4CO3Ass5

Abdullah

05/08/2021

Problem 1

1a)

```
Cata <- c(-1, 1, -1, 1, 1, -1, 1, -1, 1, -1, -1, 1, 1, 1, -1, -1)
Temp <- c(1, -1, -1, 1, -1, 1, 1, -1, -1, 1, 1, -1, 1, 1, -1, -1)
Pres <- c(1, -1, 1, -1, -1, 1, 1, -1, 1, -1, -1, 1, 1, -1, 1, -1)
Conc <- c(-1, -1, 1, -1, 1, 1, 1, -1, -1, -1, 1, 1, -1, 1, -1, 1)
y    <- c(83, 57, 55, 78, 46, 81, 74, 67, 57, 86, 85, 47, 76, 79, 64, 57)

mod1 <- lm(y~Cata*Temp*Pres*Conc)
summary(mod1)
```

```
## 
## Call:
## lm(formula = y ~ Cata * Temp * Pres * Conc)
## 
## Residuals:
## ALL 16 residuals are 0: no residual degrees of freedom!
## 
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         6.825e+01        NaN     NaN      NaN
## Cata               -4.000e+00        NaN     NaN      NaN
## Temp                1.200e+01        NaN     NaN      NaN
## Pres               -1.125e+00        NaN     NaN      NaN
## Conc               -2.750e+00        NaN     NaN      NaN
## Cata:Temp           5.000e-01        NaN     NaN      NaN
## Cata:Pres           3.750e-01        NaN     NaN      NaN
## Temp:Pres          -6.250e-01        NaN     NaN      NaN
## Cata:Conc           4.163e-17        NaN     NaN      NaN
## Temp:Conc           2.250e+00        NaN     NaN      NaN
## Pres:Conc          -1.250e-01        NaN     NaN      NaN
## Cata:Temp:Pres     -3.750e-01        NaN     NaN      NaN
## Cata:Temp:Conc      2.500e-01        NaN     NaN      NaN
## Cata:Pres:Conc     -1.250e-01        NaN     NaN      NaN
## Temp:Pres:Conc     -3.750e-01        NaN     NaN      NaN
## Cata:Temp:Pres:Conc -1.250e-01        NaN     NaN      NaN
## 
## 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
```

Therefore the equation for the linear model of the full factorial design with interactions is y = 68.25 - 4.00$Xconc$ + 12.00Xtemp - 1.125$Xpres$ - 2.75Xconc + 0.5$Xcat$Xtemp + 0.375$Xcat$Xpres - 0.625$Xtemp$Xpres + 4.16$10\hat{\ }$-17Xcat$Xconc$ + 2.25Xtemp$Xconc$ - 0.125Xpres$Xconc$ - 0.375Xcat$XTemp$XPres + 0.25$Xcat$Xtemp$Xconc$ - 0.125Xcat$Xpres$Xconc -0.375$Xtemp$XPres$Xconc$ - 0.125Xcat$Xtemp$Xpres*Xconc.

Therefore the parameters are the relationship that specific variable term has with the overall effect on the output of the model. In this model a postive coefficient the indicates that there is a postive correlation betweent the variable/interaction term it multiple with on the result on the outcome. A negative correlation indicates a negative correlation between those terms. Looking further the magnitude of the coefficient essentially indicates change in the outcome when responding to a one unit change in the input variable/interaction term.
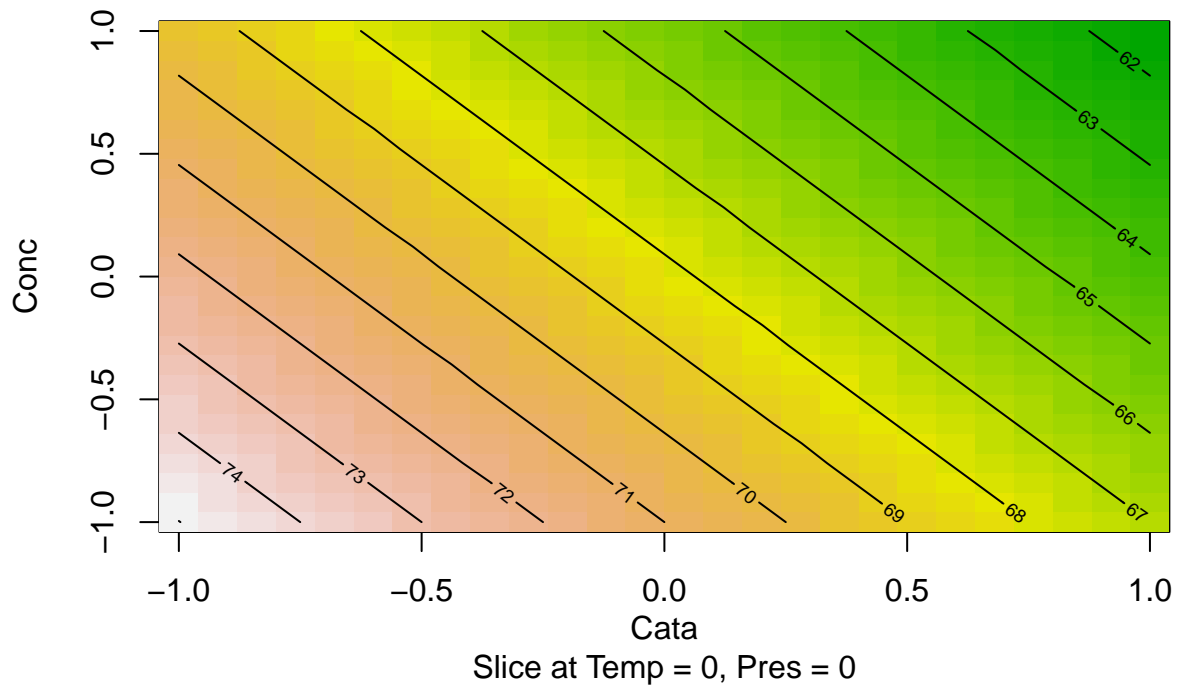
1b)

```
library("pid")
```

```
## Registered S3 method overwritten by 'DoE.base':
##   method          from
##   factorize.factor conf.design
```

```
library("rsm")
plt2 <- persp(mod1, Cata~Conc, col = "orange")
```
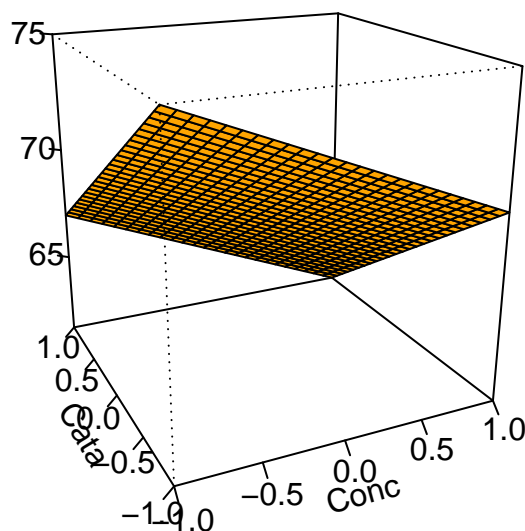

Slice at Temp = 0, Pres = 0

```
plt3 <- contour(mod1, ~Cata+Conc, image = TRUE)
```


