

# STAT502 - HW5

*Namit Chauhan*

## Q.1.

Summary of Data:

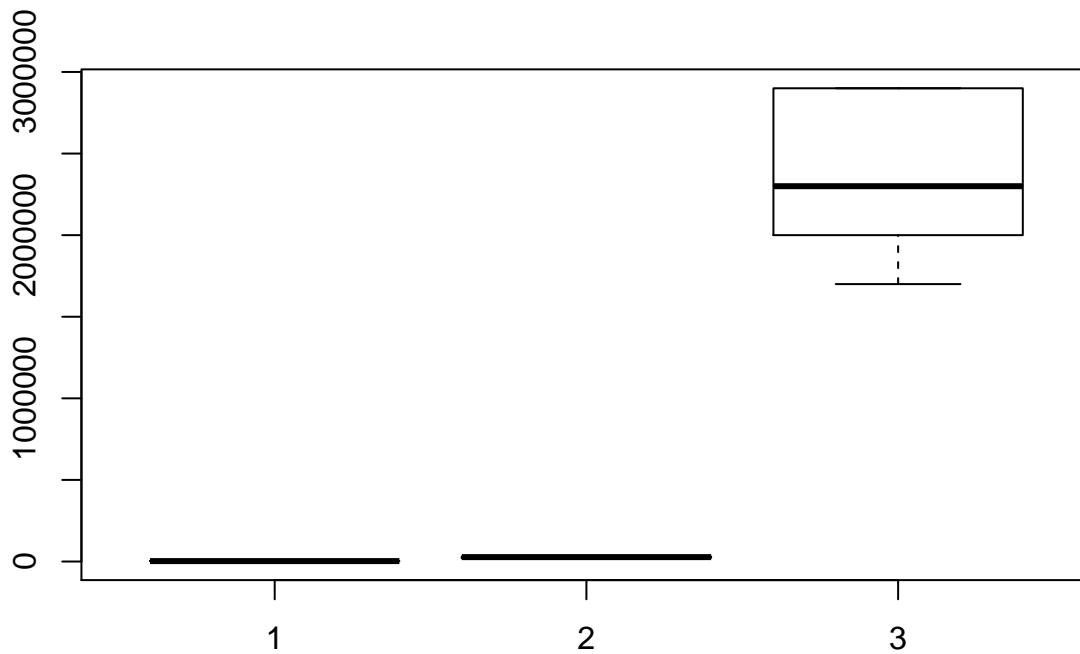
```
## Min.      :    2000   1st Qu.:    3050   Median :   27000   Mean      :  796527   3rd Qu.:1850000   Max.      :29000000
```

```
##
```

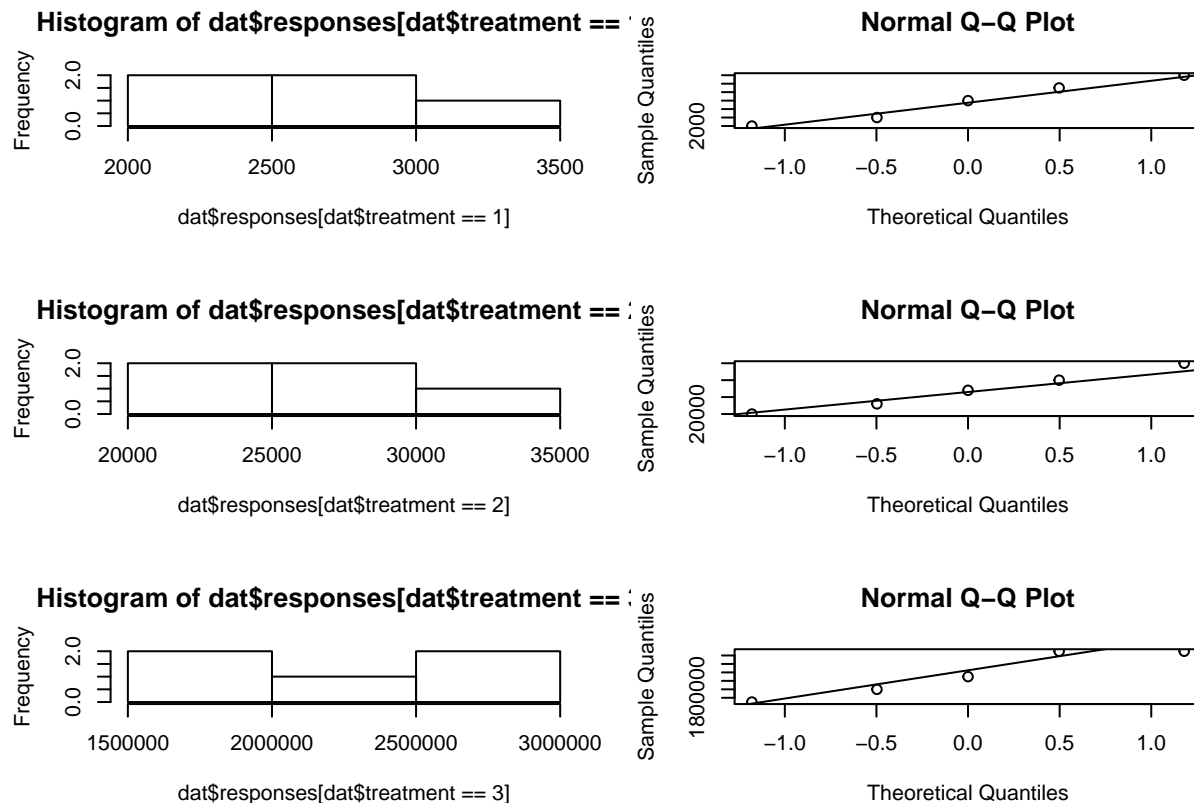
```
## group means 2580 27000 2360000
```

```
##
```

```
## group vars 242000 34500000 2.88e+11
```



## a) Univariate Normality



Normality satisfied for treatment 2, 3.

