Math 420 HW 3 solution

Fall 2023

Washington University in St. Louis

Due date: Saturday, 10/28/2023

Instruction:

Please type your answers clearly and show your work neatly. You are encouraged to use the Rmarkdown version of this assignment as a template to submit your work. Unless stated otherwise, all programming references in the assignment will be in R. For this assignment, problems roughly covers content from Fractional Factorial Designs (Ch. 6 of the text)

Problem 1 (3 points)

Create a 2^{7-3} design using `FrF2` with generators $E = ABC, F = ABD, G = ACD$

  (a) Determine the defining relation for this design. Show your work.

Defining relation is the set of columns that are held constant (I). We can identify them starting with the generators: $E = ABC \implies I = ABCE$

$F = ABD \implies I = ABDF$

$G = ACD \implies I = ACDG$

On the other hand, $I^2 = I$, and this results in :

$(ABCE)(ACDF) \implies I = BEDF$

$(ABCE)(ACDG) \implies I = BCDG$

$(ABCE)(ABDF) \implies I = CEDF$

$(ABDF)(ACDG) \implies I = BCFG$

Therefore, the defining relation is the set $I = ABCE = ABDF = ACDG = BEDF = BCDG = CEDF = BCFG$. It is a resolution IV design as the smallest number of word has length 4.

  (b) Determine the confounded strings of two-factor interactions that can be estimated from this design.

This can be done using the `aliases` and the `FrF2` package

```
## set up the fractional design
des1b <- FrF2(nruns = 16, nfactors = 7, generators = c('ABC','ABD','ACD'))
## (.)^2 tells the function to return 2-factor interactions
aliases(lm(1:16~(.)^2, data = des1b))
```

```
##
##  A:B = C:E = D:F
##  A:C = B:E = D:G
##  A:D = B:F = C:G
##  A:E = B:C = F:G
##  A:F = B:D = E:G
```

```
##   A:G = C:D = E:F
##   B:G = C:F = D:E
```

(c) Suppose that after the analysis of data from this experiment, the significant effects appear to be B, $BC + AE + FG$, and $BE + AC + DG$. What foldover fraction would you recommend in order to determine which specific interactions are causing the two strings of confounded interactions to appear significant?

We can create a foldover design by switching the signs of generators that contain factor B, this would clear the effects for B and break the confounding strings of BE and BC.

(d) List the experiments in the foldover fraction you recommend in part (c).

```r
## set up the fractional design
des1c <- fold.design(design = des1b, columns = c(2))
des1c
```

```
##      A  B  C  D     fold  E  F  G
## 1   -1 -1  1 -1 original  1 -1  1
## 2    1 -1 -1  1 original  1 -1 -1
## 3    1  1  1 -1 original  1 -1 -1
## 4   -1 -1 -1 -1 original -1 -1 -1
## 5    1 -1 -1 -1 original  1  1  1
## 6    1 -1  1  1 original -1 -1  1
## 7   -1  1  1 -1 original -1  1  1
## 8    1  1 -1  1 original -1  1 -1
## 9   -1  1 -1  1 original  1 -1  1
## 10   1  1  1  1 original  1  1  1
## 11   1  1 -1 -1 original -1 -1  1
## 12  -1 -1  1  1 original  1  1 -1
## 13  -1  1 -1 -1 original  1  1 -1
## 14  -1  1  1  1 original -1 -1 -1
## 15  -1 -1 -1  1 original -1  1  1
## 16   1 -1  1 -1 original -1  1 -1
## 17  -1  1  1 -1   mirror  1 -1  1
## 18   1  1 -1  1   mirror  1 -1 -1
## 19   1 -1  1 -1   mirror  1 -1 -1
## 20  -1  1 -1 -1   mirror -1 -1 -1
## 21   1  1 -1 -1   mirror  1  1  1
## 22   1  1  1  1   mirror -1 -1  1
## 23  -1 -1  1 -1   mirror -1  1  1
## 24   1 -1 -1  1   mirror -1  1 -1
## 25  -1 -1 -1  1   mirror  1 -1  1
## 26   1 -1  1  1   mirror  1  1  1
## 27   1 -1 -1 -1   mirror -1 -1  1
## 28  -1  1  1  1   mirror  1  1 -1
## 29  -1 -1 -1 -1   mirror  1  1 -1
## 30  -1 -1  1  1   mirror -1 -1 -1
## 31  -1  1 -1  1   mirror -1  1  1
## 32   1  1  1 -1   mirror -1  1 -1
## class=design, type= FrF2.generators.folded
```

Problem 2 (3 points)

6. Consider the experiment to study factors that affect the percent removal of arsenic from drinking water using a iron coated sand filter.