Slice at Temp = 0, Pres = 0

The perspective plot above shows that there is no significant interaction between the terms, of temperature and pressure. This was determined by looking at the plane which is flat. The contour plot signifies a similar conclusion. Since the contour lines are spaced evenly across the whole plot that means that there indeed is not a significant relationship betwenn temperature vs pressure.
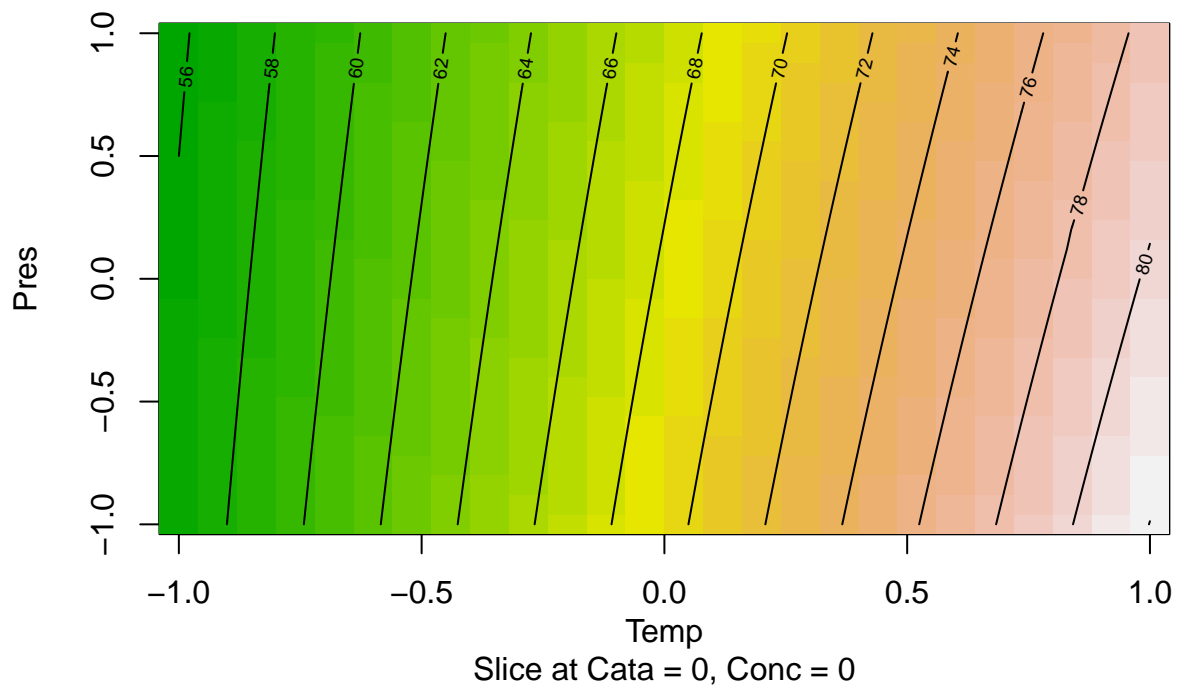
```
library("pid")
library("rsm")

plt1 <- persp(mod1, Temp~Pres, col = "orange")
```
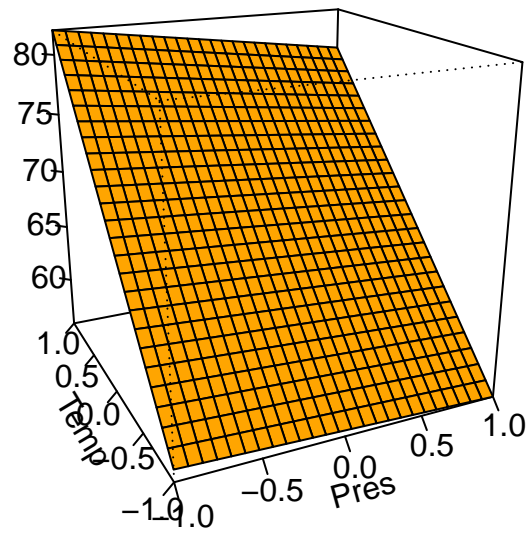
Slice at Cata = 0, Conc = 0

```
plt4 <- contour(mod1, ~Temp+Pres, image = TRUE)
```

Slice at Cata = 0, Conc = 0

The relationship of catalyst vs concentration has a similar interaction to that of temperature and pressure. Since the contour plot has lines spread evenly that is a good indication that catalyst and concentration do not have a significant interaction with each other. This is further confirmed by consulting the perspective plot which has a flat non-curved surface.

1c)

```
plt5 <- paretoPlot(mod1)
```
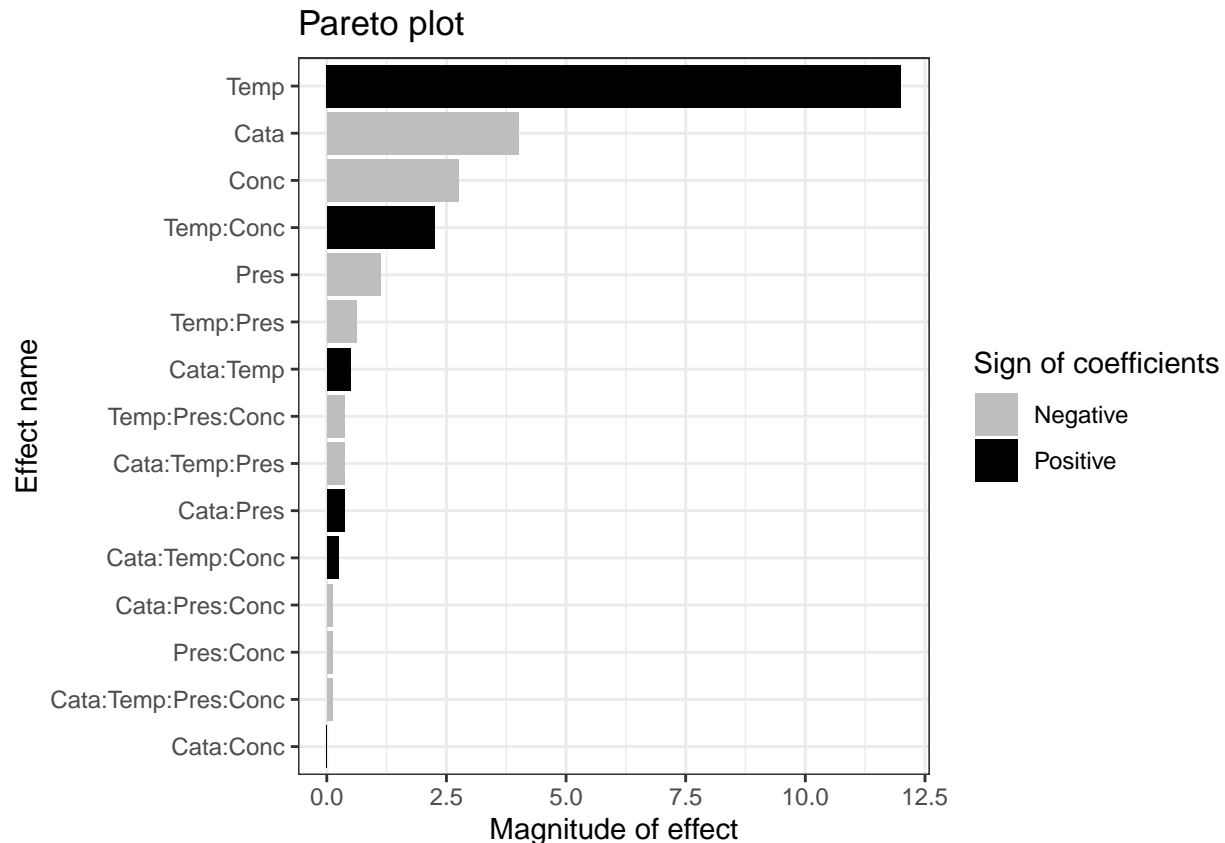
The 3 most significant factors on the conversion yield after consulting the pareto plot shows to be the temperature followed by the catalyst in second and the concentration in third.

