

# Stats\_101B\_HW\_4\_Charles\_Liu

Charles Liu (304804942)

5/18/2020

## Loading Necessary Packages

```
library(DoE.base)

## Warning: package 'DoE.base' was built under R version 3.6.3

library(unrepx)

## Warning: package 'unrepx' was built under R version 3.6.3

library(gplots)

## Warning: package 'gplots' was built under R version 3.6.3

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.6.3
```

## 1a) (Problem 6.22a)

Analyze the data from this experiment. Which factors significantly affect the customer response rate?

```
# Responses
response1 <- c(50,44,46,42,49,48,47,56,
              54,42,48,43,46,45,48,54)

# Contrasts
k=3
contrast_single1 <- contr.FrF2(2^(k))
colnames(contrast_single1) <- c("A", "B", "AB", "C", "AC", "BC", "ABC")
contrast_single1

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

# Design
contrast1 = rbind(contrast_single1, contrast_single1)
```

```
design1 <- contrast1[,2^(0:(k-1))]
design1
```

```
##      A  B  C
## 1 -1 -1 -1
## 2  1 -1 -1
## 3 -1  1 -1
## 4  1  1 -1
## 5 -1 -1  1
## 6  1 -1  1
## 7 -1  1  1
## 8  1  1  1
## 1 -1 -1 -1
## 2  1 -1 -1
## 3 -1  1 -1
## 4  1  1 -1
## 5 -1 -1  1
## 6  1 -1  1
## 7 -1  1  1
## 8  1  1  1
```

```
# Effects
df1 <- data.frame(contrast1,response1)
effect1 <- t(response1)%*%contrast1/(4*2) # 2 replicates
effect1
```

```
##           A      B    AB C AC  BC ABC
## [1,] -1.75 0.75 3.25 3  5 3.5  1
```

We can see from here that (A) and (C) Main Effects are significant to the customer responses. We also see that the Interaction Effects of (AB), (AC), and (BC) are also significant to the customer responses. For the Main Effect (A), it has a negative relation with the responses; this suggests that increasing the class from 3rd to 1st will decrease the customer responses. For the Main Effect (C), it has a positive relation with the responses; this suggests that increasing (C) from the low level (19.95) to the high level (24.95) will increase the customer responses.

