STAT430 Assignment5

Jiajun Zhang

November 28, 2019

```r
library(FrF2)
library(BsMD)
```

Question 8.2

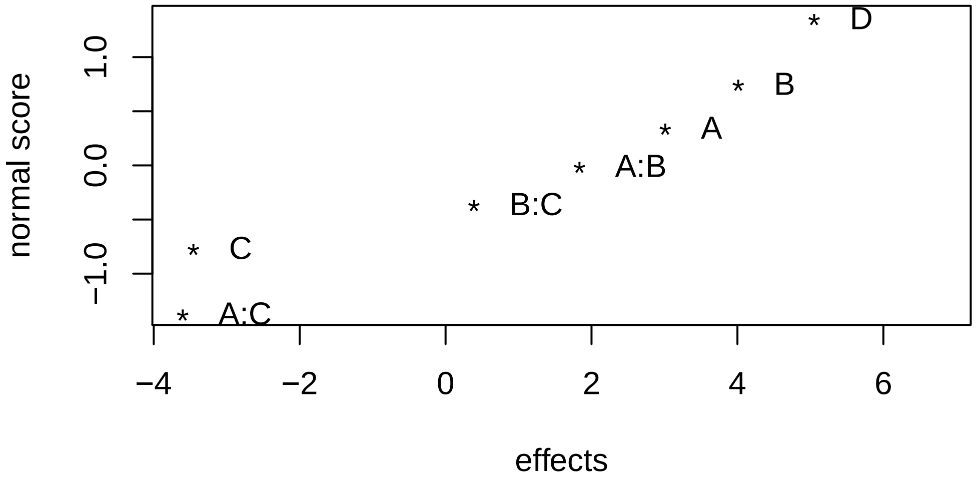
The one-half fraction of the 2^4 design is constructed as shown below.

```r
# 2^(4-1) design; 2^4 design with 8 runs
factors <- FrF2(nruns = 8, nfactors = 4, gen="ABC", randomize=FALSE)
summary(factors)

## Call:
## FrF2(nruns = 8, nfactors = 4, gen = "ABC", randomize = FALSE)
##
## Experimental design of type  FrF2.generators
## 8  runs
##
## Factor settings (scale ends):
##    A  B  C  D
## 1 -1 -1 -1 -1
## 2  1  1  1  1
##
## Design generating information:
## $legend
## [1] A=A B=B C=C D=D
##
## $generators
## [1] D=ABC
##
##
## Alias structure:
## $fi2
## [1] AB=CD AC=BD AD=BC
##
##
## The design itself:
##    A  B  C  D
## 1 -1 -1 -1 -1
## 2  1 -1 -1  1
## 3 -1  1 -1  1
## 4  1  1 -1 -1
## 5 -1 -1  1  1
## 6  1 -1  1 -1
## 7 -1  1  1 -1
## 8  1  1  1  1
## class=design, type= FrF2.generators
```

```
R1 <- c(7.037, 16.867, 13.876, 17.273, 11.846, 4.368, 9.360, 15.653)
titanium.data <- cbind.data.frame(factors, Crack.len = R1)
titanium.data.num <- titanium.data

titanium.data.num[, 1:4] <- lapply(titanium.data.num[, 1:4], as.character)
titanium.data.num[, 1:4] <- lapply(titanium.data.num[, 1:4], as.numeric)

lm.fit <- lm(Crack.len~A*B*C*D, data=titanium.data.num)
lm.fit

##
## Call:
## lm.default(formula = Crack.len ~ A * B * C * D, data = titanium.data.num)
##
## Coefficients:
## (Intercept)            A            B            C            D          A:B
##     12.0350       1.5053       2.0055      -1.7282       2.5255       0.9172
##         A:C          B:C          A:D          B:D          C:D        A:B:C
##     -1.8015       0.1943           NA           NA           NA           NA
##       A:B:D        A:C:D        B:C:D      A:B:C:D
##          NA           NA           NA           NA
```

