Garden Sprinkler Experiment

Haolan Mai

I. Introduction

Watering a garden or lawn by hand is not an easy task. If the garden is large, it may take up a lot of time. In order to promote garden watering, garden sprinklers appeared. Nowadays, there are a lot of benefits we can get from installing a garden sprinkler, and the automatic sprinkler system also is the best investment for the garden. It can bring healthy and beautiful lawns. While saving money, it also saves a lot of time, allowing people to devote more time to activities they really like.

- Convenience: We can set a sprinkler system that will automatically water the garden at fixed points and water the lawn instead of watering by ourselves, which could save valuable time and have more leisure time. Regular and quantitative watering is also more convenient to keep the lawn healthy and green, without drying up.

- Aesthetics and safety: There's nothing attractive about a garden hose stretching across the lawn. The hose is also a tripping hazard for children and pets playing in the yard. In contrast, the garden sprinkler is much more aesthetic and safe than the hose which makes it a more pleasing option.

- Water the optimal amount: Advanced garden sprinkler systems feature weather and soil moisture sensors to deliver the right amount of water right when it's needed, and watering more evenly than ourselves. In addition, the automatic irrigation system can be programmed to discharge a more precise amount of water in the target area, thereby promoting water conservation (saving money).

The goals of our project are to identify the factors that significantly affect the water consumption and spray range and figure out the optimal factors' setting to minimize the water consumption and maximize the spray range. To achieve the goals, we will design the experiment , analyze the experimental results and draw the conclusions.

Research Question:
- What are the relevant factors that drive the water consumption and spray range ?.
- What are the optimal factors' settings that maximize the spray range and minimize the water consumption?

II. Methodology

Part I: Design of Experiment

Question 1. Propose a cost-efficient experimental design. Motivate your decision in statistical and practical terms.

In this experiment, we aim to find the relevant factors and determine the best combination of garden sprinklers to minimize water consumption and maximize the spray range. We first consider eight factors. If we want a full factorial design, we need a total of 2^8(256) runs to complete all combinations. However, the maximum number of tests for the entire experiment process allowed by the budget is 20 (N). In this case, a full factorial design will not work. Therefore, as a substitute, we choose the regular part analysis factor design. The smallest regular partial factor of 8 factors that we can use is a 16-run 2^(8-4) fractional factorial design. Therefore, we finally decided to perform a 2^(8-4) fractional factorial design.

Question 2. What is the performance of your design for studying the main effects of the factors only? Can your design estimate all two-factor interactions? Why or why not?

We can visualize the aliasing in this design using a color map on correlations.

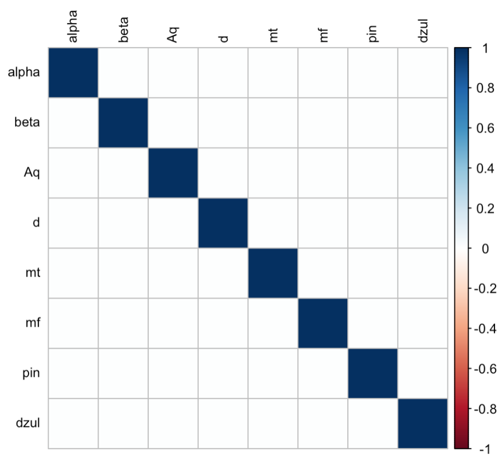

Figure 1

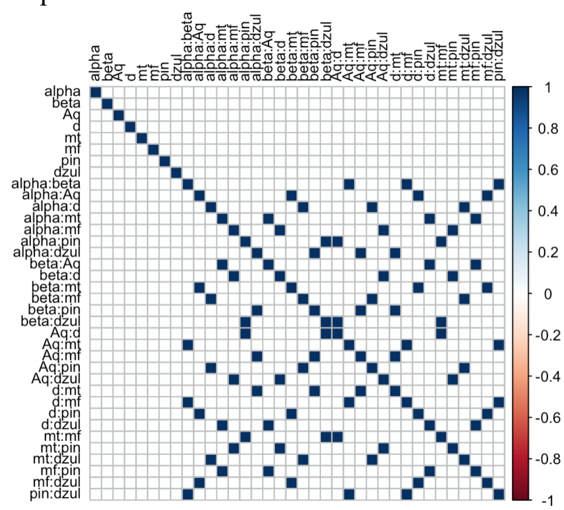

Figure 2

From Figure 1 and Figure 2, if we assume that the two-factor interactions are negligible, the aliasing structure of the design for the model involving only the main effects is excellent. There is no aliasing among the main effects.That is, the performance of the design for studying the main effects of the factors only is pretty good.

However, our design cannot estimate all two-factor interactions because there are some pairs of interaction that have a large correlation rate with other pairs, meaning they are aliasing with each other

If we consider a model including the main effects only, then we can study the variance inflation factors for the estimates of the coefficients in this model.

```
Variance the estimates when sigma^2 = 1
(Intercept)       alpha        beta          Aq           d          mt
     0.0625      0.0625      0.0625      0.0625      0.0625      0.0625
         mf         pin        dzul
     0.0625      0.0625      0.0625

 Variance inflation factors
(Intercept)       alpha        beta          Aq           d          mt
          1           1           1           1           1           1
         mf         pin        dzul
          1           1           1
```

Figure 3

From Figure 3, the result of the VIF confirms the conclusion we got from Figure 1 and Figure 2.

Question 3. The production engineers are concerned about having some failed tests in the experiment, given by sprinklers which cannot spray water. If you remove two randomly chosen test combinations, what is the performance of the resulting design?