## 1b) (Problem 6.22b)

Analyze the residuals from this experiment. Are there any indications of model inadequacy?

```
anova1 <- aov(response1~factor(A)*factor(B)*factor(C),data=df1)
summary(anova1)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## factor(A)      1  12.25    12.25   4.083 0.077971 .
## factor(B)      1   2.25     2.25   0.750 0.411694
## factor(C)      1  36.00    36.00  12.000 0.008516 **
## factor(A):factor(B)  1  42.25    42.25  14.083 0.005602 **
## factor(A):factor(C)  1 100.00   100.00  33.333 0.000418 ***
## factor(B):factor(C)  1  49.00    49.00  16.333 0.003728 **
## factor(A):factor(B):factor(C)  1  4.00     4.00   1.333 0.281537
## Residuals      8  24.00     3.00
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

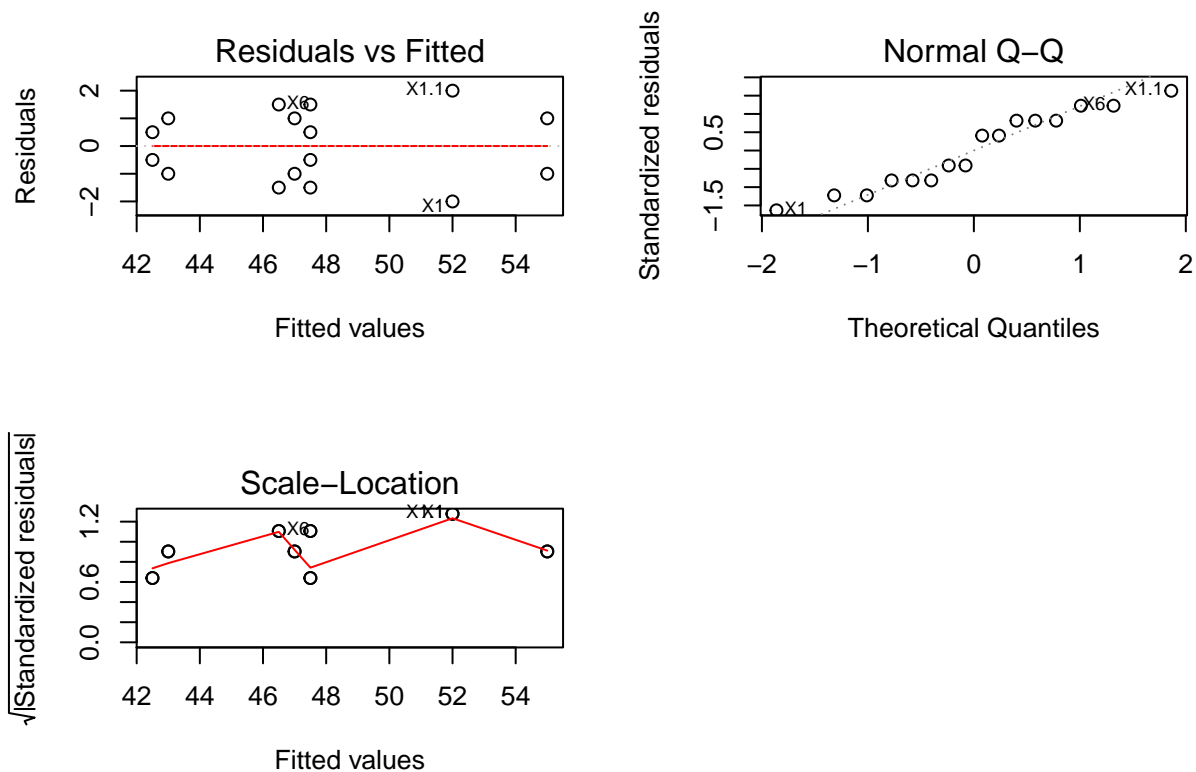
regression1 <- lm(response1 ~ A*B*C, data=df1)
summary(regression1)

##
## Call:
## lm.default(formula = response1 ~ A * B * C, data = df1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
##     -2.00    -1.00     0.00     1.00     2.00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   47.625     0.433  109.985 5.22e-14 ***
## A             -0.875     0.433   -2.021 0.077971 .
## B              0.375     0.433    0.866 0.411694
## C              1.500     0.433    3.464 0.008516 **
## A:B            1.625     0.433    3.753 0.005602 **
## A:C            2.500     0.433    5.774 0.000418 ***
## B:C            1.750     0.433    4.041 0.003728 **
## A:B:C          0.500     0.433    1.155 0.281537
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.732 on 8 degrees of freedom
## Multiple R-squared:  0.911, Adjusted R-squared:  0.8332
## F-statistic: 11.7 on 7 and 8 DF, p-value: 0.001223

par(mfrow = c(2,2))
plot(regression1)

## hat values (leverages) are all = 0.5
## and there are no factor predictors; no plot no. 5

```



We can see here that the model has Multiple R-squared = 0.911 and Adjusted R-squared = 0.8332. This indicates that our model is good. From our plots, we can see that it satisfies the Normality Assumption relatively so. As for the Residuals plot, the red dotted line is perfectly straight and no patterns from the scattered points, therefore we can say our model has no inadequacy. The Linearity Assumption is satisfied. Although, we can see the Homoscedasticity Assumption is not satisfied as we can see from the Scale-Location Plot.

### 1c) (Problem 6.22c)

What would you recommend to the company?

**Answer:** I would recommend the company should use 1st Class Mail, with Colored Brochures, and offer price of \$24.95 to maximize the average response rate.

