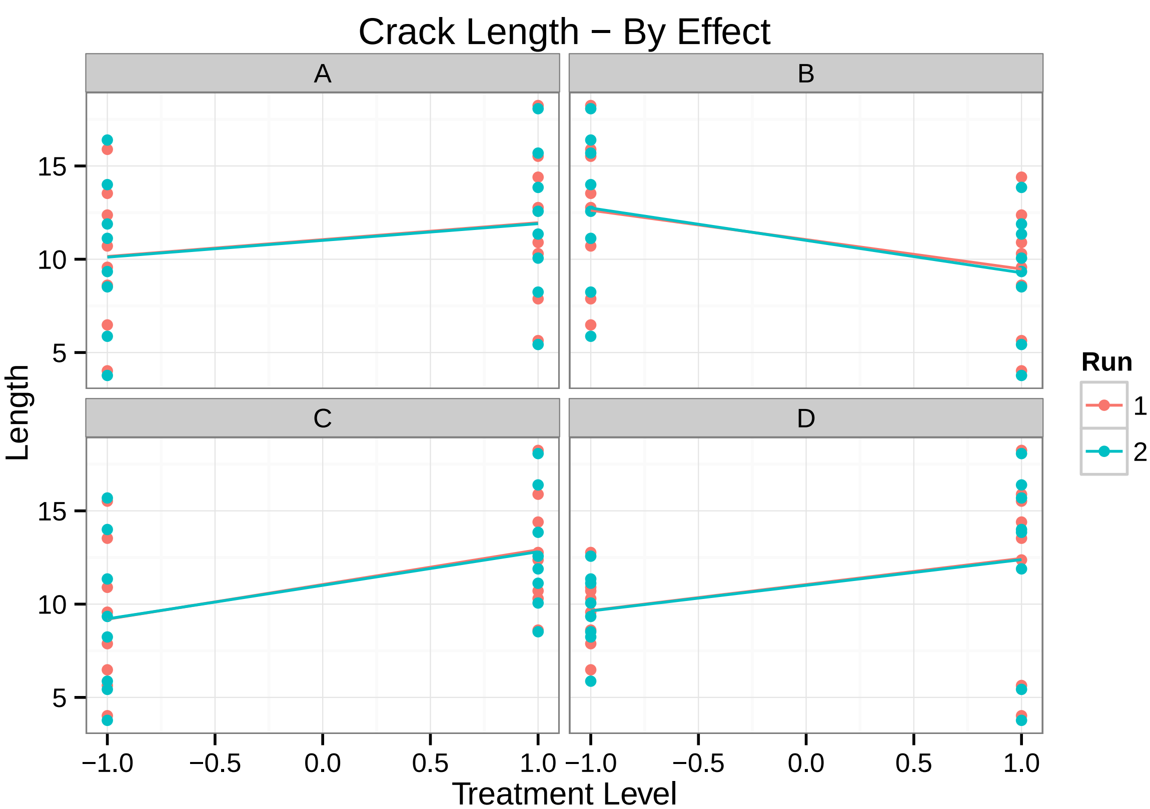
Homework 5 - 6413A

Peter Williams

4.17) *A metal alloy is used to make components for aircraft engines. Cracking is a potentially serious problem in the final part, as it can lead to a non-recoverable failure. The objective of the experiment was to identify the key factor and possibly their interactions which have effect on cracks. Four factors were considered: pouring temperature (A), titanium content (B), heat treatment method (C), and the amount of grain refiner used (D). A 2^4 experiment was conducted and the reponse of interest was the length of crack (in mm x 10^{-2}). Each trial condition was replicated twice. The data are given in Table 4.18.*

A quick visualization of the main effects of the data highlight that length tends to be slightly higher with positive settings on A, B, C, and under negative for B. There doesn't appear to be a significant difference in the main effects across the two runs completed for the experiment.

```
##     A  B  C  D      s2      lns2 Run Length
## 1 -1 -1 -1 -1 0.18605 -1.681740   1   6.48
## 2 -1 -1 -1  1 0.11045 -2.203192   1  13.53
## 3 -1 -1  1 -1 0.08405 -2.476343   1  10.71

## [1] "---"

##    A B  C  D      s2      lns2 Run Length
## 30 1 1 -1  1 0.02205 -3.814443   2   5.43
## 31 1 1  1 -1 0.02880 -3.547380   2  10.06
## 32 1 1  1  1 0.15125 -1.888821   2  13.85
```



(a) *Which main effects are significant for crack length? Are there any significant interactions?*