Problem 2)

2a)

Converting data into a standard table format allows easier recognition of high and low factors for each experiment. It makes it easier for one to show half or quarter fractional designs. It also make the interpretation of quarter, full and even half fractional designs easier, as they can simply count the number of experiments. It allows a really easy method to determine what factors correspond to other factors when there needs to be aliasing in the design.

2b)

The nomenclature for this type of fractional design is called a quarter fractional design. Or in this case more like a quarter fraction design. The standard order table indicates that there should be a total of $2^5 = 32$ experiments to have a full fractional design and the current standard order table only has 8 experiments hence the name quarter fraction design. In the case a full factorial was to be done an addition of 24 more experiments must be done on top of the current 8 experiments. There are a total of 24 experiments "missing", the values for the factors X4 and X5 are determined based off of X1, X2 and X3. In this experiment X4 is determined to be X1*X2 and X5 is determined to be X1*X3. If the cube plot was considered for this question the points cover the diagonal pairs for the experiments to ensure a mix of effects of all variables within the model. The reason for half fractional and quarter fractional designs is to simply save cost. Sometimes a test

can take thousands of dollars, and it can quite waste to invest a huge amount of the budget on the inital model which is based simply on hypothesis. A quarter fraction and half fraction design essentially has a trade off, less accuracy of the data but less amount of experiments need to be done to create the model. A half/quarter fractional design is ideal for the first model.

2c)

Factor 4(X4) and Factor5(X5) in this case were determined using a tradeoff table. Since there are less than 32 experiments trade offs need to be made on what scenarios can and cannot be added to the list of trials. To ensure maximum accuracy in the model for the terms that are important the corresponding diagnals are chosen. Factor 4 is chosen by X1*X2 and factor 5 is chosen by X1*X3. Aliasing in this case is the process of assuming data. In this case for example X4 is aliased for X1*X2.Since there are only 8 experiments, the effects of the aliasing is that the effects are now confounded by each other. Since a reduced set of experiments was done, that aliasing results in confounding of data meaning for example we cant be sure for some terms if X1 caused the change or X4. By having fewer experiments than required and assuming the rest of the data we have lost that resultion in out model to make some of these predictions. Aliasing will effect most of the larger scale interaction term's accuracy. The resolution for this tryp of experiment is resolution III which means that the main effects will be confounded with 2 factor interactions.