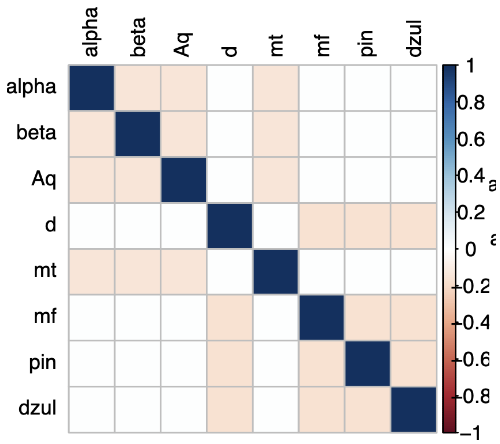

Figure 4

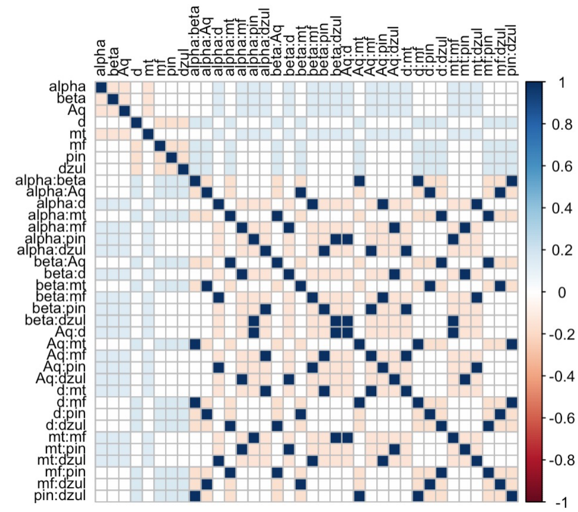

Figure 5

```
## Variance the estimates when sigma^2 = 1

## (Intercept)       alpha        beta          Aq           d          mt
##  0.08333333  0.07812500  0.07812500  0.07812500  0.08333333  0.07812500
##          mf         pin        dzul
##  0.08333333  0.08333333  0.08333333

##
##  Variance inflation factors

## (Intercept)       alpha        beta          Aq           d          mt
##     1.166667    1.093750    1.093750    1.093750    1.166667    1.093750
##          mf         pin        dzul
##     1.166667    1.166667    1.166667
```

Figure 6

The main effects are confounded with other main effects as well as interactions.The resulting design has worse performance than the previous one.

Question 4. The production engineers took an introductory course in experimental design. Using a commercial software, they came up with the experimental plan shown in Table 2. How does your full design compare with this one?

Alternative experimental design:

```
##    alpha beta aq  d mt mf pin dzul
## 1     -1    1 -1  1  1  0  -1   -1
## 2      0    0  0  0  0  0   0    0
## 3     -1    1  1  0 -1 -1   1   -1
## 4     -1   -1  1 -1  1  1   0   -1
## 5      1   -1 -1  0  1  1  -1    1
## 6     -1    0 -1 -1  1 -1   1    1
## 7      1    1 -1 -1  0  1   1   -1
## 8     -1   -1  1  1  0 -1  -1    1
## 9     -1   -1 -1  1 -1  1   1    0
## 10     1   -1  1 -1 -1  0   1    1
## 11     1    1  1 -1  1 -1  -1    0
## 12     1    1 -1  1 -1 -1   0    1
## 13    -1    1  0 -1 -1  1  -1    1
## 14     1   -1  0  1  1 -1   1   -1
## 15     0   -1 -1 -1 -1 -1  -1   -1
## 16     0    1  1  1  1  1   1    1
## 17     1    0  1  1 -1  1  -1   -1
```

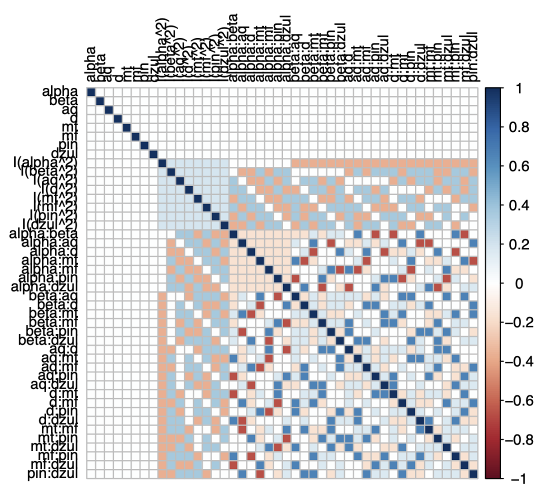

Figure 7 **Figure 8**

Full model with factor of 3 levels:

```
##       alpha beta aq  d mt mf pin dzul
## 19       -1   -1  1 -1 -1 -1  -1   -1
## 75        1   -1  1  1 -1 -1  -1   -1
## 169      -1    1 -1 -1  1 -1  -1   -1
## 225       1    1 -1  1  1 -1  -1   -1
## 235      -1   -1  1  1  1 -1  -1   -1
## 458       0    1  1  0  1  0  -1   -1
## 504       1    1  0 -1 -1  1  -1   -1
## 550      -1   -1  0  1 -1  1  -1   -1
## 678       1   -1 -1  0  1  1  -1   -1
## 1521      1    1 -1  1 -1 -1   1   -1
## 1582     -1    1  0  0  0 -1   1   -1
## 1644      1    0  1 -1  1 -1   1   -1
## 1837     -1   -1 -1  1  0  0   1   -1
## 1949      0    0 -1 -1 -1  1   1   -1
## 1992      1   -1  1  0 -1  1   1   -1
## 2023     -1    1  1  1 -1  1   1   -1
## 2125     -1   -1  1 -1  1  1   1   -1
## 2178      1    1  0  1  1  1   1   -1
## 2524     -1    0  0 -1  0  0  -1    0
## 2707     -1    1 -1  0 -1  1  -1    0
## 2943      1    1  1 -1 -1 -1   0    0
## 3700     -1   -1 -1  1 -1 -1   1    0
## 3810      1   -1 -1 -1  1 -1   1    0
## 3868     -1    1 -1  1  1 -1   1    0
## 4377      1   -1 -1 -1 -1 -1  -1    1
## 4453     -1    1  1  1 -1 -1  -1    1
## 4557      1   -1  1 -1  1 -1  -1    1
## 4835      0   -1 -1  1  1  0  -1    1
## 4885     -1    1  1 -1 -1  1  -1    1
## 4938      1    0  1  1 -1  1  -1    1
## 4942     -1   -1 -1 -1  0  1  -1    1
## 5031      1    1 -1 -1  1  1  -1    1
## 5101     -1    1  1  1  1  1  -1    1
## 5243      0    0 -1  1  0 -1   0    1
## 5419     -1   -1  1  1 -1  0   0    1
## 5763      1   -1  0 -1  1  1   0    1
## 5839     -1    1 -1 -1 -1 -1   1    1
## 5853      1   -1  1 -1 -1 -1   1    1
## 6013     -1   -1  1 -1  1 -1   1    1
## 6075      1    1  1  1  1 -1   1    1
## 6375      1   -1 -1  1 -1  1   1    1
## 6389      0    1  0  1 -1  1   1    1
## 6426      1    1  1 -1  0  1   1    1
## 6511     -1    0 -1  0  1  1   1    1
## 6554      0   -1  1  1  1  1   1    1
```

```
## 'data.frame':   45 obs. of  8 variables:
##  $ alpha: num  -1 1 -1 1 -1 0 1 -1 1 1 ...
##  $ beta : num  -1 -1 1 1 -1 1 1 -1 -1 1 ...
##  $ aq   : num  1 1 -1 -1 1 1 0 0 -1 -1 ...
##  $ d    : num  -1 1 -1 1 1 0 -1 1 0 1 ...
##  $ mt   : num  -1 -1 1 1 1 1 -1 -1 1 -1 ...
##  $ mf   : num  -1 -1 -1 -1 -1 0 1 1 1 -1 ...
##  $ pin  : num  -1 -1 -1 -1 -1 -1 -1 -1 -1 1 ...
##  $ dzul : num  -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...
```