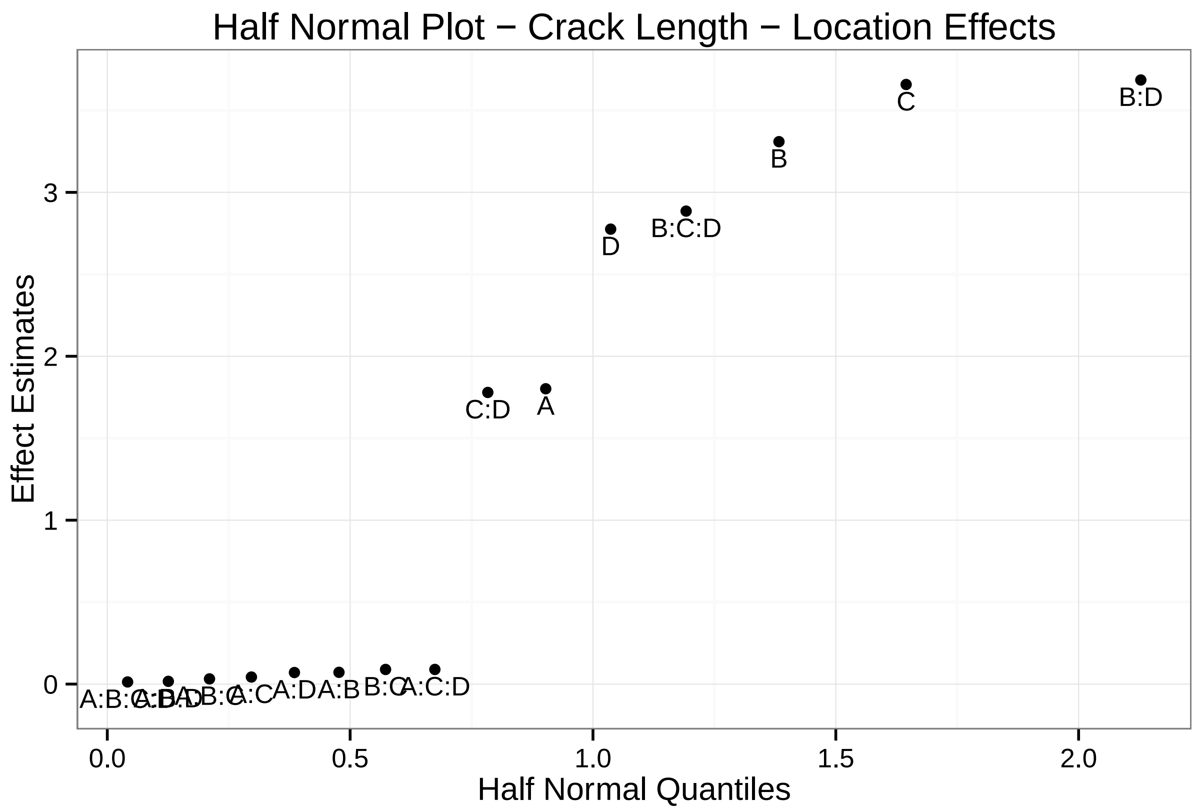
As shown below, analyzing the effects with ANOVA, all four main effects are estimated to be significant sources of variation under $\alpha = 0.001$. Second order interactions between B & D (titanium content & grain refiner amount), C & D (head treatment & grain refiner), and a third order interaction between B & C & D have effects estimated to be significant contributors to variation under $\alpha = 0.001$.