2d)

```
X1 <- c( -1, +1, -1, +1, -1, +1, -1, +1)
X2 <- c(-1, -1, 1, 1, -1, -1, 1, 1)
X3 <- c(-1, -1, -1, -1, 1, 1, 1, 1)
X4 <- c(1, -1, -1, 1, 1, -1, -1, 1)
X5 <- c(1, -1, 1 ,-1, -1, 1, -1, 1)
Y <- c(44, 53, 70, 93, 66, 55, 54, 82)

Quarter <- lm(Y~X1*X2*X3*X4*X5)
#Quarter <- lm(Y~X1+X2+X3+X4+X5)
summary(Quarter)
```

```
##
## Call:
## lm(formula = Y ~ X1 * X2 * X3 * X4 * X5)
##
## Residuals:
## ALL 8 residuals are 0: no residual degrees of freedom!
##
## Coefficients: (24 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    64.625        NaN     NaN      NaN
## X1              6.125        NaN     NaN      NaN
## X2             10.125        NaN     NaN      NaN
## X3             -0.375        NaN     NaN      NaN
## X4              6.625        NaN     NaN      NaN
## X5             -1.875        NaN     NaN      NaN
## X1:X2              NA         NA      NA       NA
## X1:X3              NA         NA      NA       NA
## X2:X3          -6.375        NaN     NaN      NaN
## X1:X4              NA         NA      NA       NA
```

```
## X2:X4                   NA        NA       NA       NA
## X3:X4                3.125       NaN      NaN      NaN
## X1:X5                   NA        NA       NA       NA
## X2:X5                   NA        NA       NA       NA
## X3:X5                   NA        NA       NA       NA
## X4:X5                   NA        NA       NA       NA
## X1:X2:X3               NA        NA       NA       NA
## X1:X2:X4               NA        NA       NA       NA
## X1:X3:X4               NA        NA       NA       NA
## X2:X3:X4               NA        NA       NA       NA
## X1:X2:X5               NA        NA       NA       NA
## X1:X3:X5               NA        NA       NA       NA
## X2:X3:X5               NA        NA       NA       NA
## X1:X4:X5               NA        NA       NA       NA
## X2:X4:X5               NA        NA       NA       NA
## X3:X4:X5               NA        NA       NA       NA
## X1:X2:X3:X4            NA        NA       NA       NA
## X1:X2:X3:X5            NA        NA       NA       NA
## X1:X2:X4:X5            NA        NA       NA       NA
## X1:X3:X4:X5            NA        NA       NA       NA
## X2:X3:X4:X5            NA        NA       NA       NA
## X1:X2:X3:X4:X5         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
```
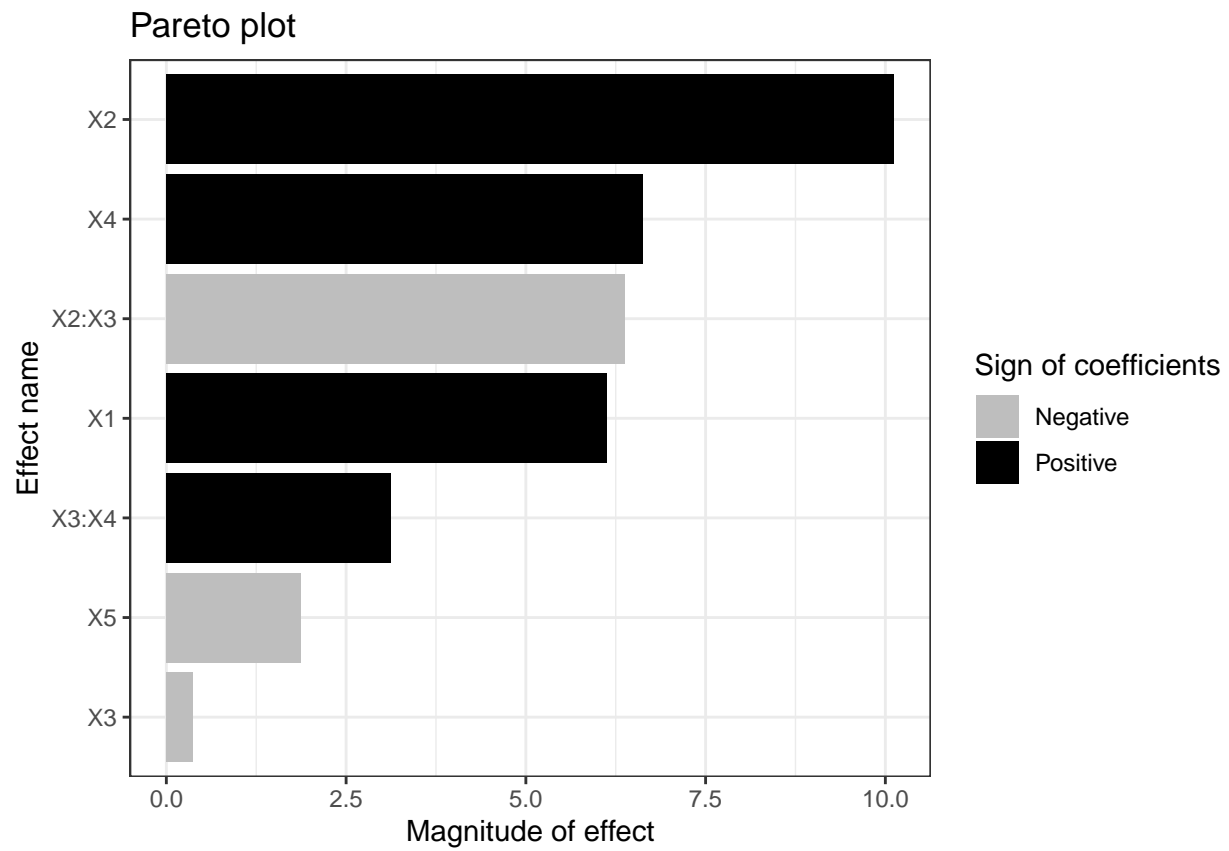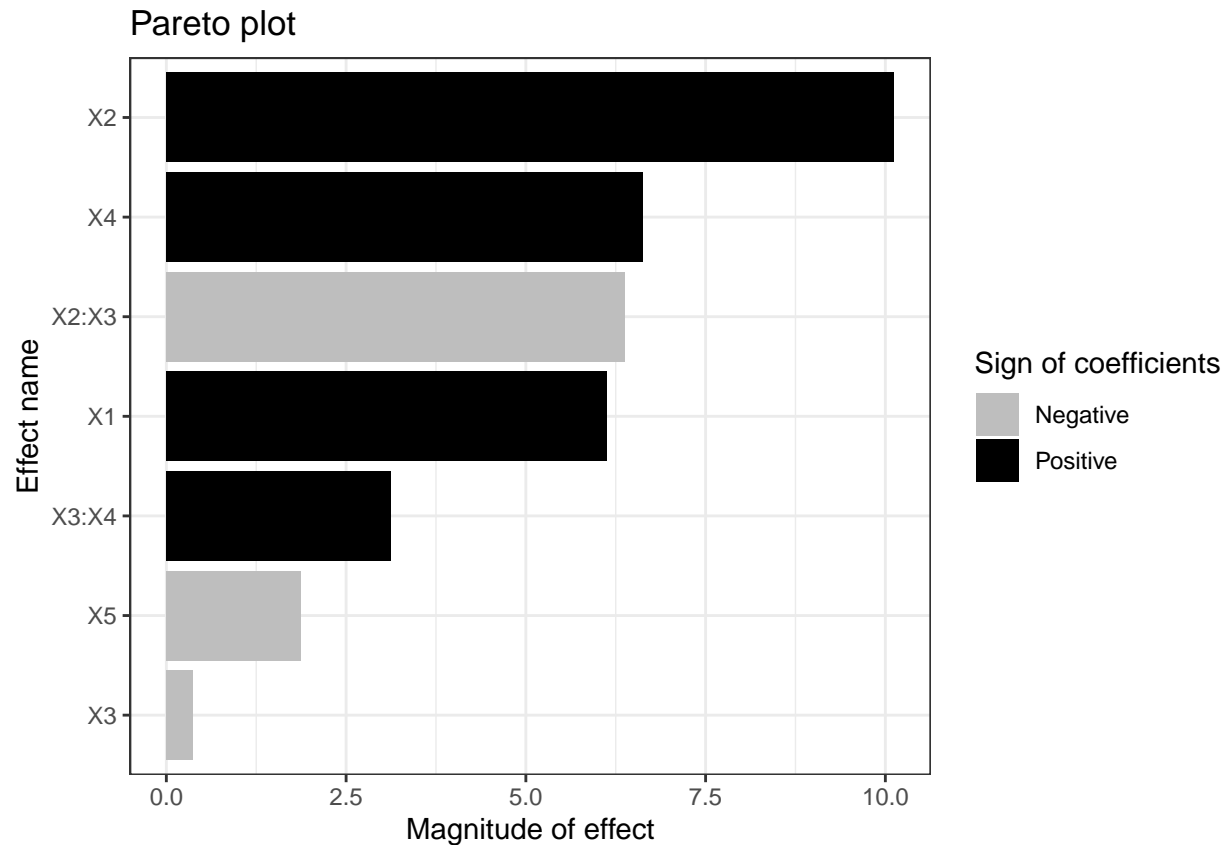
```
paretoPlot(Quarter)
```

Pareto plot

Pareto plot

The 3 most significant factors in this case are going to be X2, X4 and X1, but the interaction term for X2X3 comes as the third most significant. When consulting the pareto plot it is evident that the magnitude of effect for X2, X3 and X1 seem to be the highest out of all the coefficients. This essentially entails that the factors X2, X4, and X1 have the largest effects on the response variable y for a single unit change in their corresponding independent variables.

2e)

```
QuarterRebuilt <- lm(Y~X2*X4*X1)
summary(QuarterRebuilt)
```

```
##
## Call:
## lm(formula = Y ~ X2 * X4 * X1)
##
## Residuals:
##     1     2     3     4     5     6     7     8
## -11.0  -1.0   8.0   5.5  11.0   1.0  -8.0  -5.5
##
## Coefficients: (4 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   64.625      3.676  17.579 6.15e-05 ***
## X2            10.125      3.676   2.754   0.0512 .
## X4             6.625      3.676   1.802   0.1459
```

```
## X1                 6.125        3.676    1.666     0.1710
## X2:X4              NA           NA       NA        NA
## X2:X1              NA           NA       NA        NA
## X4:X1              NA           NA       NA        NA
## X2:X4:X1           NA           NA       NA        NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.4 on 4 degrees of freedom
## Multiple R-squared:  0.7728, Adjusted R-squared:  0.6025
## F-statistic: 4.536 on 3 and 4 DF,  p-value: 0.08909
```

```
paretoPlot(QuarterRebuilt)
```