By looking at the regression coefficients, A, B, C, D, and AC tend to be significant since they have larger effects relative to the other effects. And we can visualize them from the Daniel's plot below.

```
DanielPlot(lm.fit)
```



```
anova(lm(Crack.len~A+B+C+D+A:C, data=titanium.data))

## Analysis of Variance Table
##
## Response: Crack.len
##           Df Sum Sq Mean Sq F value  Pr(>F)
## A          1 18.126  18.126  5.1549 0.15119
## B          1 32.176  32.176  9.1505 0.09411 .
## C          1 23.895  23.895  6.7954 0.12102
## D          1 51.025  51.025 14.5110 0.06252 .
## A:C        1 25.963  25.963  7.3836 0.11295
## Residuals  2  7.033   3.516
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Then, using the results above to construct an ANOVA test, we can see that all the factors are not significant since their p-values are all greater than 0.05.

Question 8.4

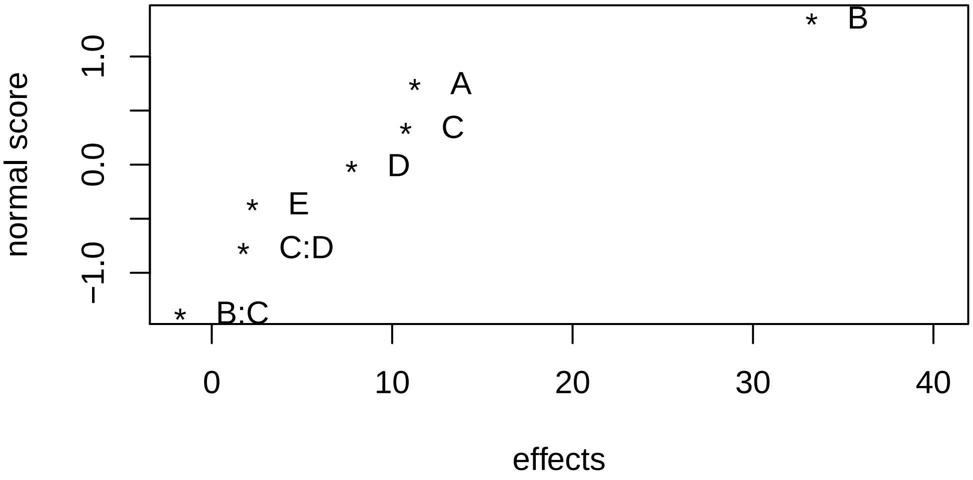
The one-quarter fraction of the 2^5 design is constructed as shown below.

```
# 2^(5-2) design; 2^5 design with 8 runs
factor1 <- FrF2(nruns = 8, nfactors = 5, gen = c("AB","AC"), randomize = FALSE)
summary(factor1)

## Call:
## FrF2(nruns = 8, nfactors = 5, gen = c("AB", "AC"), randomize = FALSE)
##
## Experimental design of type  FrF2.generators
## 8  runs
##
## Factor settings (scale ends):
##    A  B  C  D  E
## 1 -1 -1 -1 -1 -1
## 2  1  1  1  1  1
##
## Design generating information:
## $legend
## [1] A=A B=B C=C D=D E=E
##
## $generators
## [1] D=AB E=AC
##
##
## Alias structure:
## $main
## [1] A=BD=CE B=AD    C=AE    D=AB    E=AC
##
## $fi2
## [1] BC=DE BE=CD
##
##
## The design itself:
##    A  B  C  D  E
## 1 -1 -1 -1  1  1
## 2  1 -1 -1 -1 -1
## 3 -1  1 -1 -1  1
## 4  1  1 -1  1 -1
## 5 -1 -1  1  1 -1
## 6  1 -1  1 -1  1
## 7 -1  1  1 -1 -1
## 8  1  1  1  1  1
## class=design, type= FrF2.generators
```