Figure 9

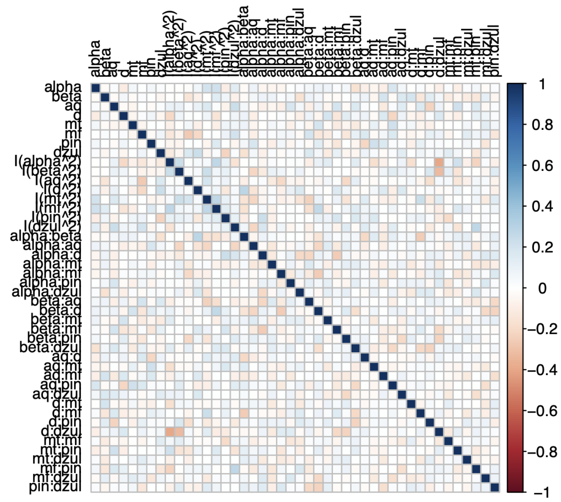

Figure 10

We generated the correlation plot for the alternative design with a coded table as Figure 7 shown. If we only consider the main effect and assume two-factor interactions are negligible, we can observe in the Figure 8 that the performance of the alternative experimental design with 17 runs is excellent since there are no aliasings between the main effects. For the full design, we also generated the full design of 3 levels with 45 runs(at least 45 runs = 1intercept + 8main effects + 8 quantitative factors + 28 two-factor interactions). As Figure 10 shows, the overall performance for the full model is good if we take the two-factor interactions into consideration since most of the color in the plot is pretty light which is approaching the value of 0. It means there are pretty weak correlations among the factors.

Part II: Analysis of the Results

Question 5. Collect data using your recommended design in Question 1. Conduct a detailed data analysis.

For spray range:

Full Model :

Data:

##	alpha	beta	Aq	d	mt	mf	pin	dzul	range
## 1	-1	-1	-1	-1	-1	-1	-1	-1	0.19236764
## 2	1	-1	-1	-1	1	1	1	-1	0.00000000
## 3	-1	1	-1	-1	1	1	-1	1	0.23450150
## 4	1	1	-1	-1	-1	-1	1	1	0.01280741
## 5	-1	-1	1	-1	1	-1	1	1	0.32894043
## 6	1	-1	1	-1	-1	1	-1	1	0.00000000
## 7	-1	1	1	-1	-1	1	1	-1	0.38104848
## 8	1	1	1	-1	1	-1	-1	-1	0.02997635
## 9	-1	-1	-1	1	-1	1	1	1	0.31772243
## 10	1	-1	-1	1	1	-1	-1	1	0.03356131
## 11	-1	1	-1	1	1	-1	1	-1	0.31974521
## 12	1	1	-1	1	-1	1	-1	-1	0.00000000
## 13	-1	-1	1	1	1	1	-1	-1	0.22686769
## 14	1	-1	1	1	-1	-1	1	-1	0.02132011
## 15	-1	1	1	1	-1	-1	-1	1	0.32358838
## 16	1	1	1	1	1	1	1	1	0.06262631

##	alpha	beta	Aq	d	mt	mf	pin	dzul	consumption
## 1	-1	-1	-1	-1	-1	-1	-1	-1	3.290304
## 2	1	-1	-1	-1	1	1	1	-1	4.619100
## 3	-1	1	-1	-1	1	1	-1	1	3.658242
## 4	1	1	-1	-1	-1	-1	1	1	4.787440
## 5	-1	-1	1	-1	1	-1	1	1	9.675282
## 6	1	-1	1	-1	-1	1	-1	1	6.610304
## 7	-1	1	1	-1	-1	1	1	-1	8.708282
## 8	1	1	1	-1	1	-1	-1	-1	6.025671
## 9	-1	-1	-1	1	-1	1	1	1	4.942020
## 10	1	-1	-1	1	1	-1	-1	1	3.447793
## 11	-1	1	-1	1	1	-1	1	-1	4.640643
## 12	1	1	-1	1	-1	1	-1	-1	3.335403
## 13	-1	-1	1	1	1	1	-1	-1	6.102466
## 14	1	-1	1	1	-1	-1	1	-1	8.633511
## 15	-1	1	1	1	-1	-1	-1	1	6.795141
## 16	1	1	1	1	1	1	1	1	9.780997

```
 alpha:beta = Aq:mt = d:mf = pin:dzul
 alpha:Aq = beta:mt = d:pin = mf:dzul
 alpha:d = beta:mf = Aq:pin = mt:dzul
 alpha:mt = beta:Aq = d:dzul = mf:pin
 alpha:mf = beta:d = Aq:dzul = mt:pin
 alpha:pin = beta:dzul = Aq:d = mt:mf
 alpha:dzul = beta:pin = Aq:mf = d:mt

Call:
lm.default(formula = range ~ (.)^2, data = data1)

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)  0.1553171         NA      NA       NA
alpha       -0.1352806         NA      NA       NA
beta         0.0152196         NA      NA       NA
Aq           0.0164789         NA      NA       NA
d            0.0078619         NA      NA       NA
mt          -0.0007897         NA      NA       NA
mf          -0.0024713         NA      NA       NA
pin          0.0252092         NA      NA       NA
dzul         0.0089014         NA      NA       NA
alpha:beta  -0.0089035         NA      NA       NA
alpha:Aq    -0.0080346         NA      NA       NA
alpha:d      0.0014786         NA      NA       NA
alpha:mt     0.0122943         NA      NA       NA
alpha:mf    -0.0019086         NA      NA       NA
alpha:pin   -0.0210572         NA      NA       NA
alpha:dzul  -0.0016891         NA      NA       NA
beta:Aq             NA         NA      NA       NA
beta:d              NA         NA      NA       NA
beta:mt             NA         NA      NA       NA
beta:mf             NA         NA      NA       NA
beta:pin            NA         NA      NA       NA
beta:dzul           NA         NA      NA       NA
Aq:d                NA         NA      NA       NA
Aq:mt               NA         NA      NA       NA
Aq:mf               NA         NA      NA       NA
Aq:pin              NA         NA      NA       NA
Aq:dzul             NA         NA      NA       NA
d:mt                NA         NA      NA       NA
d:mf                NA         NA      NA       NA
d:pin               NA         NA      NA       NA
d:dzul              NA         NA      NA       NA
mt:mf               NA         NA      NA       NA
mt:pin              NA         NA      NA       NA
mt:dzul             NA         NA      NA       NA
mf:pin              NA         NA      NA       NA
mf:dzul             NA         NA      NA       NA
pin:dzul            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

Estimated effects:
       alpha         beta           Aq            d           mt           mf
-0.270561284  0.030439257  0.032957783  0.015723701 -0.001579458 -0.004942555
         pin         dzul   alpha:beta     alpha:Aq      alpha:d     alpha:mt
 0.050418438  0.017802786 -0.017807093 -0.016069270  0.002957288  0.024588569
    alpha:mf    alpha:pin   alpha:dzul
-0.003817164 -0.042114396 -0.003378145
```