Pareto plot

Pareto plot
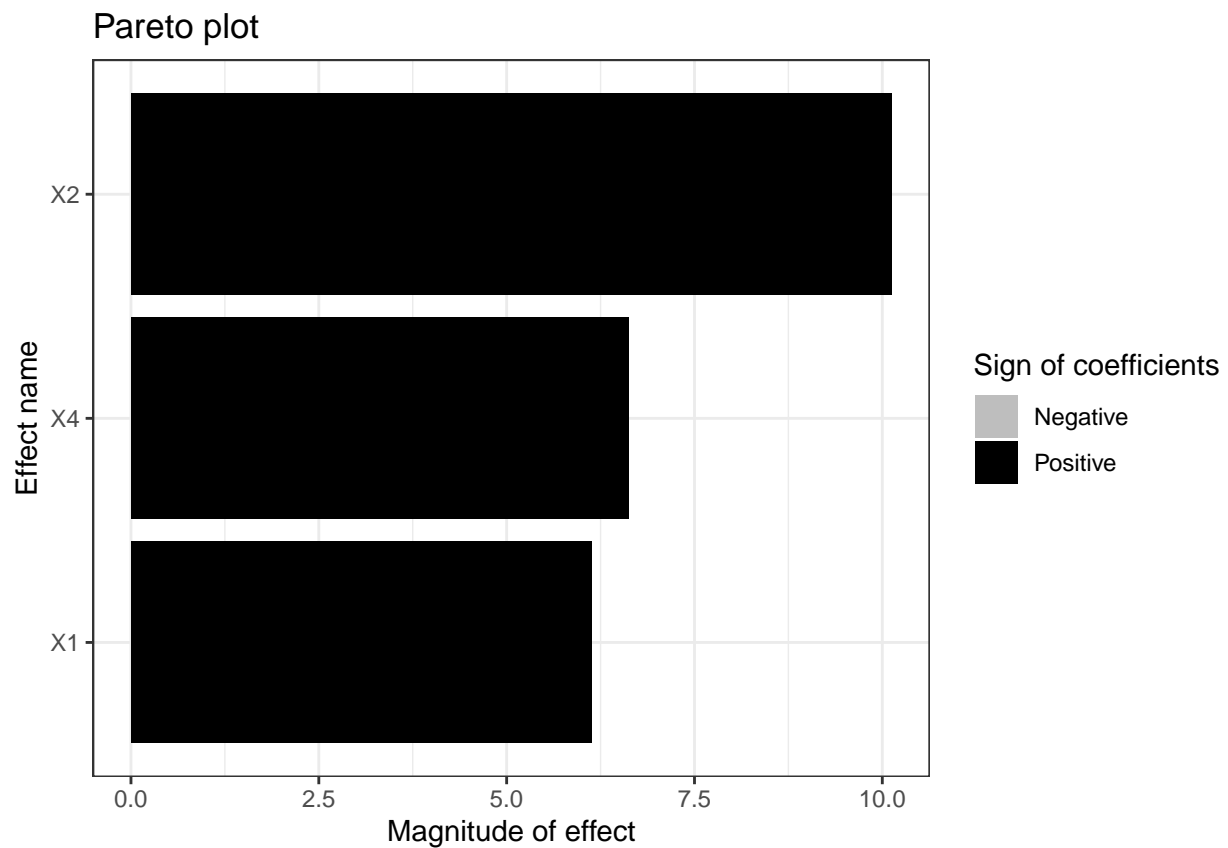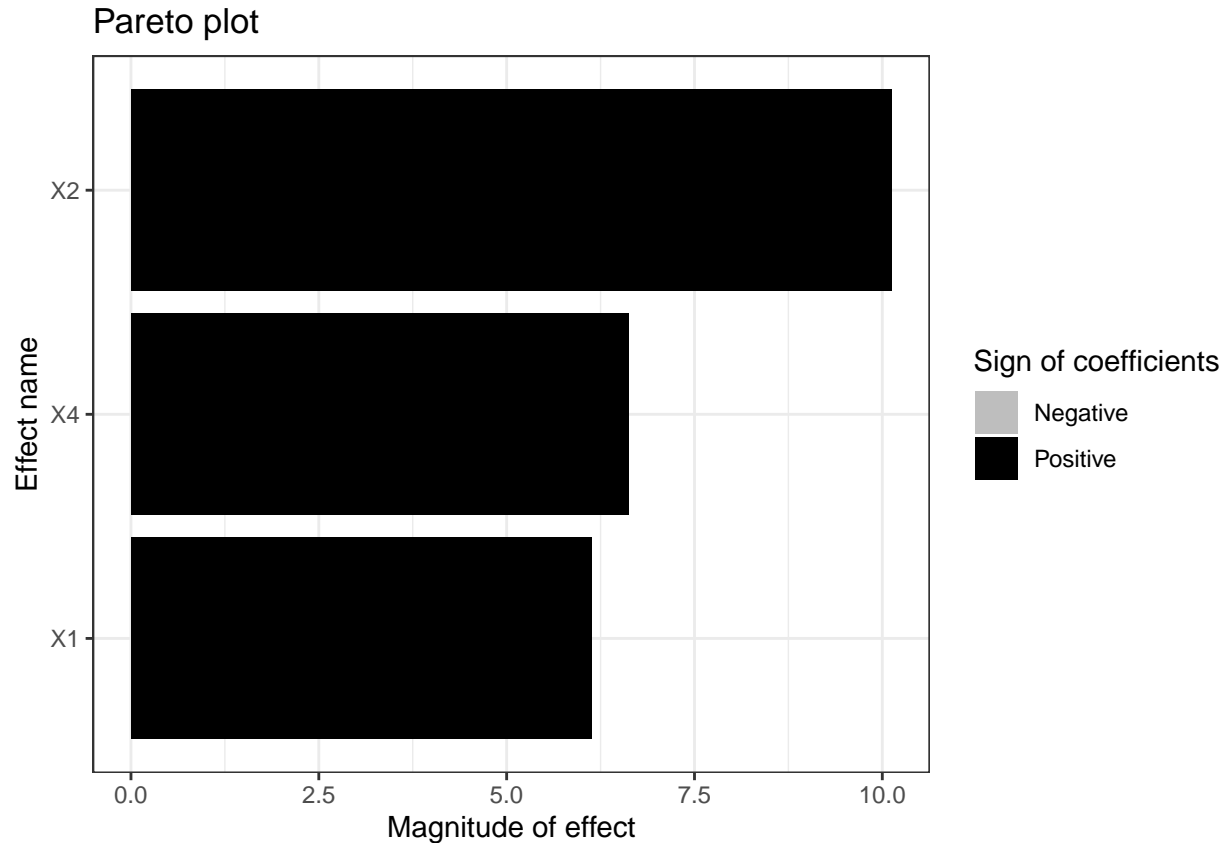
The model has been refit with the 3 important factors. The 3 important factors that were determined were X2, X4, and X1. The plot that is obtained is very similar to the previous plot, except for the lack of interaction terms. The particular ordering of the factors from one another have remained the same from the previous model and the new refitted model. When looking at the new values within the model it is clear that they are not significant values more different than standard error. None of the values achieve significance below 0.05 meaning they all essentially fail to achieve significance on their corresponding T - test's.

2f)

This is most likely a nuisance variable. Nuisance variables are usually an unwanted variable that is correlated with the hypothesized independent factors within an experimental study. In this case varying the time the study is done changes various environmental conditions, such as temperature and humidity. Also certain chemical reaction require UV light as a catalyst and the lack of it can affect the response variable. Now that we know there is a nuisance variable in this experiment there are actually many things that can be done to combat its effects on the final model. Blocking is by far the easiest and best option, in this case since the variable is known it can be added to the experimental design as another independent variable. ANCOVA/partial correlation can also be used to hold the time experiment is performed as constant. Randomization is another option which is usually used when the nuisance variable is not know which in this case may be possible, as we do not know what other confounders time of experiment and their effects on the response. This technique can be used along with blocking, to reduce the effects of remaining known variables.

Problem 3

3a)

The number of experiment that are required for a full factorial can be defined as 2^k where 2 is the amount of levels per factors and k is the number of factors. In this case to run a full factorial it would require 2^5 = 32 experiments.

3b)

The following fractional experiment has a total of 5 factors but only 8 experiments. The following equation can be used to solve for the type of design.

2^(5-p) = 8 –> p = 2.

Therefore since p is 2, the following design is a quarter fractional design. Therefore since this there are 2^(5-2) = 8 experiments, therefore the following design is resolution 3 design.