```
obs <- c(6, 9, 35, 50, 18, 22, 40, 63)
plant.data <- cbind.data.frame(factor1, Yield = obs)
plant.data.num <- plant.data

plant.data.num[, 1:5] <- lapply(plant.data.num[, 1:5], as.character)
plant.data.num[, 1:5] <- lapply(plant.data.num[, 1:5], as.numeric)

lm.fit <- lm(Yield~A*B*C*D*E, data=plant.data.num)
lm.fit

## 
## Call:
## lm.default(formula = Yield ~ A * B * C * D * E, data = plant.data.num)
## 
## Coefficients:
## (Intercept)            A            B            C            D            E  
##      30.375        5.625       16.625        5.375        3.875        1.125  
##         A:B          A:C          B:C          A:D          B:D          C:D  
##          NA           NA       -0.875           NA           NA        0.875  
##         A:E          B:E          C:E          D:E        A:B:C        A:B:D  
##          NA           NA           NA           NA           NA           NA  
##       A:C:D        B:C:D        A:B:E        A:C:E        B:C:E        A:D:E  
##          NA           NA           NA           NA           NA           NA  
##       B:D:E        C:D:E      A:B:C:D      A:B:C:E      A:B:D:E      A:C:D:E  
##          NA           NA           NA           NA           NA           NA  
##     B:C:D:E    A:B:C:D:E
##           NA           NA
```

By looking at the regression coefficients, A, B, C, and D tend to be significant since they have larger effects relative to the other effects.

```
DanielPlot(lm.fit)
```



```
anova(lm(Yield~A+B+C+D, data=plant.data))

## Analysis of Variance Table
##
## Response: Yield
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## A          1  253.13  253.13  33.938 0.0100735 *
## B          1 2211.12 2211.12 296.464 0.0004268 ***
## C          1  231.13  231.13  30.989 0.0114387 *
## D          1  120.12  120.12  16.106 0.0277670 *
## Residuals  3   22.37    7.46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Using the results above to construct an ANOVA test, we can see that all the factors A, B, C, and D are significant since their p-values are less than 0.05.

Question 8.11

(a)

```r
matrix(c("-", "-", "-", "-", "+",
         "+", "-", "-", "+", "-",
         "-", "+", "-", "+", "+",
         "+", "+", "-", "-", "-",
         "-", "-", "+", "+", "-",
         "+", "-", "+", "-", "+",
         "-", "+", "+", "-", "-",
         "+", "+", "+", "+", "+"), nrow=8, ncol=5, byrow=T,
       dimnames=list(c("e", "ad", "bde", "ab", "cd", "ace", "bc", "abcde"),
                     c("A", "B", "C", "D", "E")))
```

```
##       A   B   C   D   E  
## e     "-" "-" "-" "-" "+"
## ad    "+" "-" "-" "+" "-"
## bde   "-" "+" "-" "+" "+"
## ab    "+" "+" "-" "-" "-"
## cd    "-" "-" "+" "+" "-"
## ace   "+" "-" "+" "-" "+"
## bc    "-" "+" "+" "-" "-"
## abcde "+" "+" "+" "+" "+"
```

From the output table above, we can see that for each row, the combinations of A and C result in E since the signs in E are the products of A and C signs. Similiarly, we can confirm that the product signs of B and D also have the same signs in E.

(b)

Since we have design generators ACE and BDE, then we can obtain the other generator ACEBDE = ABCDE^2 = ABCD. Then, the complete defining relation is

$$I = ACE = BDE = ABCD$$

Then, we compute the aliases in this design. For example, for facotr A, we will have aliases AACE = A^2CE = CE, ABDE, and AABCD = A^2BCD = BCD. Also, for factor B, we can have ABCE, DE, and ACD. Similarly, we can compute the rest of aliases. Then, the final aliases structure can be written as:

$$A = CE = ABDE = BCD$$

$$B = ABCE = DE = ACD$$

$$C = AE = BCDE = ABD$$

$$D = ACDE = BE = ABC$$

$$E = AC = BD = ABCDE$$

$$AB = BCE = ADE = CD$$

$$BC = ABE = CDE = AD$$

(c)

```
# 2^(5-2) design; 2^5 design with 8 runs
factor2 <- FrF2(nruns = 8, nfactors = 5, gen = c("ABC","AC"), randomize = FALSE)
# summary(factor2)

obs <- c(23.2, 16.9, 16.8, 15.5, 23.8, 23.4, 16.2, 18.1)
Chem.data <- cbind.data.frame(factor2, Yield = obs)
Chem.data.num <- Chem.data

Chem.data.num[, 1:5] <- lapply(Chem.data.num[, 1:5], as.character)
Chem.data.num[, 1:5] <- lapply(Chem.data.num[, 1:5], as.numeric)

lm.fit <- lm(Yield~A*B*C*D*E, data=Chem.data.num)
#Estimated effects
2 * coef(lm.fit)

## (Intercept)           A           B           C           D           E
##      38.475      -1.525      -5.175       2.275      -0.675       2.275
##         A:B         A:C         B:C         A:D         B:D         C:D
##       1.825          NA      -1.275          NA          NA          NA
##         A:E         B:E         C:E         D:E       A:B:C       A:B:D
##          NA          NA          NA          NA          NA          NA
##       A:C:D       B:C:D       A:B:E       A:C:E       B:C:E       A:D:E
##          NA          NA          NA          NA          NA          NA
##       B:D:E       C:D:E     A:B:C:D     A:B:C:E     A:B:D:E     A:C:D:E
##          NA          NA          NA          NA          NA          NA
##     B:C:D:E   A:B:C:D:E
##          NA          NA
```

Since $E = AC$ and $D = BE = ABC$, and from the results above, we can conclude that the estimated main effects are: $A = -1.525$, $B = -5.175$, $C = 2.275$, $AB = 1.825$, $AC = E = 2.275$, $BC = -1.275$, $ABC = D = -0.675$.

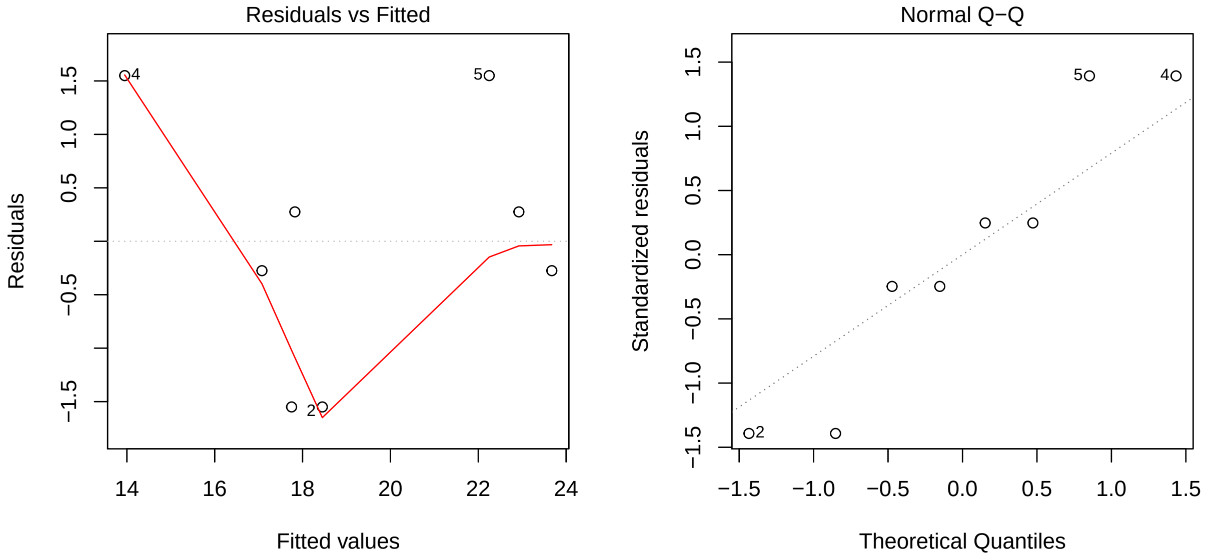
(d)

```
anova(lm(Yield~A+B+C+D+E, data=Chem.data))

## Analysis of Variance Table
##
## Response: Yield
##           Df Sum Sq Mean Sq F value Pr(>F)
## A          1  4.651   4.651  0.9385 0.4349
## B          1 53.561  53.561 10.8068 0.0814 .
## C          1 10.351  10.351  2.0885 0.2853
## D          1  0.911   0.911  0.1839 0.7098
## E          1 10.351  10.351  2.0885 0.2853
## Residuals  2  9.913   4.956
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA test is performed above. We can see that all factors are not significant at 5% significance level. Since AB and AD are both aliased with other factors, and we can see that all two and three factor interactions are negligible from the main effects output. Then, we can discard them and use them as error.

(e)