### 2a) (Problem 6.17a)

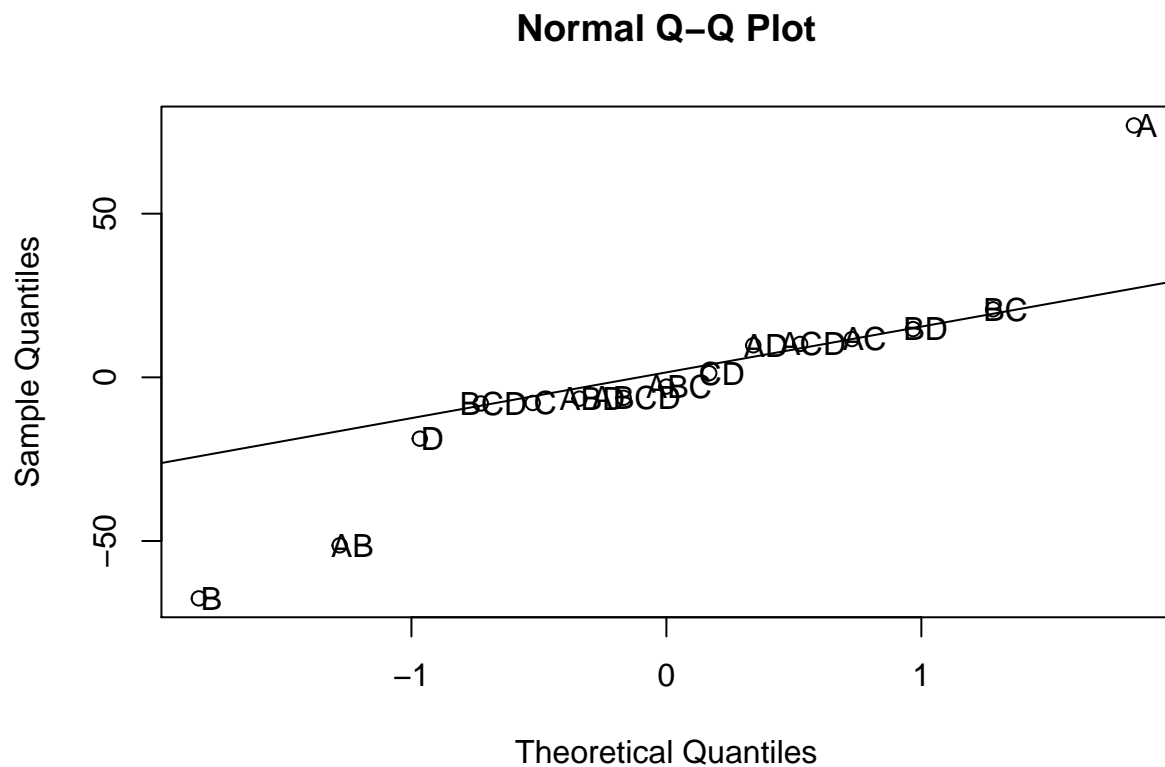
Construct a normal probability plot of these effects.

```
# Effects
effect2 <- c(76.95,-67.52,-7.84,-18.73,
            -51.32,11.69,9.78,20.78,14.74,1.27,
            -2.82,-6.50,10.20,-7.98,-6.25)
names(effect2) <- c("A","B","C","D","AB","AC","AD","BC","BD","CD","ABC",
                  "ABD","ACD","BCD","ABCD")
effect2
```

##	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC
----	---	---	---	---	----	----	----	----	----	----	-----

```
## 76.95 -67.52 -7.84 -18.73 -51.32 11.69 9.78 20.78 14.74 1.27 -2.82
## ABD ACD BCD ABCD
## -6.50 10.20 -7.98 -6.25
```

```
# Normality Probability Plot
qqnorm(effect2)
text(qqnorm(effect2)$x + 0.05, qqnorm(effect2)$y + 0.05, names(effect2))
qqline(effect2)
```



We can see here that the Normality Probability Plot is relatively normal, and therefore, it satisfies the Normality Assumption.

## 2b) (Problem 6.17b)

Identify a tentative model, based on the plot of the effects in part (a).