(a) Modify the R code in Section 6.5.1 to produce the design shown in Table 6.8.

```
## original design
des2a <- FrF2(nruns = 8, nfactors = 7,
              generators = c("AB","AC","BC","ABC"),
              randomize = FALSE)
## foldover design
des2a_fold <- fold.design(des2a, columns = 'full')
## convert it into a dataframe for easier manipulation
des2a_fold <- data.frame(des2a_fold)[,c('fold','A','B','C','D','E','F','G')]
## rename 'origina'/'mirror' as a 'block' variable
des2a_fold[,'block']<-ifelse(des2a_fold$fold=='original',1,2)
des2a_fold[,'run'] <- 1:16
des2a_fold[,'y']<-c(69.95,58.65,56.25,53.25,
        94.40,73.45,10.00,2.11,
        16.20,52.85,9.05,31.10,
        7.40,9.90,10.85,48.75)
des2a_fold <- des2a_fold[,c('run','block','A','B','C','D','E','F','G','y')]
```

(b) Determine the defining relation for the experiments in the first block and show the confounding of main effects with strings of two-factor interactions.

The generators of the original design are: $D = AB, E = AC, F = BC$, and $G = ABC$. This implies:

$$I = ABD = ACE = BCF = ABCG$$

On the other hand,

$$(ABD)(ACE) = BCDE, (ABD)(BCF) = ACDF, (ABD)(ABCG) = CDG$$

$$(ACE)(BCF) = ABEF, (ACE)(ABCG) = BEG, (BCF)(ABCG) = AFG$$

Altogether, we have $I = ABD = ACE = BCF = CDG = BEG = ABEF = ABCG = BCDE = ACDF$, and this is a resolution III design

The confounding strings for 2-factor interactions can be obtained using the following code

```
aliases(lm(1:8~(.)^2, data= des2a))
```

```
##
##  A = B:D = C:E = F:G
##  B = A:D = C:F = E:G
##  C = A:E = B:F = D:G
##  D = A:B = C:G = E:F
##  E = A:C = B:G = D:F
##  F = A:G = B:C = D:E
##  G = A:F = B:E = C:D
```

(c) Calculate the effects and make a half-normal plot of them using only the data from block 1 in Table 6.8.