This significance of these effects can also be shown visually by the half normal plot below:

```r
#ybar
effModel <- aov(Length~A*B*C*D,data=crackData)
summary(effModel)
```

```
##             Df Sum Sq Mean Sq  F value   Pr(>F)    
## A            1  25.97   25.97  377.959 1.48e-12 ***
## B            1  87.62   87.62 1274.932  < 2e-16 ***
## C            1 107.06  107.06 1557.801  < 2e-16 ***
## D            1  61.63   61.63  896.843 1.76e-15 ***
## A:B          1   0.04    0.04    0.601    0.449    
## A:C          1   0.01    0.01    0.216    0.648    
## B:C          1   0.06    0.06    0.930    0.349    
## A:D          1   0.04    0.04    0.581    0.457    
## B:D          1 108.67  108.67 1581.311  < 2e-16 ***
## C:D          1  25.33   25.33  368.578 1.80e-12 ***
## A:B:C        1   0.01    0.01    0.118    0.735    
## A:B:D        1   0.00    0.00    0.033    0.858    
## A:C:D        1   0.06    0.06    0.930    0.349    
## B:C:D        1  66.61   66.61  969.337 9.56e-16 ***
## A:B:C:D      1   0.00    0.00    0.020    0.889    
## Residuals   16   1.10    0.07                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

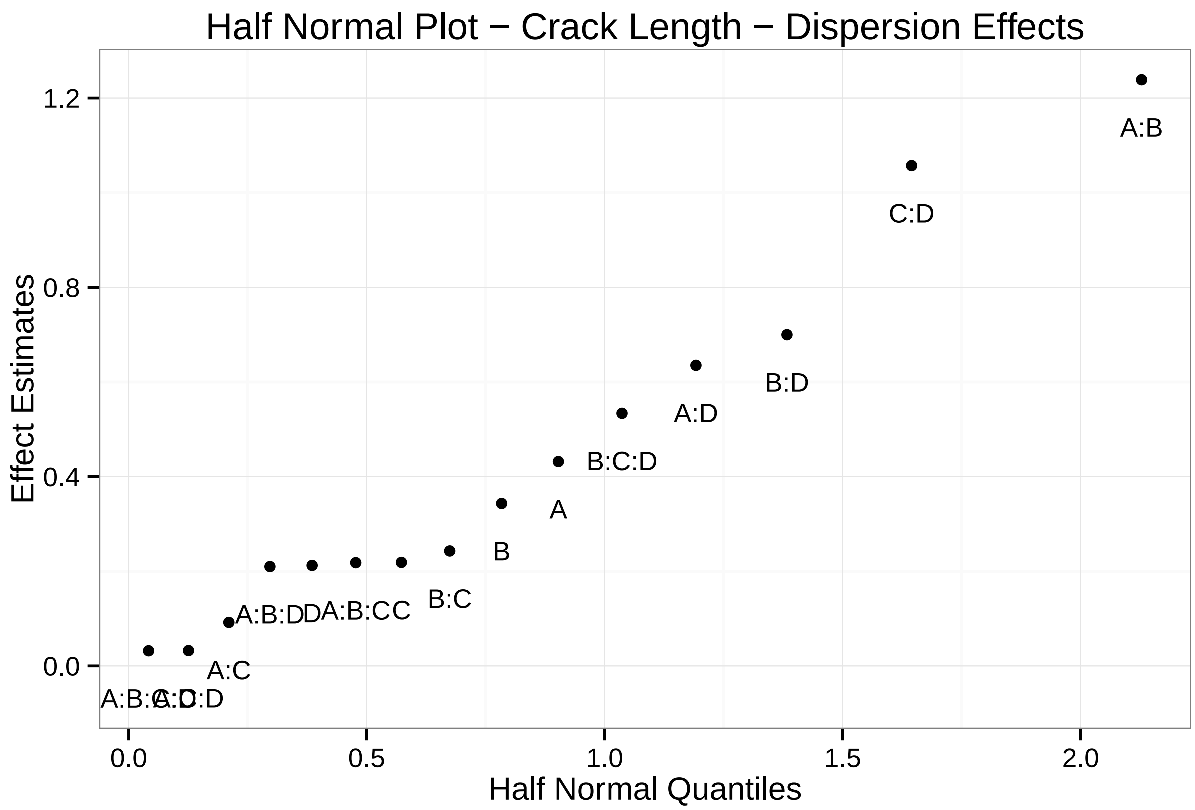
```r
#lns2
dispModel <- aov(lns2~A*B*C*D,data=crackData)


coefData <- data.frame(coef(effModel)*2);
coefData$lns2 <- coef(dispModel)*2
colnames(coefData) <- c('ybar','lns2');
coefData$Effect <- row.names(coefData);
row.names(coefData) <- NULL;
coefData <- coefData[2:nrow(coefData),]#take out intercept
coefData$absYbar <- abs(coefData$ybar)
coefData <- coefData[order(coefData$absYbar),] #rank order
coefData$rankOrder <- 1:nrow(coefData)
coefData$qNorm <- qnorm(0.5+0.5*(coefData$rankOrder-0.5)/nrow(coefData))


#half normal plot - effect
ggplot(coefData, aes(x=qNorm,y=absYbar)) + geom_point() +
  geom_text(aes(x=qNorm,y=absYbar-0.1,label=Effect,size=2)) + theme_bw() +
  xlab('Half Normal Quantiles') +
  ylab('Effect Estimates') +
  ggtitle('Half Normal Plot - Crack Length - Location Effects') +
  theme(legend.position='none')
```




```
#half normal plot - dispersions
```

(b) *What factors affect the variability of the crack length?* Below is a half-normal plot of the dispersion effects. It highlights variability sourced from second level interactions largest for AB, and CD.

