(response ~ A * B * C * D * E)
main <- 2 * coef(model3)[2:6]
main
```

```
##      A      B      C      D      E
## -1.525 -5.175  2.275 -0.675  2.275
```

Part (d): Prepare an analysis of variance table. Verify that the AB and AD interactions are available to use as error.

```
effect3 <- na.omit(2 * coef(model3)[-1])
effect3
```

```
##      A      B      C      D      E    A:B    B:C
## -1.525 -5.175  2.275 -0.675  2.275  1.825 -1.275
## attr(,"na.action")
##       A:C       A:D       B:D       C:D       A:E       B:E       C:E       D:E
##         7         9        10        11        12        13        14        15
##     A:B:C     A:B:D     A:C:D     B:C:D     A:B:E     A:C:E     B:C:E     A:D:E
##        16        17        18        19        20        21        22        23
##     B:D:E     C:D:E   A:B:C:D   A:B:C:E   A:B:D:E   A:C:D:E B:C:D:E A:B:C:D:E
##        24        25        26        27        28        29        30        31
## attr(,"class")
## [1] "omit"
```

```
summary(aov(model3))
```

```
##              Df Sum Sq Mean Sq
## A             1   4.65    4.65
## B             1  53.56   53.56
## C             1  10.35   10.35
## D             1   0.91    0.91
## E             1  10.35   10.35
## A:B           1   6.66    6.66
## B:C           1   3.25    3.25
```