Using only data from block 1:

```
des2a1 <- des2a_fold[des2a_fold$block==1,c('A','B','C','D','E','F','G','y')]

mod2c <- lm(y~(.)^2, data = des2a1)

## model effects
summary(mod2c)

##
## Call:
```

```
## lm.default(formula = y ~ (.)^2, data = des2a1)
##
## Residuals:
## ALL 8 residuals are 0: no residual degrees of freedom!
##
## Coefficients: (21 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   52.258        NaN     NaN      NaN
## A1            -5.393        NaN     NaN      NaN
## B1           -21.855        NaN     NaN      NaN
## C1            -7.268        NaN     NaN      NaN
## D1             2.670        NaN     NaN      NaN
## E1            -1.817        NaN     NaN      NaN
## F1           -17.080        NaN     NaN      NaN
## G1             0.595        NaN     NaN      NaN
## A1:B1             NA         NA      NA       NA
## A1:C1             NA         NA      NA       NA
## A1:D1             NA         NA      NA       NA
## A1:E1             NA         NA      NA       NA
## A1:F1             NA         NA      NA       NA
## A1:G1             NA         NA      NA       NA
## B1:C1             NA         NA      NA       NA
## B1:D1             NA         NA      NA       NA
## B1:E1             NA         NA      NA       NA
## B1:F1             NA         NA      NA       NA
## B1:G1             NA         NA      NA       NA
## C1:D1             NA         NA      NA       NA
## C1:E1             NA         NA      NA       NA
## C1:F1             NA         NA      NA       NA
## C1:G1             NA         NA      NA       NA
## D1:E1             NA         NA      NA       NA
## D1:F1             NA         NA      NA       NA
## D1:G1             NA         NA      NA       NA
## E1:F1             NA         NA      NA       NA
## E1:G1             NA         NA      NA       NA
## F1:G1             NA         NA      NA       NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:     NaN
## F-statistic:   NaN on 7 and 0 DF,  p-value: NA
```

```
## halfnormal plot:
halfnorm(effects = mod2c$effects[2:8],
labs = labels(mod2c),
alpha = 0.2, refline=TRUE)
```
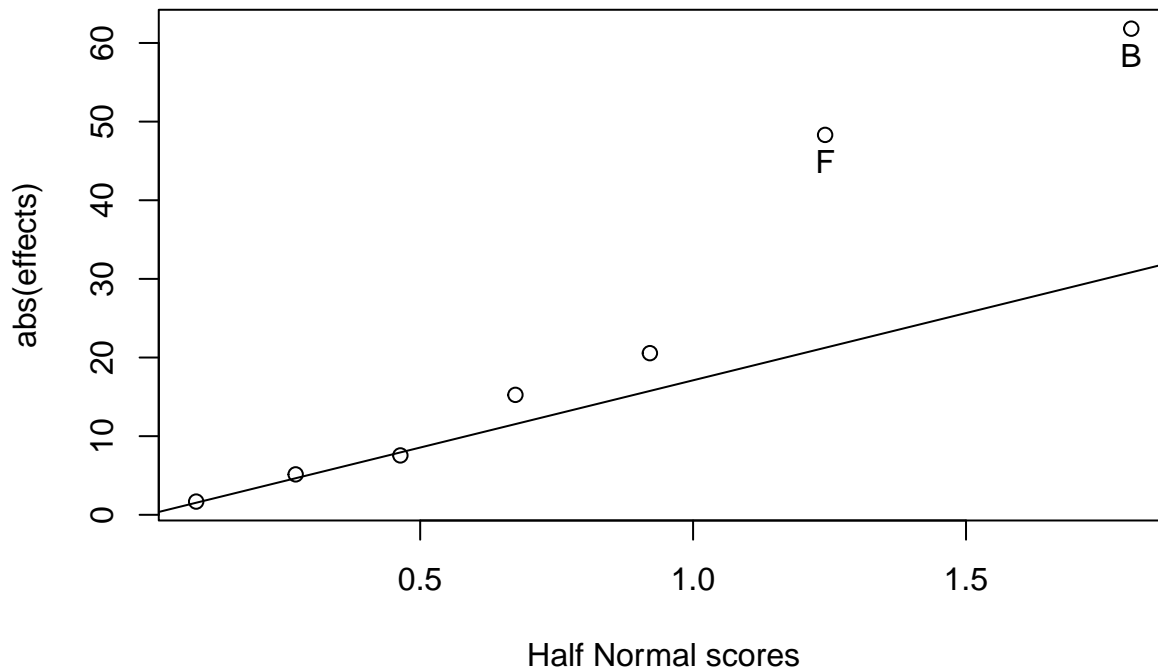
```
## zscore= 0.08964235 0.27188 0.4637078 0.6744898 0.920823 1.241867 1.802743effp= 1.682914 5.140666 7.55
```

(d) Determine the defining relation for the combined experiments.

We can refer back to part (b) and see that $ABCG, BCDE, ACDF, ABEF$ will result in a column of identity in the full design. Notice that, since the full design contains a augmented runs that were generated from a full foldover (i.e. flipping the signs of all the factors), this becomes a resolution IV design as opposed to resolution III in the first block.

(e) Calculate the effects, including the block effect, and make a half-normal plot of them using all the data in Table 6.8. What interactions are confounded with the block effect?