```
par(mfrow = c(1,2))
res <- aov(Yield~A+B+C+D+E, data=Chem.data)
plot(res,1); plot(res,2)
```



There is a sort of pattern that the constant variance assumption not satisfied just looking at the residuals vs fitted values plot. But, we might need a further test for that. And normality plot looks nothing unusual.

Question 8.13

By having two four-factor interactions as generators, we can choose $F = CDE$, and $G = ABC$. Then, we will have $I = CDEF = ABCG$. Also, $CDEFABCG = ABDEFG$. Therefore, the complete defining relation is:

$$I = CDEF = ABCG = ABDEFG$$

Then, this is a resolution IV design. The design can be constructed as below.

```
A <- rep(c("-", "+"), 16)
B <- rep(c("-", "-", "+", "+"), 8)
C <- rep(c(rep("-",4), rep("+",4)), 4)
D <- rep(c(rep("-",8), rep("+",8)), 2)
E <- rep(c(rep("-",16), rep("+",16)))
#F=CDE
F <- c(rep("-", 4), rep("+", 8), rep("-", 4), rep("+", 4), rep("-", 8), rep("+", 4))
#G=ABC
G <- c("-", "+", "+", "-", "+", "-", "-", "+", "-", "+", "+", "-", "+", "-", "-", "+",
       "-", "+", "+", "-", "+", "-", "-", "+", "-", "+", "+", "-", "+", "-", "-", "+")
Trt <- c("(1)", "ag", "bg", "ab", "cfg", "acf", "bcf", "abcfg", "df", "adfg", "bdfg",
         "abdf", "cdg", "acd", "bcd", "abcdg", "ef", "aefg", "befg", "abef", "ceg",
         "ace", "bce", "abceg", "de", "adeg", "bdeg", "abde", "cdefg", "acdef", "bcdef",
         "abcdefg")
design <- data.frame(A=A, B=B, C=C, D=D, E=E, F=F, G=G, Trt.Combination=Trt)
design
```

```
##    A B C D E F G Trt.Combination
## 1  - - - - - - -             (1)
## 2  + - - - - - +              ag
## 3  - + - - - - +              bg
## 4  + + - - - - -              ab
## 5  - - + - - + +             cfg
## 6  + - + - - + -             acf
## 7  - + + - - + -             bcf
## 8  + + + - - + +           abcfg
## 9  - - - + - + -              df
## 10 + - - + - + +            adfg
## 11 - + - + - + +            bdfg
## 12 + + - + - + -            abdf
## 13 - - + + - - +             cdg
## 14 + - + + - - -             acd
## 15 - + + + - - -             bcd
## 16 + + + + - - +           abcdg
## 17 - - - - + + -              ef
## 18 + - - - + + +            aefg
## 19 - + - - + + +            befg
## 20 + + - - + + -            abef
## 21 - - + - + - +             ceg
## 22 + - + - + - -             ace
## 23 - + + - + - -             bce
## 24 + + + - + - +           abceg
## 25 - - - + + - -              de
## 26 + - - + + - +            adeg
## 27 - + - + + - +            bdeg
## 28 + + - + + - -            abde
## 29 - - + + + + +           cdefg
## 30 + - + + + + -           acdef
## 31 - + + + + + -           bcdef
## 32 + + + + + + +         abcdefg
```

Then, we can write down the aliases structure for this design as:

$$A = ACDEF = BCG = BDEFG$$

$$B = BCDEF = ACG = ADEFG$$

$$C = DEF = ABG = ABCDEFG$$

$$D = CEF = ABCDG = ABEFG$$

$$E = CDF = ABCEG = ABDFG$$

$$F = CDE = ABCFG = ABDEG$$

$$G = CDEFG = ABC = ABDEF$$

$$AB = ABCDEF = CG = DEFG$$

$$AC = ADEF = BG = BCDEFG$$

$$AD = ACEF = BCDG = BEFG$$

$$AE = ACDF = BCEG = BDFG$$

$$AF = ACDE = BCFG = BDEG$$

$$AG = ACDEFG = BC = BDEF$$

$$BD = BCEF = ACDG = AEFG$$

$$BE = BCDF = ACEG = ADFG$$

$$BF = BCDE = ACFG = ADEG$$

$$CD = EF = ABDG = ABCEFG$$

$$CE = DF = ABEG = ABCDFG$$

$$CF = DE = ABFG = ABCDEG$$

$$DG = CEFG = ABCD = ABEF$$

$$EG = CDFG = ABCE = ABDF$$

$$FG = CDEG = ABCF = ABDE$$

Then, we can outline the ANOVA table as shown below. Here, AB is also CG. Similiarly, $AC = BG$, $AG = BC$, $CD = EF$, $CE = DF$, $CF = DE$.