3c)

After consulting the tradeoff table the generators used in a resolution 3 fractional design is +-D = AB & +-E = AC. The following relationship has 2^p words. Since p is 2, there are a total of 4 words.

1 –> D = AB, E = AC

2 –> I = ACE, I = ABD

3 –> I = ABD = ACE = (ABD)(ACE) = BCDE

Therefore the 4 words that are generated are ABD, ACE, BCDE and I

3d)

The 2 factor interaction of BC will be confounded with A. The reason being since this is a resolution 3 quarter factorial design, the main effects are aliased with 2 factor interactions. In this case the 2 factor interaction being confounded with A.

3e)

```
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)
y <- c(83, 72, 67, 80, 71, 81, 67, 82)

model.1 <- lm(y~A+B+C+(A*B)+(A*C))

summary(model.1)


##
## Call:
## lm(formula = y ~ A + B + C + (A * B) + (A * C))
##
```

13

```
## Residuals:
##       1       2       3       4       5       6       7       8
##    3.00   -1.75   -3.00    1.75   -3.00    1.75    3.00   -1.75
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      75.375      1.737  43.405  0.00053 ***
## A                 3.375      1.737   1.944  0.19141
## B                -1.375      1.737  -0.792  0.51147
## C                -0.125      1.737  -0.072  0.94917
## A:B               3.625      1.737   2.087  0.17210
## A:C               2.875      1.737   1.656  0.23964
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.912 on 2 degrees of freedom
## Multiple R-squared:  0.8519, Adjusted R-squared:  0.4818
## F-statistic: 2.302 on 5 and 2 DF,  p-value: 0.3301
```

```
library("pid")
```

```
paretoPlot(model.1)
```
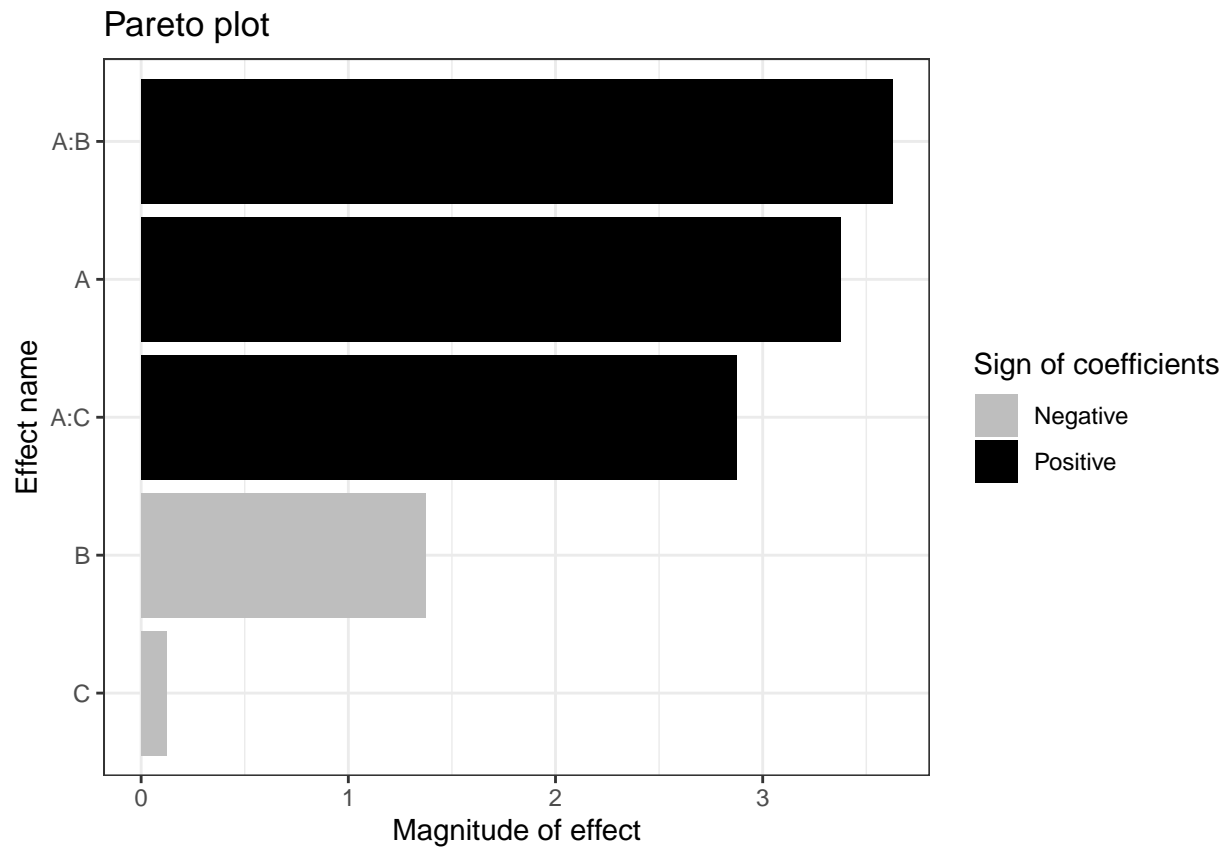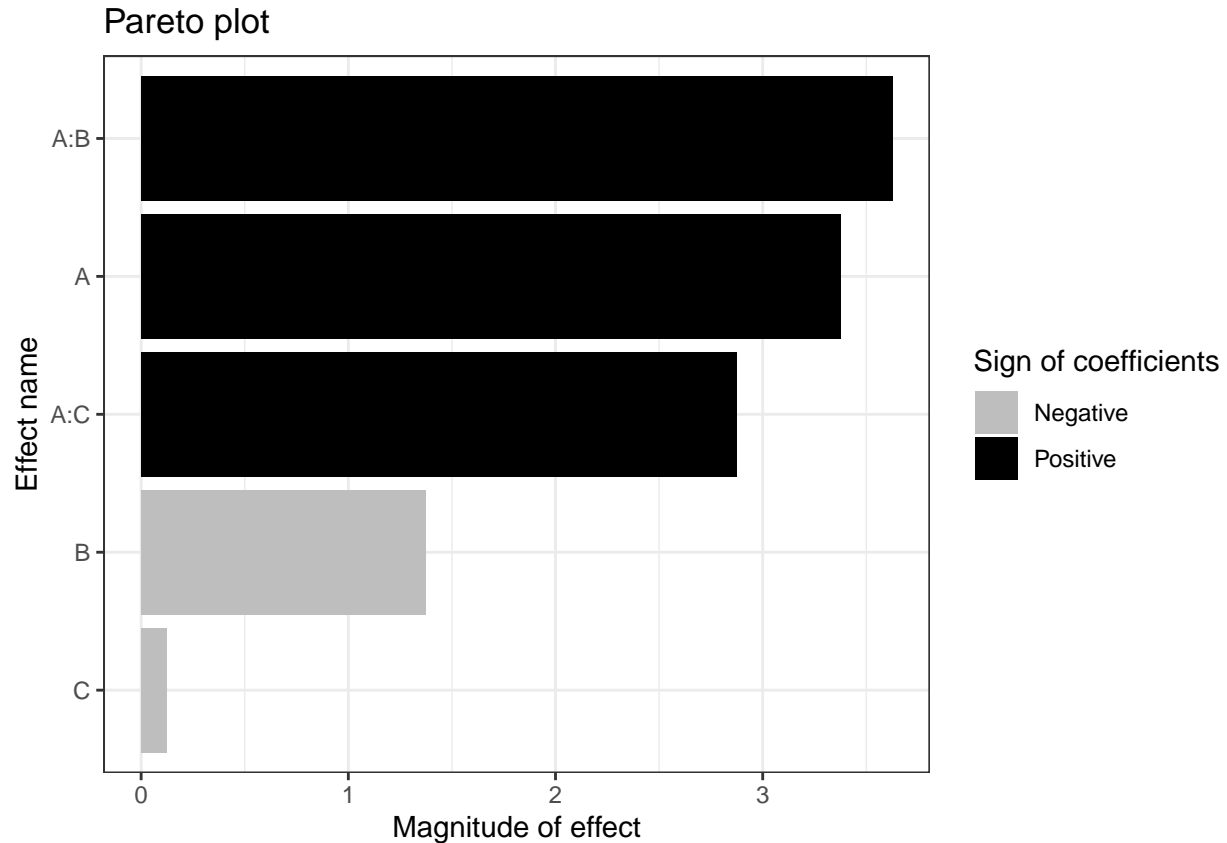
Pareto plot