The block factor appears to be a significant effect in the data, this implies that the results generated from the two runs are significantly different. We also see that the block factor confounds with each of the main effects, suggesting that it may be challenging to discern the true effect of factors A-G in this set of augmented runs.

```
mod2e<-lm(y~(.)^2, data = des2a_fold[,c('block','A','B','C','D','E','F','G', 'y')])

## model effect
summary(mod2e)

##
## Call:
## lm.default(formula = y ~ (.)^2, data = des2a_fold[, c("block",
##     "A", "B", "C", "D", "E", "F", "G", "y")])
##
## Residuals:
## ALL 16 residuals are 0: no residual degrees of freedom!
##
## Coefficients: (21 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   81.253        NaN     NaN      NaN
## block        -28.995        NaN     NaN      NaN
## A1             1.603        NaN     NaN      NaN
## B1           -42.035        NaN     NaN      NaN
```

```
## C1             -18.572         NaN    NaN     NaN
## D1               7.940         NaN    NaN     NaN
## E1              -5.923         NaN    NaN     NaN
## F1             -25.260         NaN    NaN     NaN
## G1               7.440         NaN    NaN     NaN
## block:A1        -6.995         NaN    NaN     NaN
## block:B1        20.180         NaN    NaN     NaN
## block:C1        11.305         NaN    NaN     NaN
## block:D1        -5.270         NaN    NaN     NaN
## block:E1         4.105         NaN    NaN     NaN
## block:F1         8.180         NaN    NaN     NaN
## block:G1        -6.845         NaN    NaN     NaN
## A1:B1              NA           NA     NA      NA
## A1:C1              NA           NA     NA      NA
## A1:D1              NA           NA     NA      NA
## A1:E1              NA           NA     NA      NA
## A1:F1              NA           NA     NA      NA
## A1:G1              NA           NA     NA      NA
## B1:C1              NA           NA     NA      NA
## B1:D1              NA           NA     NA      NA
## B1:E1              NA           NA     NA      NA
## B1:F1              NA           NA     NA      NA
## B1:G1              NA           NA     NA      NA
## C1:D1              NA           NA     NA      NA
## C1:E1              NA           NA     NA      NA
## C1:F1              NA           NA     NA      NA
## C1:G1              NA           NA     NA      NA
## D1:E1              NA           NA     NA      NA
## D1:F1              NA           NA     NA      NA
## D1:G1              NA           NA     NA      NA
## E1:F1              NA           NA     NA      NA
## E1:G1              NA           NA     NA      NA
## F1:G1              NA           NA     NA      NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:     NaN
## F-statistic:   NaN on 15 and 0 DF,  p-value: NA
## halfnormal plot:
halfnorm(effects = mod2e$effects[2:16],
labs = labels(mod2e),
alpha = 0.2, refline=TRUE)
```
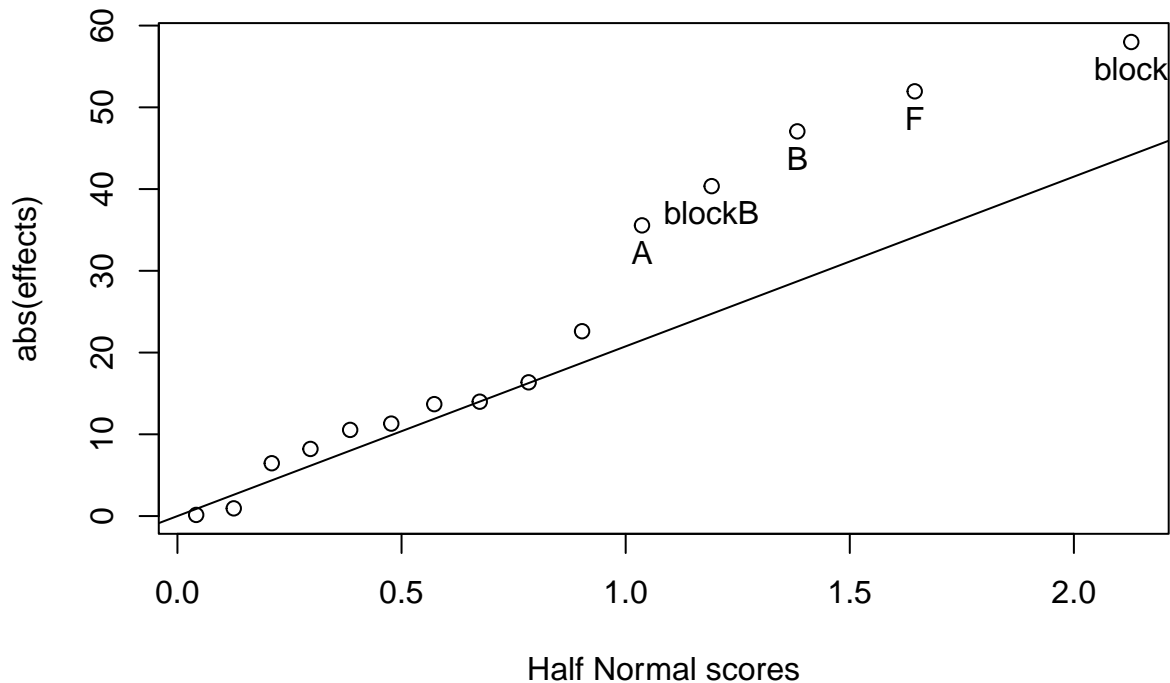
```
## zscore= 0.0417893 0.1256613 0.2104284 0.2967378 0.3853205 0.4770404 0.5729675 0.6744898 0.7835004 0.
```

```
## check confounding patterns
```

```
X<-as.matrix(data.frame(model.matrix(mod2e)))
XtX <-t(X)%*%X
```