Figure 11

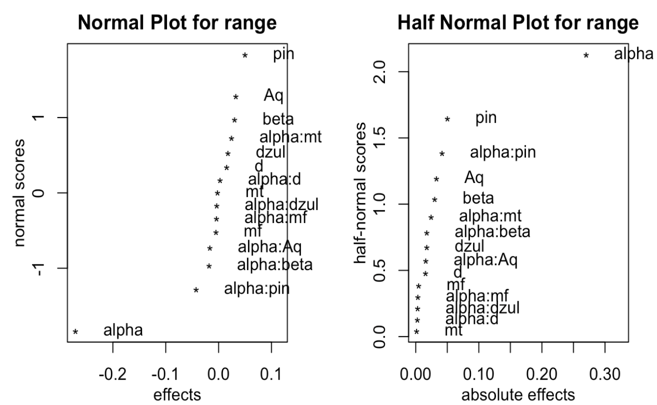

Figure 12

From Figure 11 and Figure 12, we could find that alpha, pin and alpha:pin are active(significant), thus we are going to refine our model using alpha,pin and alpha:pin.

Refine Model:

```
Call:
lm.default(formula = range ~ alpha + pin + alpha:pin, data = data1)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.05196 -0.01721 -0.01060  0.01499  0.07926 

Coefficients:
             Estimate Std. Error t value       Pr(>|t|)    
(Intercept)  0.155317   0.008919  17.414 0.000000000698 ***
alpha       -0.135281   0.008919 -15.167 0.000000003423 ***
pin          0.025209   0.008919   2.826         0.0153 *  
alpha:pin   -0.021057   0.008919  -2.361         0.0360 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03568 on 12 degrees of freedom
Multiple R-squared:  0.9531,	Adjusted R-squared:  0.9413 
F-statistic:  81.2 on 3 and 12 DF,  p-value: 0.00000003076
```

Figure 13

Model Adequacy Checking:

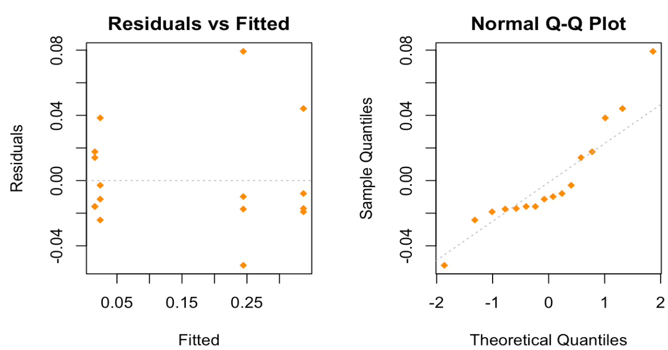

Figure 14

In Figure 14 Normal Q-Q plot, most of the points are close to the dashed line so the residual is generally distributed as normal. The normality assumption is satisfied.

In Figure 14 Residuals vs Fitted plot, there is no pattern(relationship) found(i.e. residuals are distributed randomly and independently around zero), so the constant-variance assumption is satisfied.

There is nothing unusual about the residual plots.We conclude that the assumptions for analysis of variance are satisfied.

For consumption:

```
Call:
lm.default(formula = consumption ~ (.)^2, data = data2)

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)  5.940788         NA      NA       NA
alpha       -0.035760         NA      NA       NA
beta         0.025690         NA      NA       NA
Aq           1.850669         NA      NA       NA
d            0.018959         NA      NA       NA
mt           0.052987         NA      NA       NA
mf           0.028814         NA      NA       NA
pin          1.032622         NA      NA       NA
dzul         0.271365         NA      NA       NA
alpha:beta   0.051660         NA      NA       NA
alpha:Aq     0.006924         NA      NA       NA
alpha:d      0.375439         NA      NA       NA
alpha:mt     0.010376         NA      NA       NA
alpha:mf     0.152609         NA      NA       NA
alpha:pin    0.017613         NA      NA       NA
alpha:dzul  -0.019759         NA      NA       NA
beta:Aq            NA         NA      NA       NA
beta:d             NA         NA      NA       NA
beta:mt            NA         NA      NA       NA
beta:mf            NA         NA      NA       NA
beta:pin           NA         NA      NA       NA
beta:dzul          NA         NA      NA       NA
Aq:d               NA         NA      NA       NA
Aq:mt              NA         NA      NA       NA
Aq:mf              NA         NA      NA       NA
Aq:pin             NA         NA      NA       NA
Aq:dzul            NA         NA      NA       NA
d:mt               NA         NA      NA       NA
d:mf               NA         NA      NA       NA
d:pin              NA         NA      NA       NA
d:dzul             NA         NA      NA       NA
mt:mf              NA         NA      NA       NA
mt:pin             NA         NA      NA       NA
mt:dzul            NA         NA      NA       NA
mf:pin             NA         NA      NA       NA
mf:dzul            NA         NA      NA       NA
pin:dzul           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


 alpha:beta = Aq:mt = d:mf = pin:dzul
 alpha:Aq = beta:mt = d:pin = mf:dzul
 alpha:d = beta:mf = Aq:pin = mt:dzul
 alpha:mt = beta:Aq = d:dzul = mf:pin
 alpha:mf = beta:d = Aq:dzul = mt:pin
 alpha:pin = beta:dzul = Aq:d = mt:mf
 alpha:dzul = beta:pin = Aq:mf = d:mt
Estimated effects:
      alpha        beta          Aq           d          mt          mf
-0.07152010  0.05137985  3.70133848  0.03791822  0.10597360  0.05762863
        pin        dzul  alpha:beta    alpha:Aq     alpha:d    alpha:mt
 2.06524385  0.54272968  0.10332074  0.01384806  0.75087868  0.02075188
   alpha:mf   alpha:pin  alpha:dzul
 0.30521870  0.03522520 -0.03951761
```