```
qqnorm(effect3, plot.it = F)
text(qqnorm(effect3)$x, qqnorm(effect3)$y, names(effect3))
qqline(effect3)
```
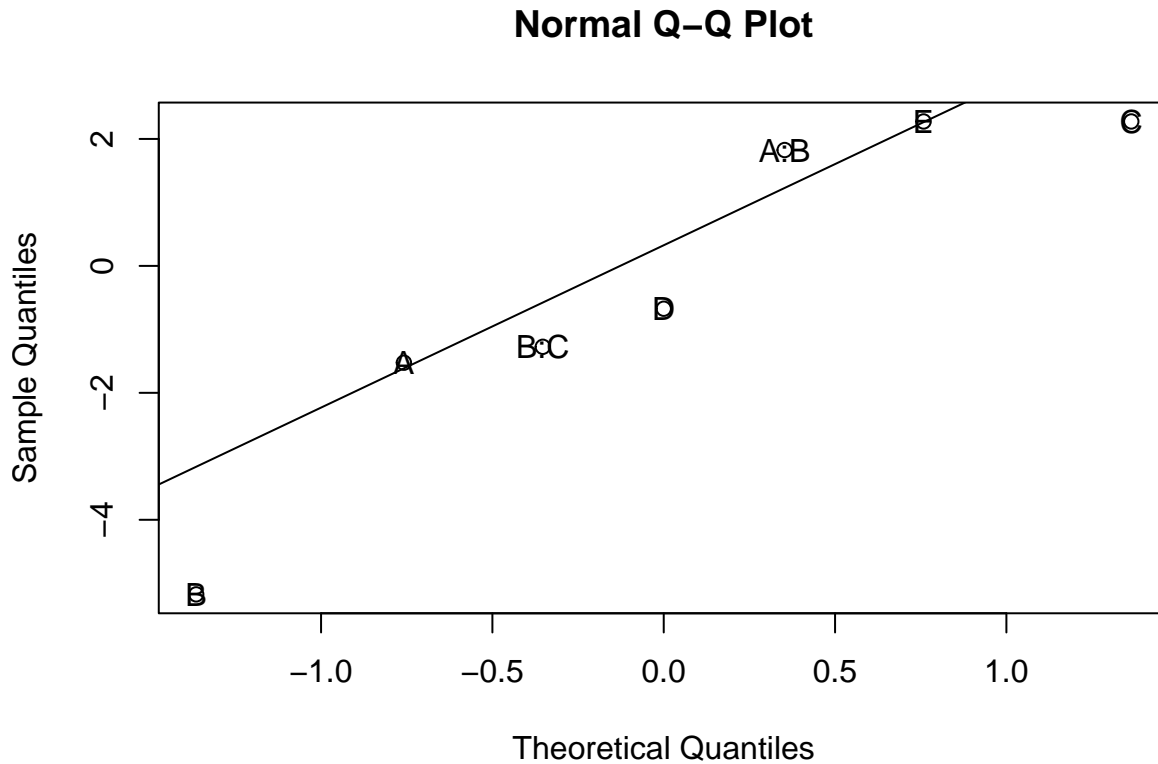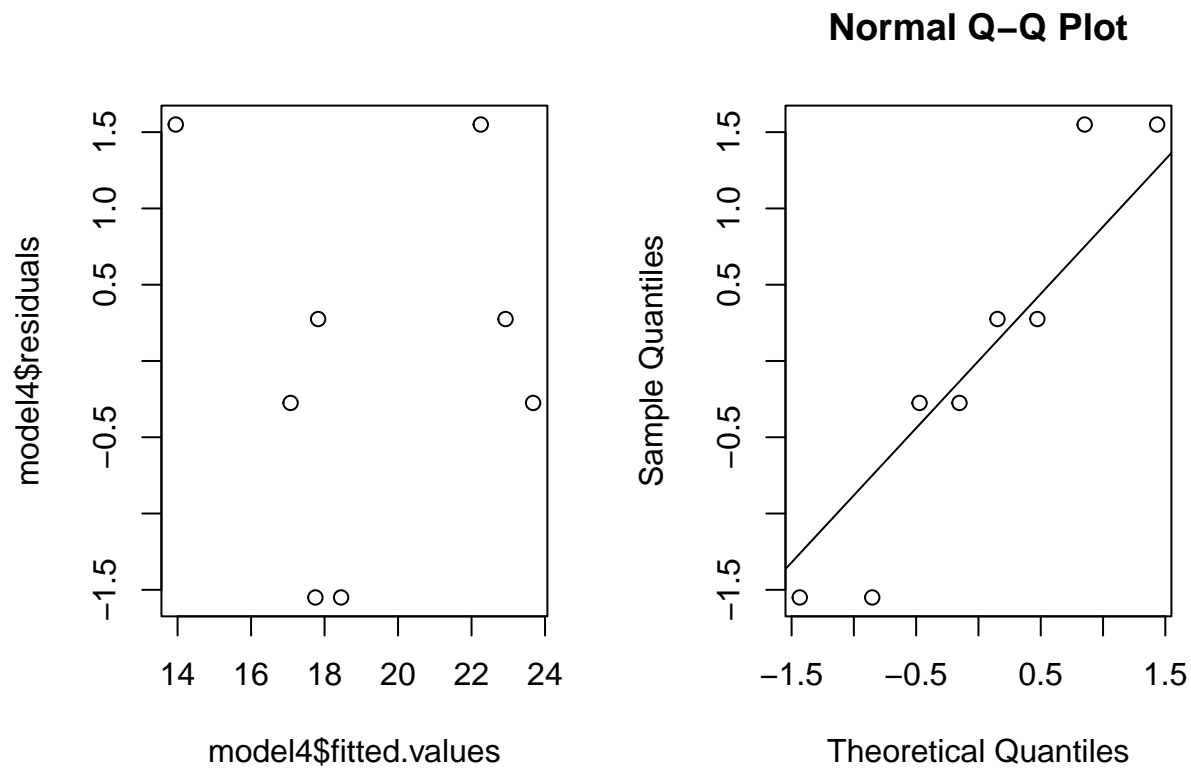
Normal Q–Q Plot

- From the plot, we can observe that the AB interaction does stray from the normal Q-Q line. Due to this, we can say that the AB interaction can be used as error. However, the AD interaction is not plotted nor is in the ANOVA table, indicating that this interaction has little to no significance so it not available to use as an error for the model.

Part (e): Plot the residuals versus the fittted values. Also construct a normal probability plot of the residuals. Comment on the results.

```
model4 <- aov(response ~ factor(A) + factor(B) + factor(C) +
    factor(D) + factor(E))
summary(model4)

##              Df Sum Sq Mean Sq F value Pr(>F)
## factor(A)     1   4.65    4.65   0.938 0.4349
## factor(B)     1  53.56   53.56  10.807 0.0814 .
## factor(C)     1  10.35   10.35   2.089 0.2853
## factor(D)     1   0.91    0.91   0.184 0.7098
## factor(E)     1  10.35   10.35   2.089 0.2853
## Residuals     2   9.91    4.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
par(mfrow = c(1, 2))
plot(model4$fitted.values, model4$residuals)
qqnorm(model4$residuals)
qqline(model4$residuals)
```

Normal Q−Q Plot

- Based on the residuals vs fitted values plot, we can observe that the points are randomly scattered and do not follow a pattern. This indicates that the model does not violate the homoscedasticity assumption. The normal Q-Q plot also displays points relatively close to the Q-Q line, which supports the assumption that the residuals of this model are normally distributed. However, there is a slight S-shaped curve, which may indicate that the tails of the distribution may be heavy.