The 2 factors with the greatest effect on the response variable are factor a with the greatest effect followed by factor B. If the output from R shown above is considered it is evident that factor A has the greatest magnitude. The pareto plot ranks A as first and B behind it.

3f)

3g)

Considering the factors A and B, the direction which we should move next shoudl be towards the upper left of the contour plot. Since this is a minimization probelm the goal is to move towards a region with lower response values. The upper left region shows a region which will minimize the response.

3h)

The addition of replicates to the experiment will result in an increase in the degrees of freedom. Since replicates add more data points the model gets clearer giving a more accurate depiction of the standard error and confidence intervals. Also the removal of insignificant factors from the model will remove any unnecessary affects.
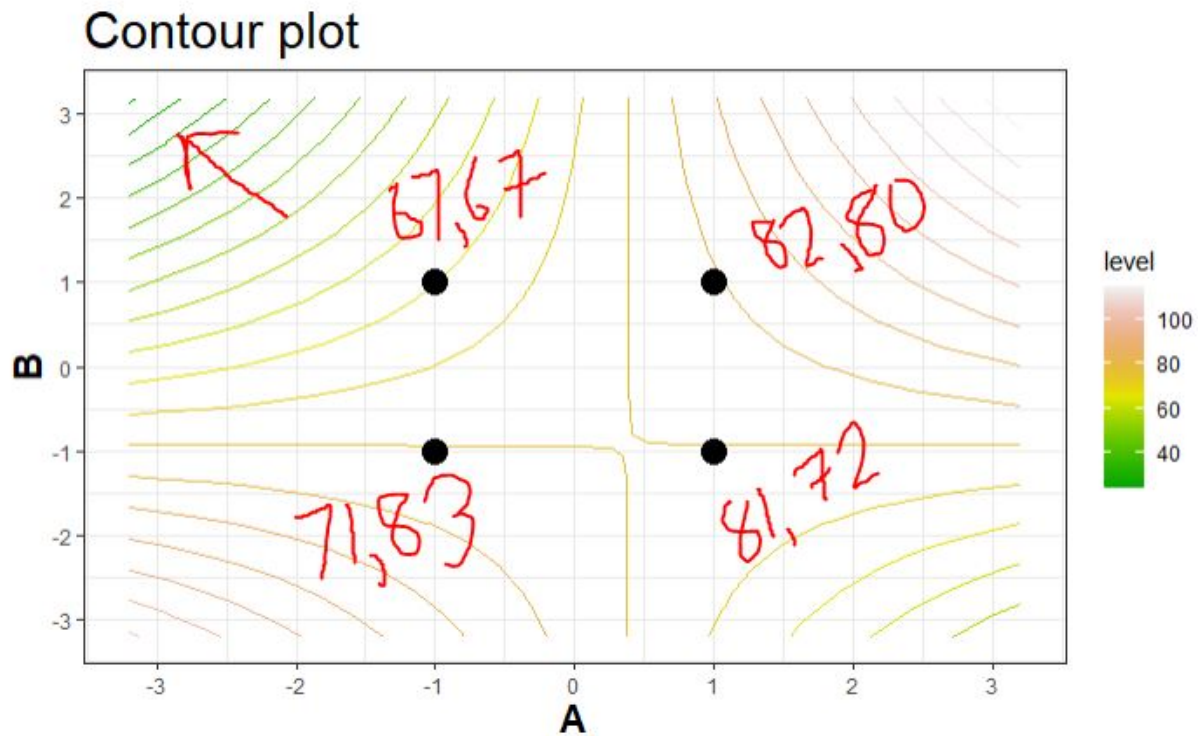
Figure 1: Contour Plot

Problem 4

4a)

```
A <- c(1, -1, 0, -1, 1)
B <- c(1, -1, 0, 1, -1)
Y <- c(25.5, 30.1, 27.5, 27.0, 28.0)

model1 <- lm(Y~A*B)
summary(model1)
```

```
##
## Call:
## lm(formula = Y ~ A * B)
##
## Residuals:
##     1     2     3     4     5
##  0.03  0.03 -0.12  0.03  0.03
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.62000    0.06000 460.333  0.00138 **
## A           -0.90000    0.06708 -13.416  0.04736 *
## B           -1.40000    0.06708 -20.870  0.03048 *
```

16

```
## A:B              0.15000     0.06708    2.236   0.26772
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1342 on 1 degrees of freedom
## Multiple R-squared:  0.9984, Adjusted R-squared:  0.9936
## F-statistic: 206.9 on 3 and 1 DF,  p-value: 0.05106
```

Therefore the linear squares prediction model for the data is y = 27.62 - 0.90Xa - 1.40Xb + 0.15XaXb

4b)

x = (real value - centerpoint)/0.5*range

Xa = 12-10/0.5*10 = 2/10 = 0.4

Xb = 900-500/0.5*500 = 400/250 = 1.6

4c)

Predicted value calculation

y = 27.62 - 0.9$0.4$ -$1.4$1.6 + 0.15$0.4$1.6 = 25.12 minutes

Therefore the predicted value of y at the given conditions is 25.12 minutes.

4d)

deltaA = 2 * (10/2) = 10 minutes

A(6) = BaselineValue + deltaA = 10 + 10 = 20

deltaXb = (bB/bA)$deltaxA$ = -(1.4/0.9)2 = 3.11

Overall to solve for the change in b we multiply the steepest slope by the value for change in A.

Therefore the overall change in Xa is +2 units and Xb is +3.11 units.

4e)

The given model is still consistent with this observation because the goal here is minimization. Since the goal is to minimize an additional experiment with a lower response value is a good indicator that the current model is still relevant and consistent. In the case the experiment resulted in a response value which was above the previous experiment that would mean the current model is no longer relative and a new model must be calculated with the previous experiment as the baseline. In a way this is a pivot in which the direction to the minimized model is found.