Figure 15

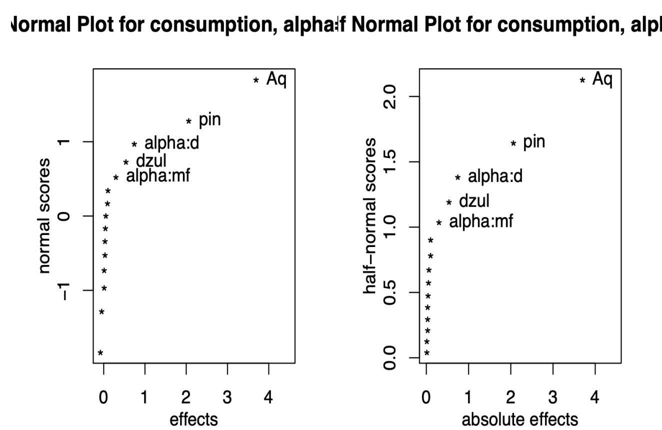

Figure 16

Figure 15 and Figure 16 above suggest Aq,pin, dzul,alpha:d,alpha:mf are active(significant).However, since alpha, d and mf are inactive(insignificant), by hierarchical rule, we are going refine our model using Aq,pin, dzul only.

Refine Model:

```
Call:
lm.default(formula = consumption ~ Aq + pin + dzul, data = data2)

Residuals:
     Min       1Q   Median       3Q      Max
-0.60666 -0.39373 -0.06497  0.37308  0.68555

Coefficients:
            Estimate Std. Error t value            Pr(>|t|)
(Intercept)   5.9408     0.1203  49.377 0.00000000000000312 ***
Aq            1.8507     0.1203  15.382 0.00000000291415406 ***
pin           1.0326     0.1203   8.583 0.00000181748860332 ***
dzul          0.2714     0.1203   2.255              0.0436 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4813 on 12 degrees of freedom
Multiple R-squared:  0.9633,    Adjusted R-squared:  0.9542
F-statistic: 105.1 on 3 and 12 DF,  p-value: 0.000000007003
```

Figure 17

Model Adequacy Checking:

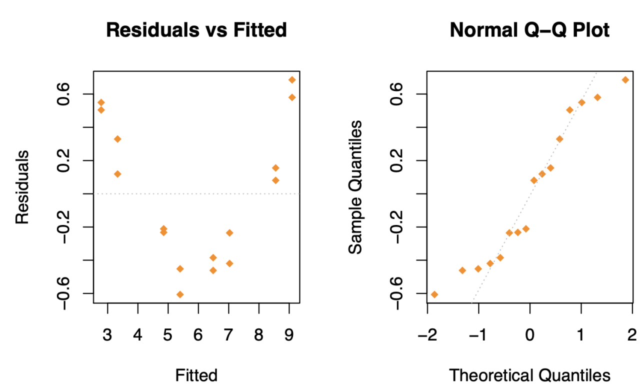

Figure 18

As we can see from Figure 18 Residuals vs Fitted plot, there seems to appear a bowl curve, so the constant-variance assumption might be violated.

To be safe, we will try to use log transformation to improve the model.

Log-transform the consumption:

Refine Model with log transformation :

```
Call:
lm.default(formula = consumption ~ Aq + pin + dzul, data = data2)

Residuals:
      Min        1Q    Median        3Q       Max
-0.043080 -0.012951  0.001646  0.014074  0.032772

Coefficients:
            Estimate Std. Error t value             Pr(>|t|)
(Intercept) 1.714913   0.006118 280.309 < 0.0000000000000002 ***
Aq          0.320024   0.006118  52.309  0.000000000000000157 ***
pin         0.172432   0.006118  28.185  0.00000000000246582 ***
dzul        0.041754   0.006118   6.825  0.00001836675894324 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02447 on 12 degrees of freedom
Multiple R-squared:  0.9967,	Adjusted R-squared:  0.9958
F-statistic:  1192 on 3 and 12 DF,  p-value: 0.000000000000000409
```

Figure 19

Model Adequacy Checking:

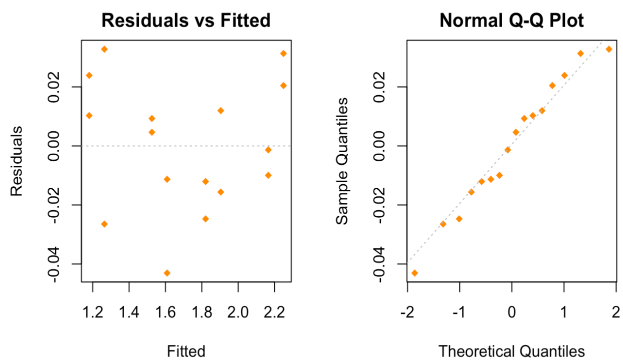

Figure 20

In Figure 20 Normal Q-Q plot, most of the points are close to the dashed line so the residual is generally distributed as normal. The normality assumption is satisfied.

In Figure 20 Residuals vs Fitted plot,there is no pattern(relationship) found(i.e. residuals are distributed randomly and independently around zero), so the constant-variance assumption is satisfied.

There is nothing unusual about the residual plots.We conclude that the assumptions for analysis of variance are satisfied. The model has been improved.

Question 6. What are the most influential factors?

```
Analysis of Variance Table

Response: range
          Df   Sum Sq  Mean Sq  F value         Pr(>F)
alpha      1 0.292814 0.292814 230.0486 0.000000003423 ***
pin        1 0.010168 0.010168   7.9885        0.01528 *
alpha:pin  1 0.007094 0.007094   5.5738        0.03599 *
Residuals 12 0.015274 0.001273
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 21

Define hypothesis test for main effects as below :
H_0: The main effect[Alpha/pin] is not statistically significant
H_1: The main effect[Alpha/pin] is statistically significant

Define hypothesis test for interaction effects as below :
H_0: The interaction effect[Alpha:pin] is not statistically significant
H_1: The interaction effect[Alpha:pin] is statistically significant

As per the ANOVA table :
The p-value for pin = $0.01528 < 0.1$
The p-value for alpha = $3.423e-09 < 0.001$
The p-value for pin:alpha = $0.03599 < 0.05$

Figure 22

According to Figure 21 and Figure 22, for the response value of range,we can confirm that the two-factor interaction pin:alpha , and factor alpha pin are significant.That is, alpha and pin are the most influential factors for the spray range model.

For consumption:

```
Analysis of Variance Table

Response: consumption
          Df  Sum Sq Mean Sq  F value              Pr(>F)
Aq         1 1.63865 1.63865 2736.247 0.000000000000001566 ***
pin        1 0.47573 0.47573  794.376 0.000000000002465821 ***
dzul       1 0.02789 0.02789   46.579 0.000018366758943236 ***
Residuals 12 0.00719 0.00060
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 23

Define hypothesis test for main effects as below :

H_0: The main effect[Aq/pin/dzul] is not statistically significant
H_1: The main effect[Aq/pin/dzul] is statistically significant

As per the ANOVA table :
The p-value for Aq = 1.566e-15 $<$ 0.001
The p-value for pin = 2.466e-12 $<$ 0.001
The p-value for dzul = 1.837e-05 $<$ 0.001

Figure 24

According to Figure 23 and Figure 24, for the response value of consumption, we can confirm that the main effects Aq pin dzul are significant. That is, Aq, pin, dzul are the most influential factors for the water consumption model.

Question 7. Recommend the settings of the factors that optimize the water consumption and spray range simultaneously.

For range :

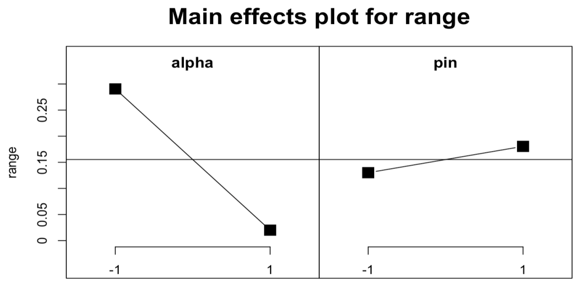

Figure 25

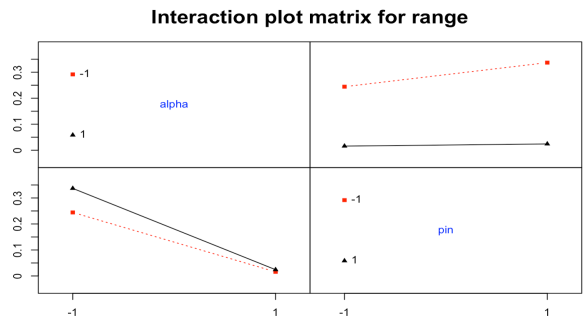

Figure 26

From Figure 25 and Figure 26, the main-effects plot and the interaction plot, we recommend alpha = -1 and pin = 1 as the settings of the factors that optimize(maximize) the spray range.

```
$par
[1] -1  1

$value
[1] -0.325864

$counts
function gradient
       8        8

$convergence
[1] 0

$message
[1] "CONVERGENCE: NORM OF PROJECTED GRADIENT <= PGTOL"
```

We can also use optim() to confirm our conclusion above.
From Figure 27, the results of optim() are the same as the main-effects plot and the interaction plot;we recommend alpha = -1 and pin = 1 as the settings of the factors that optimize(maximize) the spray range.

Figure 27

For Consumption :

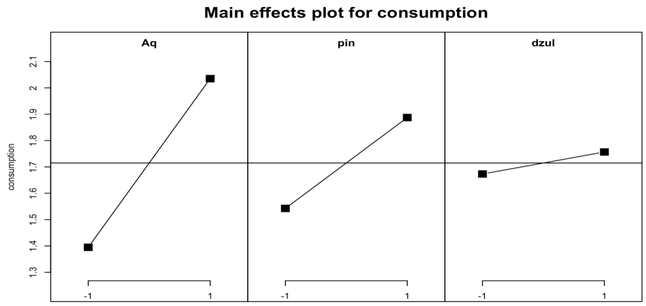

Figure 28

```
$par
[1] -1 -1 -1

$value
[1] 2.7861

$counts
function gradient
       3        3

$convergence
[1] 0

$message
[1] "CONVERGENCE: NORM OF PROJECTED GRADIENT <= PGTOL"
```

Figure 29

From Figure 28, the main-effects plot, we recommend Aq = -1,pin = -1, dzul = -1 as the settings of the factors that optimize(minimize) the water consumption.We can also use optim() to confirm our conclusion above.

From Figure 29, the results of optim() are the same as the main-effects plot and the interaction plot;we recommend Aq = -1,pin = -1, dzul = -1 as the settings of the factors that optimize(minimize) the water consumption.

Overall, as we can see, the optimal setting of the factor pin for the range and the consumption is conflicted. To achieve a better overall quality, we would choose the setting of pin = -1 since it has a stronger affect in the consumption rather than the range. Additionally, the p-value of pin is 2.466e-12 which is much smaller than 0.01. It is more statistically significant than the performance in the range.

Therefore, to minimize the water consumption and maximize the spray range simultaneously, the recommended setting of the factors should be Aq = -1 ; pin = -1 ; dzul = -1 ;alpha = -1.

Question 8. Conduct confirmation experiments using your recommended settings. Are your predictions accurate?

According to the confirmation experiments of section 8.2 from textbook, a simple confirmation experiment is to use the model equation to predict the response at a point of interest in the design space (this should not be one of the runs in the current design) and then actually find/run that treatment combination (perhaps several times), comparing the predicted and observed responses.

For range :

```
##
## Call:
## lm.default(formula = range ~ alpha + pin + alpha:pin, data = data1)
##
## Coefficients:
## (Intercept)        alpha          pin    alpha:pin
##     0.15532     -0.13528      0.02521     -0.02106
```

The final fitted equation for range is :

$$\hat{y} = 0.15532 - 0.13528x_1 + 0.02521x_7 - 0.02106x_1x_7$$

where $\hat{y}$ is the predicted response and x_1 and x_7 denote the coded level of factors alpha and pin respectively.

From Question 7, we know that the optimal setting of x_1 and x_7 for maximizing the spray range are -1 and -1.

Then $\hat{y} = 0.15532 - 0.13528(-1) + 0.02521(-1) - 0.02106(-1)(-1) = 0.24433$

The observed response under the condition that [alpha=pin=-1] are 0.1923676 0.2345015 0.2268677 0.3235884.

```
## [1] 0.1923676 0.2345015 0.2268677 0.3235884
```

We already conduct confirmation experiment for the model of "range"; comparing the predicted response 0.24433 with the observed response [0.1923676 0.2345015 0.2268677 0.3235884], there isn't big difference so we can say the prediction is accurate.