Problem 3 (4 points)

8. Prince (2007) performed a 2^{5-2} fractional factorial in the process of optimizing a clinical assay to detect Streptococcus pyogenes with real-time PCR. Optimization of this process would allow hospitals to detect Strep infections in less than 30 minutes with 99 accuracy. The factors he studied were: A=Number of S. pyogenes colonies (1 or 4), B=Boiling time (1 min. or 3 min.), C=Centrifuge time (0 min. or 2 min.), D=cycling temperature (60 or 61), E=Cycling time (5/10 sec. or 8/13 sec.).

The generators for the design were $D = AC$ and $E = BC$.

(a) The response data (in standard order) from the eight experiments were: 1.31091, 1.43201, 1.29951, 1.37199, 1.33566, 1.46820, 1.39023, 1.41531. Calculate the effects and make a half-normal plot of the effects. Identify any effects you believe may be significant.

The design and dataset is assembled in the code below. Setting $\alpha = 0.2$, we can see that effects from A and C appear significant according to the half normal plot.

```
des3a <- FrF2(nruns = 8, nfactors = 5,
generators = c('AC','BC'),randomize = FALSE)
```

```
## aliases structure:
aliases(lm(1:8~(.)^4, data = des3a))
```

```
##
##  A = C:D = B:D:E = A:B:C:E
##  B = C:E = A:D:E = A:B:C:D
##  C = B:E = A:D
##  D = A:B:E = B:C:D:E = A:C
```

```
##  E = B:C = A:B:D = A:C:D:E
##  A:B = D:E = A:C:E = B:C:D
##  A:E = B:D = A:B:C = C:D:E
```

```
y3a <- c(1.31091, 1.43201, 1.29951, 1.37199, 1.33566, 1.46820, 1.39023, 1.41531)

des3a<- add.response(des3a, y3a)
mod3a <- lm(y3a ~ (.)^2, data = des3a)
summary(mod3a)
```

```
##
## Call:
## lm.default(formula = y3a ~ (.)^2, data = des3a)
##
## Residuals:
## ALL 8 residuals are 0: no residual degrees of freedom!
##
## Coefficients: (8 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.377977        NaN     NaN      NaN
## A1           0.043900        NaN     NaN      NaN
## B1          -0.008717        NaN     NaN      NaN
## C1           0.024373        NaN     NaN      NaN
## D1          -0.004495        NaN     NaN      NaN
## E1           0.009137        NaN     NaN      NaN
## A1:B1       -0.019510        NaN     NaN      NaN
## A1:C1             NA         NA      NA       NA
## A1:D1             NA         NA      NA       NA
## A1:E1       -0.007355        NaN     NaN      NaN
## B1:C1             NA         NA      NA       NA
## B1:D1             NA         NA      NA       NA
## B1:E1             NA         NA      NA       NA
## C1:D1             NA         NA      NA       NA
## C1:E1             NA         NA      NA       NA
## D1:E1             NA         NA      NA       NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:     NaN
## F-statistic:   NaN on 7 and 0 DF,  p-value: NA
```

```
## half normal plot, skipping overall effect
halfnorm(effects = mod3a$effects[2:8],
labs = labels(mod3a),
alpha = 0.2, refline=TRUE)
```
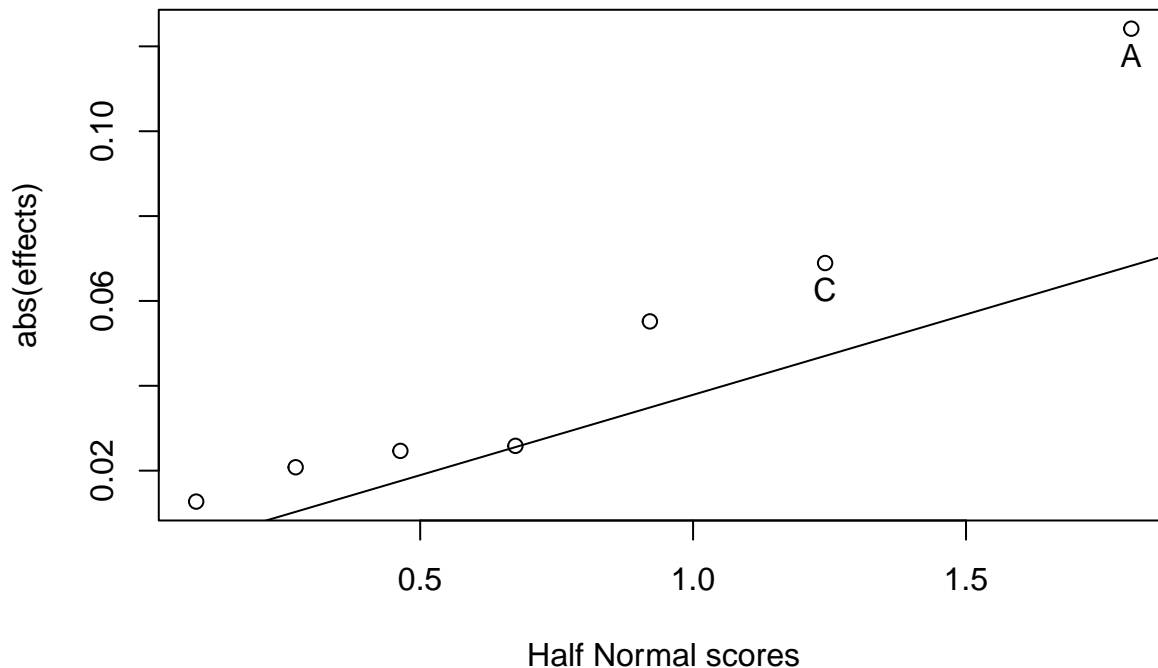
```
## zscore= 0.08964235 0.27188 0.4637078 0.6744898 0.920823 1.241867 1.802743effp= 0.01271378 0.02080308
```

(b) Determine the confounding pattern or alias structure for the design.

See above in (a)

(c) What are the aliases of the largest effect in absolute value?

The largest effect for this experiment appears to be factor A which has aliases CD and some other higher order terms.

(d) Prince performed eight more experiments according to a foldover (similar to that shown in Table 6.8) and the resulting data were: 1.31702, 1.38881, 1.32222, 1.36248, 1.33826, 1.32654, 1.32654, 1.34635. Combining this data with the data from the original 8 runs, calculate the 15 effects including a block effect for the differences in the two groups of experiments.