```
Trt.comb <- c("A", "B", "C", "D", "E", "F", "G", "AB", "AC", "AD", "AE", "AF",
              "AG", "BD", "BE", "CD", "CE", "CF", "DG", "EG", "FG", "Error", "Total")
Df <- c(rep(1, 21), 9, 31)
ANOVA <- data.frame(Treatments=Trt.comb, DF=Df)
ANOVA
```

```
##    Treatments DF
## 1           A  1
## 2           B  1
## 3           C  1
## 4           D  1
## 5           E  1
## 6           F  1
## 7           G  1
## 8          AB  1
## 9          AC  1
## 10         AD  1
## 11         AE  1
## 12         AF  1
## 13         AG  1
## 14         BD  1
## 15         BE  1
## 16         CD  1
## 17         CE  1
## 18         CF  1
## 19         DG  1
## 20         EG  1
## 21         FG  1
## 22      Error  9
## 23      Total 31
```

Question 10.1

```
hardwood.data <- data.frame(Strength=c(160, 171, 175, 182, 184, 181, 188, 193, 195, 200),
                            Percent.Hardwood=c(10, 15, 15, 20, 20, 20, 25, 25, 28, 30))
```

(a)

```
lm.fit <- lm(Strength~Percent.Hardwood, data=hardwood.data)
lm.fit

##
## Call:
## lm.default(formula = Strength ~ Percent.Hardwood, data = hardwood.data)
##
## Coefficients:
##      (Intercept)  Percent.Hardwood
##          143.824             1.879
```

A linear regression model can be written as:

$$\hat{y}_{Strength} = 143.824 + 1.879x_{PercentHardwood}$$

(b)

```
summary(lm.fit)

##
## Call:
## lm.default(formula = Strength ~ Percent.Hardwood, data = hardwood.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7903 -1.3206 -0.2903  1.8080  2.9961
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      143.8244     2.5215   57.04 9.91e-12 ***
## Percent.Hardwood   1.8786     0.1165   16.12 2.20e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.203 on 8 degrees of freedom
## Multiple R-squared:  0.9701,Adjusted R-squared:  0.9664
## F-statistic:   260 on 1 and 8 DF,  p-value: 2.197e-07
```

Based on the summary result, we can conclude that the predictor of percent hardwood is statistically significant in the model since its p-value is less than 0.05.

(c)

```
confint(lm.fit, 'Percent.Hardwood', level = 0.95, interval = 'confident')
```

```
##                     2.5 %   97.5 %
## Percent.Hardwood 1.609968 2.147303
```

A 95% confidence interval on β_1 is [1.610, 2.147].

Question 10.6