For consumption :

```
##
## Call:
## lm.default(formula = consumption ~ Aq + pin + dzul, data = data2)
##
## Coefficients:
## (Intercept)           Aq          pin         dzul
##     1.71491      0.32002      0.17243      0.04175
```

The final fitted equation for water consumption is :

$$\hat{y} = 1.71491 + 0.32002x_3 + 0.17243x_7 + 0.04175x_8$$

where $\hat{y}$ is the log predicted response and x_3, x_7 and x_8 denote the coded level of factors Aq,pin and dzul respectively.

From Question 7, we know that the optimal setting of x_3, x_7 and x_8 for minimizing the water consumption are -1,-1 and -1.

Then $\hat{y} = 1.71491 + 0.32002(-1) + 0.17243(-1) + 0.04175(-1) = 1.18071$

$e^{\hat{y}} = 3.256686$

The exponential observed response of consumption under the condition that (Aq=pin=dzul=-1) are 3.290304 3.335403

We already conduct confirmation experiment for the model of "consumption"; comparing the predicted value 3.256686 with the observed value [3.290304 3.335403], they are quite close so we can say the prediction is accurate.

III. Conclusions and Recommendations

To achieve the goal of minimizing the water consumption and the spray range of the garden sprinkle at its high quality, we propose the 2^(8-4) fractional factorial design in the first place. Because it could use only 16 runs to cover much information to analyze the data which is at a lower expense instead of 256 runs for all combinations among 8 factors. To compare the performance with main effects and the two-factor interactions , we generate the correlation plot of our proposed design. It turned out the performance of main effects is much better than it is in the two-factor interactions since there is no aliasing among the main effects and the value of variance inflation is 1 .

To analyze the data in a further way, we built the linear models based on our proposed design structure for the response value of the spray range and the water consumption respectively. It turned out that alpha, pin and alpha:pin are significant for the spray range. And the factors of Aq, pin, dzul, alpha:d and alpha:mf are significant for water consumption. To identify the influential factors, we rebuilt the linear model with those significant factors. By observing p-value in the Anova table and a series of hypothesis tests, it came out that the factors of alpha and pin are the most influential factors for the spray range model. And the factors of Aq, pin, dzul are the most influential factors for the water consumption model. To choose the settings that optimize the water consumption and the spray range simultaneously, we built the main-effects plot and interaction plot. For the spray range, alpha = -1 and pin = 1 should be the settings of the factors that optimize(maximize). For the water consumption, Aq = -1,pin = -1, dzul = -1 should be the optimal settings of the factors that optimize(minimize). However, the optimal setting of the factor pin is conflicted. To achieve a better overall quality, we would choose the setting of pin = -1 since it has a stronger effect in the consumption rather than the range. As we conducted more confirmation experiments, it turned out that our recommended setting worked out well with the accurate prediction. Finally, the recommended setting of the factors should be Aq = -1 ; pin = -1 ; dzul = -1 ;alpha = -1 with the 2^(8-4) fractional factorial design. In other words, we recommend setting the factors Nozzle profile = 2e-6 , Entrance pressure = 1 bar, Diameter flow line = 5mm , Vertical nozzle angle = 15 degree so that we can optimize the effect of the garden sprinkler.

IV. Appendix

```
library(corrplot)
library(readr)
library(FrF2)
library(AlgDesign)

# Part I
### Question 1
factor.names=list(alpha=c(0,90),
            beta=c(0,90),
            Aq=c(2e-06,4e-06),
            d=c(0.1,0.2),
            mt=c(0.01,0.02),
            mf=c(0.01,0.02),
            pin=c(1,2),
            dzul=c(5,10))
# Design to create txt
design <- FrF2(16, 8, factor.names = factor.names, randomize = FALSE)
design.info(design)$catlg.entry
design.df <- data.frame(design)
design.df
write.table(design.df,"design.txt",sep="\t",quote=FALSE,dec=".",row.names=FALSE)
# Read output from web
result <- read_csv("result.txt")
# Append output to design
data1 <- data.frame(desnum(design))
data1

### Question 2
```

```
generators(design)
design.info(design)$aliased
X.one <- model.matrix(~(alpha + beta + Aq + d + mt + mf + pin + dzul)^2-1, data.frame(data1))
contrast.vectors.correlations.one <- cor(X.one)
par(mfrow=c(1,2))
corrplot(contrast.vectors.correlations.one[1:8,1:8],
        type = "full", addgrid.col = "gray",
        tl.col = "black", tl.srt = 90, method = "color", tl.cex=0.8)
corrplot(contrast.vectors.correlations.one, type = "full", addgrid.col = "gray",
        tl.col = "black", tl.srt = 90, method = "color", tl.cex=0.8)


X.opt.me <- model.matrix(~(alpha+beta+Aq+d+mt+mf+pin+dzul), data.frame(data1))
XtX <- t(X.opt.me)%*%X.opt.me
inv.XtX <- solve(XtX)
var.eff <- diag(inv.XtX)
cat("Variance the estimates when sigma^2 = 1 \n")
print(var.eff)
cat("\n Variance inflation factors \n")
print(nrow(data1)*var.eff)



### Question 3
set.seed(888)
data1.removed <- data1[-sample(1:16,2,replace = F),]

X.one.removed <- model.matrix(~(alpha + beta + Aq + d + mt + mf + pin + dzul)^2-1,
data.frame(data1.removed))
contrast.vectors.correlations.one.removed <- cor(X.one.removed)
par(mfrow=c(1,2))
corrplot(contrast.vectors.correlations.one.removed[1:8,1:8],
        type = "full", addgrid.col = "gray",
        tl.col = "black", tl.srt = 90, method = "color", tl.cex=0.8)
corrplot(contrast.vectors.correlations.one.removed, type = "full", addgrid.col = "gray",
        tl.col = "black", tl.srt = 90, method = "color", tl.cex=0.8)

X.opt.me.removed <- model.matrix(~(alpha+beta+Aq+d+mt+mf+pin+dzul), data.frame(data1.removed))
XtX.removed <- t(X.opt.me.removed)%*%X.opt.me.removed
inv.XtX.removed <- solve(XtX.removed)
var.eff.removed <- diag(inv.XtX.removed)
cat("Variance the estimates when sigma^2 = 1 \n")
print(var.eff.removed)
cat("\n Variance inflation factors \n")
print(nrow(data1.removed)*var.eff.removed)


### Question 4

A <- c(-1,0,-1,-1,1,-1,1,-1,-1,1,1,1,-1,1,0,0,1)
B <- c(1,0,1,-1,-1,0,1,-1,-1,-1,1,1,1,-1,-1,1,0)
C <- c(-1,0,1,1,-1,-1,-1,1,-1,1,1,-1,0,0,-1,1,1)
D <- c(1,0,0,-1,0,-1,-1,1,1,-1,-1,1,-1,1,-1,1,1)
```

```
E <- c(1,0,-1,1,1,1,0,0,-1,-1,1,-1,-1,1,-1,1,-1)
F. <- c(0,0,-1,1,1,-1,1,-1,1,0,-1,-1,1,-1,-1,1,1)
G <- c(-1,0,1,0,-1,1,1,-1,1,1,-1,0,-1,1,-1,1,-1)
H <- c(-1,0,-1,-1,1,1,-1,1,0,1,0,1,1,-1,-1,1,-1)

design17.coded <- data.frame("alpha"=A, "beta"=B, "aq"=C, "d"=D, "mt"=E, "mf"=F., "pin"=G, "dzul"=H)

print(design17.coded)

# Visualize the aliasing in the design.
D.three.level <- design17.coded # Extract the design.
# Create the model matrix including main effects and two-factor interactions.
X.three.level <- model.matrix(~quad(alpha,beta,aq,d,mt,mf,pin,dzul)-1, data.frame(D.three.level))

# Create color map on pairwise correlations.
contrast.vectors.correlations.three <- cor(X.three.level)
corrplot(contrast.vectors.correlations.three, type = "full", addgrid.col = "gray",
        tl.col = "black", tl.srt = 90, method = "color", tl.cex=0.8)



candidate.set <- gen.factorial(levels = 3, nVars = 8,
                           varNames = c("alpha","beta","aq","d","mt","mf","pin","dzul"))

# 1 + 8 +8 +8(7)/2 =45runs
three.level.design <- optFederov(~quad(alpha,beta,aq,d,mt,mf,pin,dzul), candidate.set, nTrials = 45,
                              nRepeats = 100)
print.data.frame(three.level.design$design)

# Visualize the aliasing in the design.
D.three.level <- three.level.design$design # Extract the design.
str(D.three.level)
# Create the model matrix including main effects and two-factor interactions.
X.three.level <- model.matrix(~quad(alpha,beta,aq,d,mt,mf,pin,dzul)-1, data.frame(D.three.level))

# Create color map on pairwise correlations.
contrast.vectors.correlations.three <- cor(X.three.level)
corrplot(contrast.vectors.correlations.three, type = "full", addgrid.col = "gray",
        tl.col = "black", tl.srt = 90, method = "color", tl.cex=0.8)
```