The code below shows the combined data and the effects of the data. Overall, the block:A interaction, the block effect, as well as effects from factors A, C all appear as significant based on an alpha value of α.

```
## reusing code from problem 2
des3d <- fold.design(des3a, columns = 'full')
des3d_fold <- data.frame(des3d)[,c('fold','A','B','C','D','E')]
des3d_fold[,'block']<-ifelse(des3d_fold$fold=='original',1,2)
des3d_fold[,'run'] <- 1:16
des3d_fold <- des3d_fold[,c('run','block','A','B','C','D','E')]
y3d <- c(1.31091, 1.43201, 1.29951, 1.37199, 1.33566, 1.46820, 1.39023, 1.41531,
1.31702, 1.38881, 1.32222, 1.36248, 1.33826, 1.32654, 1.32654, 1.34635)
des3d_fold[,'y']<-y3d
des3d_fold <- des3d_fold[,c('run','block','A','B','C','D','E','y')]


mod3d <- lm(y ~ (.)^3, data = des3d_fold[,c('block','A','B','C','D','E','y')])
summary(mod3d)

##
## Call:
```

9

```
## lm.default(formula = y ~ (.)^3, data = des3d_fold[, c("block",
##     "A", "B", "C", "D", "E", "y")])
##
## Residuals:
## ALL 16 residuals are 0: no residual degrees of freedom!
##
## Coefficients: (26 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.41493        NaN     NaN      NaN
## block       -0.03695        NaN     NaN      NaN
## A1           0.10282        NaN     NaN      NaN
## B1          -0.01907        NaN     NaN      NaN
## C1           0.04214        NaN     NaN      NaN
## D1          -0.02198        NaN     NaN      NaN
## E1           0.02193        NaN     NaN      NaN
## block:A1    -0.05892        NaN     NaN      NaN
## block:B1     0.01035        NaN     NaN      NaN
## block:C1    -0.01777        NaN     NaN      NaN
## block:D1     0.01749        NaN     NaN      NaN
## block:E1    -0.01279        NaN     NaN      NaN
## A1:B1       -0.03902        NaN     NaN      NaN
## A1:C1            NA          