**Answer:**

$$\hat{y} = \text{Intercept} + 38.475x_a - 33.76x_b - 25.66x_a x_b$$

## 2c) (Problem 6.17c)

- (c) Identify the significant effects through constructing 95% confidence intervals. Do this first:
- (d) if it is reasonable to assume that the interactions involving three or more factors are negligible, and then
- (ii) using Lenth's method.

```

# (i) Assume that interactions involving 3 or more factors are negligible
# (ii) Use Lenth's Method

# The estimate of tau (by hand)
effect_abs2 <- abs(effect2)
tau2 <- 1.5*median(effect_abs2)
tau2

## [1] 15.3

# Truncated effect (by hand)
truncated_effect_abs2 <- effect_abs2[effect_abs2<(2.5*tau2)]
tau_tilde2.1 <- 1.5*median(truncated_effect_abs2)
tau_tilde2.1

## [1] 13.32

# Truncated effect (by R function)
tau_tilde2.2 <- PSE(effect2, method = "Lenth")
tau_tilde2.2 # This is the same exact value found by doing it "by hand"

## Lenth_PSE
##      13.32

### T-tests and C.I.
t_value_lenth2 <- round(qt((1-0.025),5),3)
t_value_lenth2

## [1] 2.571

## t-tests
t_tests2 <- eff.test(effect2, method = "Lenth")
t_tests2[, -5]

##      effect Lenth_PSE t.ratio p.value
## A      76.95     13.32   5.777 0.0017
## B     -67.52     13.32  -5.069 0.0026
## AB    -51.32     13.32  -3.853 0.0082
## BC     20.78     13.32   1.560 0.1215
## D     -18.73     13.32  -1.406 0.1540
## BD     14.74     13.32   1.107 0.2482
## AC     11.69     13.32   0.878 0.3541
## ACD     10.20     13.32   0.766 0.4151
## AD       9.78     13.32   0.734 0.4350
## BCD     -7.98     13.32  -0.599 0.5826
## C       -7.84     13.32  -0.589 0.5893
## ABD     -6.50     13.32  -0.488 0.6532
## ABCD    -6.25     13.32  -0.469 0.6647
## ABC     -2.82     13.32  -0.212 0.8424
## CD       1.27     13.32   0.095 0.9277

## 95% CIs
names_effects2 <- c("A", "B", "C", "D", "AB", "AC", "AD", "BC", "BD", "CD", "ABC",
                    "ABD", "ACD", "BCD", "ABCD")
ci2 <- cbind(t(effect2-t_value_lenth2*tau_tilde2.2), t(effect2+t_value_lenth2*tau_tilde2.2))
ci2 <- matrix(c(names_effects2, ci2), nrow=15, ncol=3, byrow=FALSE)
ci2

```

```
##      [,1]      [,2]      [,3]
## [1,] "A"      "42.70428"  "111.19572"
## [2,] "B"      "-101.76572" "-33.27428"
## [3,] "C"      "-42.08572"  "26.40572"
## [4,] "D"      "-52.97572"  "15.51572"
## [5,] "AB"     "-85.56572"  "-17.07428"
## [6,] "AC"     "-22.55572"  "45.93572"
## [7,] "AD"     "-24.46572"  "44.02572"
## [8,] "BC"     "-13.46572"  "55.02572"
## [9,] "BD"     "-19.50572"  "48.98572"
## [10,] "CD"    "-32.97572"  "35.51572"
## [11,] "ABC"   "-37.06572"  "31.42572"
## [12,] "ABD"   "-40.74572"  "27.74572"
## [13,] "ACD"   "-24.04572"  "44.44572"
## [14,] "BCD"   "-42.22572"  "26.26572"
## [15,] "ABCD" "-40.49572"  "27.99572"
```

```
# The significant factor from C.I.'s
significant_ci2 <- (ci2[,1])[c(which(ci2[,2]>0), which(ci2[,3]<0))]
significant_ci2
```

```
## [1] "A" "B" "AB"
```

- (i) It is reasonable to assume that the interactions involving three or more factors are negligible. Therefore, we need to use (ii) Lenth's method. We also see that the Effects (A), (B), and (AB) are significant.