Part II: Analysis of the Results

```
data1$range <- result$range
data1
data2 <- data.frame(desnum(design))
data2$consumption <- result$consumption
data2
```

Question 5. Collect data using your recommended design in Question 1. Conduct a detailed data analysis.

```
m1.full <- lm(range ~  (.)^2, data=data1)
aliases(m1.full)
summary(m1.full)
effects <- 2*(coef(m1.full)[-1])
cat("Estimated effects: \n")
print(effects[!is.na(effects)])
par(mfrow=c(1,2))
DanielPlot(m1.full, half = F,autolab = FALSE)
DanielPlot(m1.full, half = T,autolab=FALSE)

m1.simp <- lm(range ~ alpha + pin + alpha:pin,data=data1)
summary(m1.simp)

Residuals <- m1.simp$residuals
Fitted <- m1.simp$fitted.values
par(mfrow = c(1,2),oma = c(0, 0, 2, 0))
plot(Fitted,Residuals,main="Residuals vs Fitted",pch =18,col="dark orange")
abline(h = mean(Residuals),lty=3,col="grey")
qqnorm(Residuals,pch =18,col="dark orange")
qqline(Residuals,lty=3,col="grey")


m2.full <- lm(consumption ~  (.)^2, data=data2)
summary(m2.full)
aliases(m2.full)
effects <- 2*(coef(m2.full)[-1])
cat("Estimated effects: \n")
print(effects[!is.na(effects)])
par(mfrow=c(1,2))
DanielPlot(m2.full, half = F)
DanielPlot(m2.full, half = T)

m2.simp <- lm(consumption ~ Aq + pin + dzul,data=data2)
summary(m2.simp)

Residuals <- m2.simp$residuals
Fitted <- m2.simp$fitted.values
par(mfrow = c(1,2),oma = c(0, 0, 2, 0))
plot(Fitted,Residuals,main="Residuals vs Fitted",pch =18,col="dark orange")
abline(h = mean(Residuals),lty=3,col="grey")
qqnorm(Residuals,pch =18,col="dark orange")
qqline(Residuals,lty=3,col="grey")

data2$consumption <- log(data2$consumption)
data2

m2.simp.log <- lm(consumption ~ Aq + pin + dzul,data=data2)
summary(m2.simp.log)


Residuals <- m2.simp.log$residuals
Fitted <- m2.simp.log$fitted.values
par(mfrow = c(1,2),oma = c(0, 0, 2, 0))
```

```
plot(Fitted,Residuals,main="Residuals vs Fitted",pch =18,col="dark orange")
abline(h = mean(Residuals),lty=3,col="grey")
qqnorm(Residuals,pch =18,col="dark orange")
qqline(Residuals,lty=3,col="grey")


### Question 6. What are the most influential factors?

anova(m1.simp)

anova(m2.simp.log)

### Question 7. Recommend the settings of the factors that optimize the water consumption and spray
range simultaneously.

MEPlot(m1.simp)
IAPlot(m1.simp)


obj_func <- function(x){
  pred.y <- 0.144317 - 0.135281*x[1]+ 0.025209*x[2] - 0.021057*x[1]*x[2]
  return(-1*pred.y)
}

optim(par = c(0, 0), fn = obj_func, lower = -1, upper = 1, method = "L-BFGS-B")

MEPlot(m2.simp.log)

obj_func <- function(x){
  pred.y <- 5.9408 + 1.8507*x[1]+ 1.0326*x[2] + 0.2714*x[3]
  return(1*pred.y)
}

optim(par = c(0, 0, 0), fn = obj_func, lower = -1, upper = 1, method = "L-BFGS-B")

### Question 8. Conduct confirmation experiments using your recommended settings. Are your predictions
accurate?
m1.simp
data1[(data1$alpha==-1)&(data2$pin == -1),9]
m2.simp.log
#exp(1.18071)
#exp(data2[(data2$Aq==-1)&(data2$pin == -1)&(data2$dzul==-1),9])
```