NA      NA       NA
## A1:D1            NA          NA      NA       NA
## A1:E1       -0.02259        NaN     NaN      NaN
## B1:C1            NA          NA      NA       NA
## B1:D1            NA          NA      NA       NA
## B1:E1            NA          NA      NA       NA
## C1:D1            NA          NA      NA       NA
## C1:E1            NA          NA      NA       NA
## D1:E1            NA          NA      NA       NA
## block:A1:B1  0.01951        NaN     NaN      NaN
## block:A1:C1      NA          NA      NA       NA
## block:A1:D1      NA          NA      NA       NA
## block:A1:E1  0.01524        NaN     NaN      NaN
## block:B1:C1      NA          NA      NA       NA
## block:B1:D1      NA          NA      NA       NA
## block:B1:E1      NA          NA      NA       NA
## block:C1:D1      NA          NA      NA       NA
## block:C1:E1      NA          NA      NA       NA
## block:D1:E1      NA          NA      NA       NA
## A1:B1:C1         NA          NA      NA       NA
## A1:B1:D1         NA          NA      NA       NA
## A1:B1:E1         NA          NA      NA       NA
## A1:C1:D1         NA          NA      NA       NA
## A1:C1:E1         NA          NA      NA       NA
## A1:D1:E1         NA          NA      NA       NA
## B1:C1:D1         NA          NA      NA       NA
## B1:C1:E1         NA          NA      NA       NA
## B1:D1:E1         NA          NA      NA       NA
## C1:D1:E1         NA          NA      NA       NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:    NaN
## F-statistic:   NaN on 15 and 0 DF,  p-value: NA
```

```
## half normal plot for the 15 effects available in the model, skipping overall effect
halfnorm(effects = mod3d$effects[2:16],
labs = labels(mod3d),
alpha = 0.35, refline=TRUE)
```
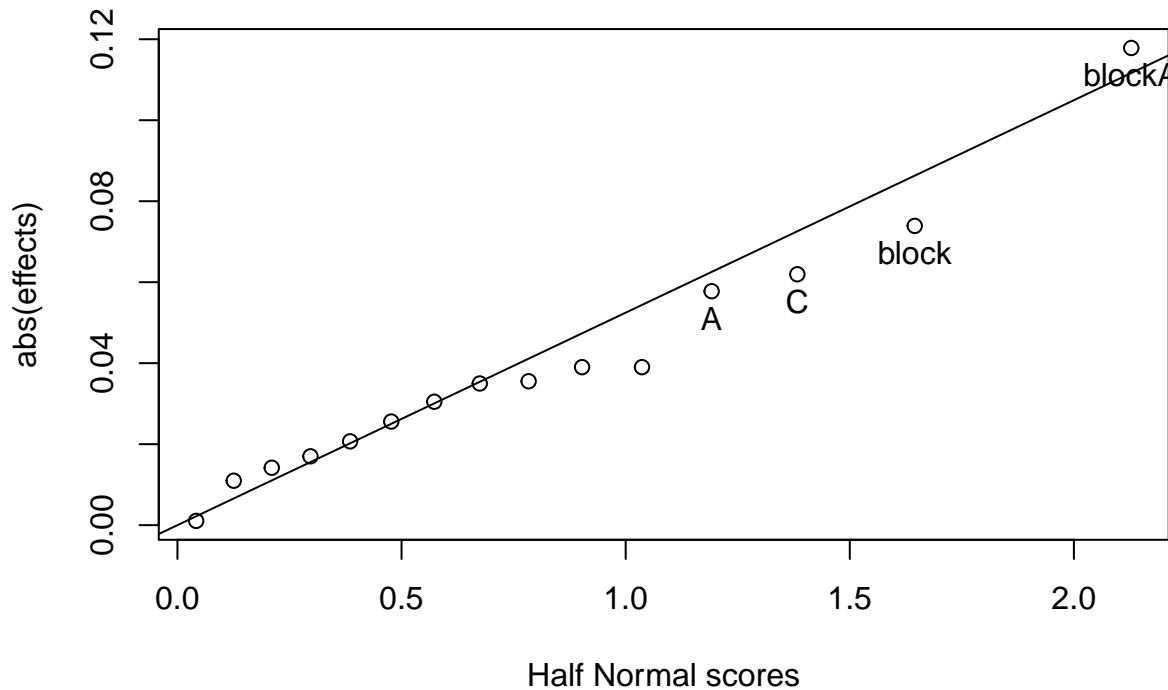