```
bearing.data <- data.frame(Bearing=c(193, 230, 172, 91, 113, 125),
                           Oil.Vis=c(1.6, 15.5, 22.0, 43.0, 33.0, 40.0),
                           Load=c(851, 816, 1058, 1201, 1357, 1115))
```

(a)

```
lm.fit <- lm(Bearing~., data=bearing.data)
lm.fit

##
## Call:
## lm.default(formula = Bearing ~ ., data = bearing.data)
##
## Coefficients:
## (Intercept)      Oil.Vis         Load
##    350.9943      -1.2720      -0.1539
```

A multiple linear regression model can be written as:

$$\hat{y} = 350.9943 - 1.2720x_1 - 0.1539x_2$$

where $\hat{y}$ is the esimated bearing, x_1 is the oil viscosity and x_2 is the load.

(b)

```
summary(lm.fit)

##
## Call:
## lm.default(formula = Bearing ~ ., data = bearing.data)
##
## Residuals:
##       1       2       3       4       5       6
## -24.987  24.307  11.820 -20.460  12.830  -3.511
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 350.99427   74.75307   4.695   0.0183 *
## Oil.Vis      -1.27199    1.16914  -1.088   0.3562
## Load         -0.15390    0.08953  -1.719   0.1841
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.5 on 3 degrees of freedom
## Multiple R-squared:  0.8618,Adjusted R-squared:  0.7696
## F-statistic: 9.353 on 2 and 3 DF,  p-value: 0.05138
```

From the summary result, we can see that there are no significant relationships between both predictors and the response since the p-values of both predictors are greater than 0.05.

(c)

```
#t statistics
summary(lm.fit)[["coefficients"]][, "t value"]

## (Intercept)     Oil.Vis        Load
##    4.695382   -1.087974   -1.719030

#Critical values
qt(0.025, 6-1)

## [1] -2.570582
```

First, we will have our null hypothesis stated as there is no linear relationship between x and y and alternative hypothesis is there is linear relationship. Since -$t_{0.025,5}$ = -2.57 and both predictors have t statistics greater than -2.57, then we can conclude that there are no linear relationships between the predictors and the response.

Question 10.20

(a)

$$DF_{model} = \frac{534.66}{133.67} = 4$$

$$SS_{error} = 590.25 - 534.66 = 55.59$$

$$DF_{error} = 19 - 4 = 15$$

$$MS_{error} = \frac{55.59}{15} = 3.71$$

$$F = \frac{133.67}{3.71} = 36.03$$

(b)

There are 4 predictors in the model since df_{model} is 4.

(c)

```r
#F critical
qf(0.999, 4, 15)

## [1] 8.252684
```

Since F_0=36.03 $>$ $F_{0.001,4,15}$=8.25, then the p-value must be smaller than 0.001.

(d)

```r
abs(qt(0.0005, 15))

## [1] 4.072765
```

Since $t_{\alpha/2,n-k-1} = t_{0.0005,15} = 4.073$ and $6.11 > 4.073$, then we can conclude that the p-value is less than 0.001 in this case.

(e)

$$R^2 = \frac{SS_R}{SS_T} = \frac{534.66}{590.25} = 90.58\%$$

Question 10.22

(a)

$$SS_{error} = 275.60 - 245.86 = 29.74$$

$$DF_{error} = (45 - 1) - (3 - 1) = 42$$

$$MS_{error} = \frac{29.74}{42} = 0.708$$

(b)

$$MS_{model} = \frac{245.86}{2} = 122.93$$

(c)

$$F = \frac{122.93}{0.708} = 173.63$$

```
#F critical
qf(0.999, 2, 42)

## [1] 8.179405
```

Since F_0 is way larger than 8.179, we can conclude that p-value is less than 0.001 which is highly significant.

(d)

$$R^2 = \frac{245.86}{275.60} = 89.2\%$$

(e)

$$R^2_{adj} = 1 - (\frac{n-1}{n-p})(1 - R^2) = 1 - (\frac{44}{42})(1 - 0.892) = 88.7\%$$