4f)

```
A <- c(1, -1, 0, -1, 1, 2)
B <- c(1, -1, 0, 1, -1, 3.11)
Y <- c(25.5, 30.1, 27.5, 27.0, 28.0, 22.4)

model1 <- lm(Y~A*B)
summary(model1)
```

```
##
## Call:
## lm(formula = Y ~ A * B)
##
## Residuals:
##           1          2          3          4          5          6
##   2.979e-02  2.997e-02 -1.200e-01  3.008e-02  3.011e-02  6.998e-05
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.62001    0.04213 655.613 2.33e-06 ***
## A           -0.89997    0.04574 -19.674 0.002574 **
## B           -1.39995    0.04324 -32.380 0.000952 ***
## A:B          0.15011    0.02697   5.567 0.030787 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09487 on 2 degrees of freedom
## Multiple R-squared:  0.9995, Adjusted R-squared:  0.9987
## F-statistic:  1255 on 3 and 2 DF,  p-value: 0.0007965
```
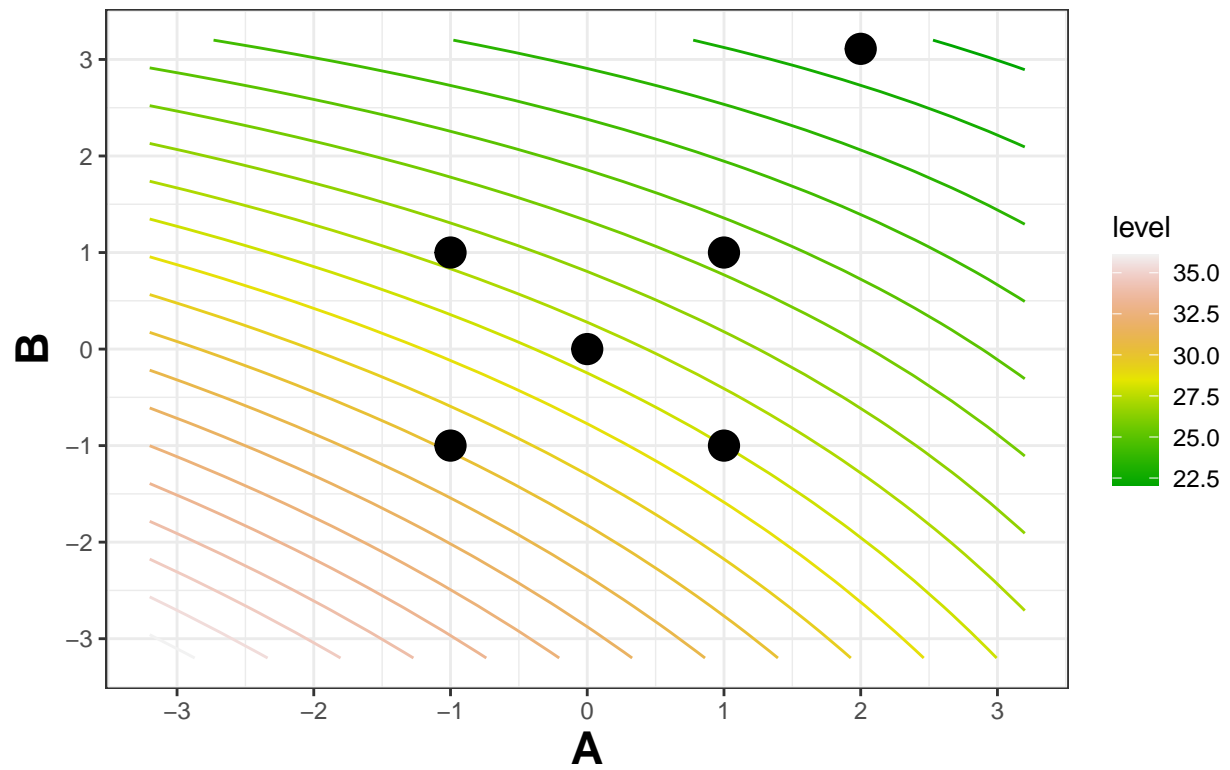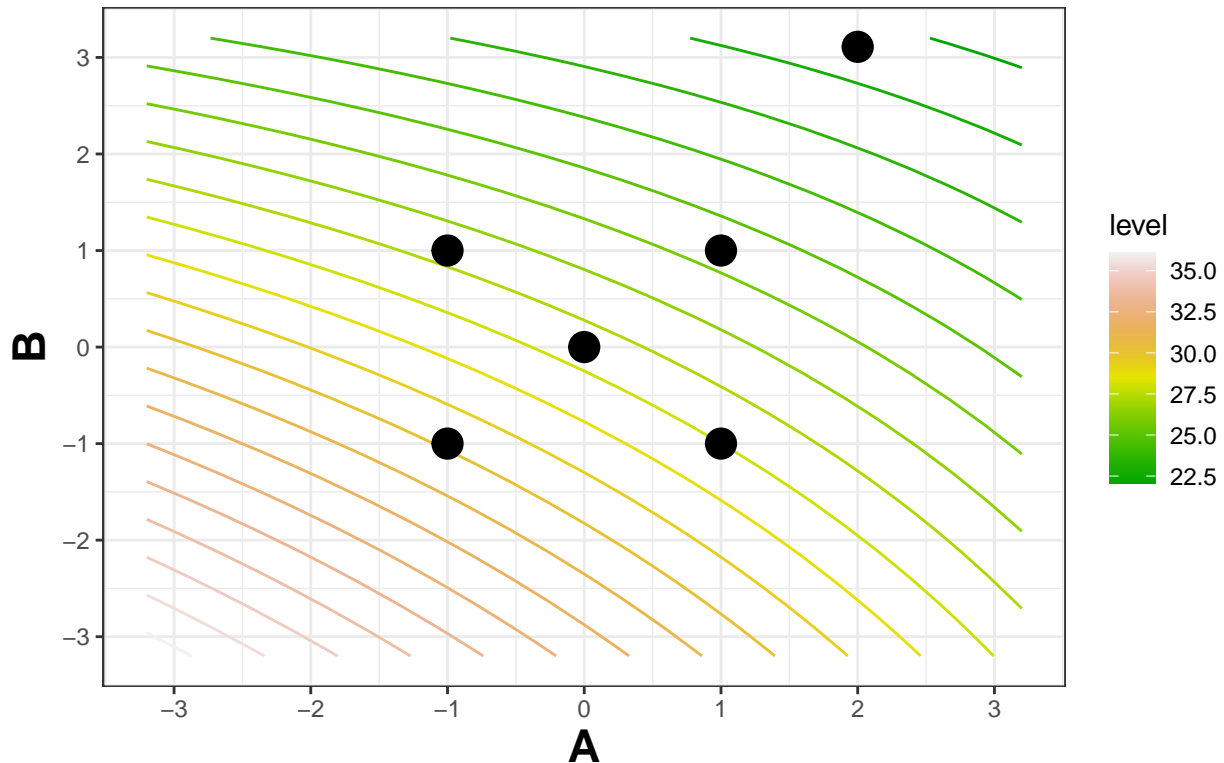
```
library("pid")
contourPlot(model1)
```

Contour plot

Contour plot

4g)

As we move along the response surface there are a couple scenarios that indicate that the optimum has been reached. If the response variable begins to plateau it can be a good indicator that the optimum has been reached. In theory you could keep pivoting and creating new models forever. So it is very important that there should be a cutoff point where the increase in optimization of the model is insignificant. This plateau of change should be determined via manual analysis and comparison of changes while also accounting for range of response values. Another method is in the case the response variable stays the same over many jumps, in many experiments this is not possible, but this is a surefire way of knowing that the optimization of the model has been reached. In the case the response variable increases/slope change direction; this indicates that the current model is insignificant. In this case the optimized model has been reached, sometimes multiple pivots are required if surefire precision is required. Finally if the interaction terms becomes very large and is within the range of the main effect terms it is usually a sign that the optimum has been reached.