Half Normal scores

```
## zscore= 0.0417893 0.1256613 0.2104284 0.2967378 0.3853205 0.4770404 0.5729675 0.6744898 0.7835004 0.9
```

(e) What is the defining relationship and alias structure for the complete set of 16 experiments?

The generators of the original design are $D = AC$ and $E = BC$. This implies $I = ACD = BEC = ABDE$, combining with a augmented runs generated by $D = -AC$, $E = -BC$, we can check and see that the only combination left that result in a column of identify for the full dataset is $I = ABDE$. This is a resolution IV design and the alias structure can be obtained using the following code

```
aliases(mod3d)
```

```
##
##  block = -A:C:D = -B:C:E
##  A = -block:C:D = B:D:E
##  B = -block:C:E = A:D:E
##  C = -block:B:E = -block:A:D
##  D = A:B:E = -block:A:C
##  E = -block:B:C = A:B:D
##  block:A = -C:D
##  block:B = -C:E
##  block:C = -A:D = -B:E
##  block:D = -A:C
##  block:E = -B:C
##  A:B = D:E
##  A:E = B:D
##  block:A:B = block:D:E = -A:C:E = -B:C:D
##  block:A:E = block:B:D = -A:B:C = -C:D:E
```

(f) Make a half-normal plot of the 15 effects and determine what appears to be significant from the
```

combined set of data.

Skip grading this - this seems to be the same ask as part (d)

(g) Provide an interpretation for all effects and interactions you determine are significant, and discuss how the conclusions and interpretation after 16 experiments differs from what could have been concluded after the first 8 experiments.

Factors A (number of S. pyogenes colonies) and C (centrifuge time) both appear to have significant in both experiments. In the first run, however there were a few confounding patterns (e.g. C confounded with AD) which can make the interpretation more difficult. Since if A has significant effect, and if C is confounded with AD, then it is possible that the significance from C is partially due to A.

In the full set of 16 runs, the design has cleared all main effects. This implies that the significance from effects of factors A and C should not be confounded with any other 2-factor interactions. However, one caveat is that we are seeing the block factor showing up as significant, indicating that there could be some nuisance factors associated with the way each set of experiments were run that may not have been addressed.