```
coefData$absLns2 <- abs(coefData$lns2)
coefData <- coefData[order(coefData$absLns2),] #rank order
coefData$rankOrder <- 1:nrow(coefData)
coefData$qNorm <- qnorm(0.5+0.5*(coefData$rankOrder-0.5)/nrow(coefData))
ggplot(coefData, aes(x=qNorm,y=absLns2)) + geom_point() +
  geom_text(aes(x=qNorm,y=absLns2-0.1,label=Effect,size=2)) + theme_bw() +
  xlab('Half Normal Quantiles') +
  ylab('Effect Estimates') +
  ggtitle('Half Normal Plot - Crack Length - Dispersion Effects') +
  theme(legend.position='none')
```



(c) *Find the optimum process conditions to minimize mean crack length*

Based on the factorial effect fits, crack length is estimated to be minimized under:

1. Pouring temperature, negative setting, A(-),
2. Titanium content, positive setting, B(+),
3. Heat treatment, negative setting C(-), and
4. Grain refiner, positive setting D(+)

These experimental conditions are demonstrated in the 6th (Run1), and 22nd (Run2) record in the dataset.

4.18) *A biologist performed a* 2^5 *factorial design in four blocks of equal size by making use of the blocking variables:* $B_1 = 1234, B_2 = 2345$. *If block effects are present the experiment is performed, how will the main effect of variable 5 be affected? How will the interaction 15 be affected?*

This blocking scheme confounds with the following interactions: $1234, 2345, 15$, since $B_1 B_2 = 1$ X 2 X 2 X 3 X 3 X 4 X 4 X 5 = 15. So this blocking scheme will not be confounded with any of the main effects including 5, but if there were a blocking effect it would be confounded with interaction 15.

4.21) *Difficulties with Nonorthogonal Blocking*

a) *Suppose that the* 2^3 *design in the Table 4.12 is arranged in two blocks with runs 1,2,5,7 in block I and runs 3,4,6,8 in block II. Show that the corresponding blocking variable is not identical to any of the first seven columns of the table and, therefore, it is not confounded with any of the factorial effects.*

Here is the design matrix, with A = 1, B=2, C = 3 for table 4.12.

```
dat <- expand.grid(A = c(-1,1), B=c(-1,1), C=c(-1,1))
dat <- dat[order(dat$A),]
```

```r
dat <- dat[c(1,3,2,4,5,7,6,8),]
dat <- transform(dat, AB = A*B, AC = A*C, BC = B*C, ABC = A*B*C)
dat$AB <- dat$A*dat$B
dat$AC <- dat$A*dat$C
dat$BC <- dat$B*dat$C
dat$ABC <- dat$A*dat$B*dat$C
dat$Block <- c(1,1,-1,-1,1,-1,1,-1) # (+, 1,2,5,7, -, 3,4,6,8)
row.names(dat) <- NULL;
dat
```

```
##    A  B  C AB AC BC ABC Block
## 1 -1 -1 -1  1  1  1  -1     1
## 2 -1 -1  1  1 -1 -1   1     1
## 3 -1  1 -1 -1  1 -1   1    -1
## 4 -1  1  1 -1 -1  1  -1    -1
## 5  1 -1 -1 -1 -1  1   1     1
## 6  1 -1  1 -1  1 -1  -1    -1
## 7  1  1 -1  1 -1 -1  -1     1
## 8  1  1  1  1  1  1   1    -1
```

This blocking scheme can be viewed as a column vector of the form $(++--+-+-)$. Which as demostrated below, is not identical to any of the 7 factorical effects (since there are true and false matches in all columnwise comparisons here):

```r
print('Block vector vs. factorial effects - comparison of vector elements:')
```

```
## [1] "Block vector vs. factorial effects - comparison of vector elements:"
```

```r
lapply(dat[,1:7], function(a) a == dat$Block)
```

```
## $A
## [1] FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE FALSE
##
## $B
## [1] FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE
##
## $C
## [1] FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE
##
## $AB
## [1]  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE
##
## $AC
## [1]  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
##
## $BC
## [1]  TRUE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE
##
## $ABC
## [1] FALSE  TRUE FALSE  TRUE  TRUE  TRUE FALSE FALSE
```

b) *Show that the block effect is not orthogonal to some of the factorial effects. Identify these effects. How does it affect the estimation of these effects? Discuss why it is undesirable to use this blocking scheme.*

Using the alias function in R to find linearly dependent terms highlights that the interaction AC and the Block variable are not orthogonal, and thus cannot be estimated together. That is, 13 (AC) can be expressed as a linear combination of 2 (B), 3 (C), 12 (B) and the Block variable:

```
dat$y <- rnorm(8) #arbitrary response
alias(y~Block+A*B*C,data = dat)
```

```
## Model :
## y ~ Block + A * B * C
##
## Complete :
##     (Intercept) Block A  B  C  A:B B:C A:B:C
## A:C  0          -2     0 -1 -1  1   0   0
```

```
#
dat$Block*-2 + (dat$B + dat$C) * -1 + dat$AB == dat$AC
```

```
## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

Thus the block effect is confounded with the second order interaction 13 (AC). The blocking effect is not clear or strongly clear, and is undesirable since it is confounded with a lower order interaction, whose effect could not be estimated from the model.