## Bartleys test for equal variances:

Equal variances condition not satisfied.

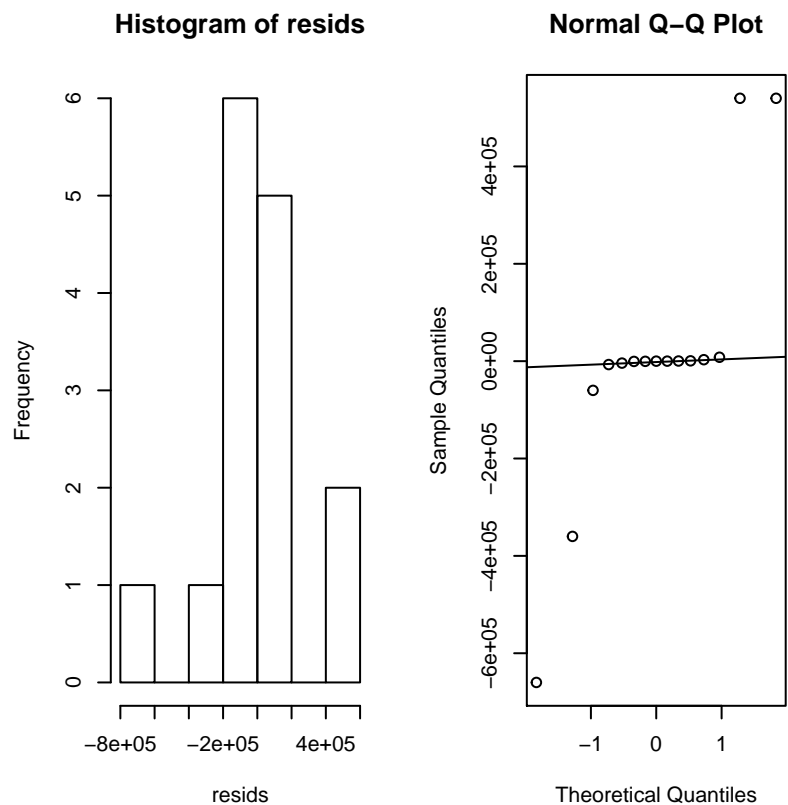
```
##
## Bartlett test of homogeneity of variances
##
## data: dat$response by dat$treatment
## Bartlett's K-squared = 71.006, df = 2, p-value = 3.813e-16
```

Also ratio of max\_var is > min\_var by 7 (thumbrule).

```
##
## max_var/min_var 1190083
```

## b) Perform anova and analyse residuals:

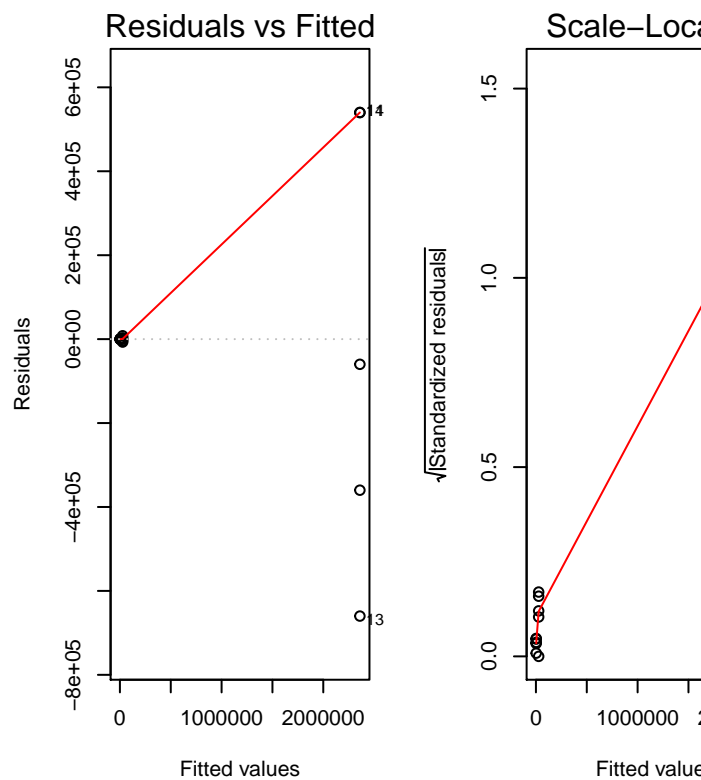
```
##           Df    Sum Sq  Mean Sq F value    Pr(>F)
## treatment   2 1.833e+13  9.167e+12   95.48 4.27e-08 ***
## Residuals  12 1.152e+12  9.601e+10
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Residuals are not normal based on qq plot:

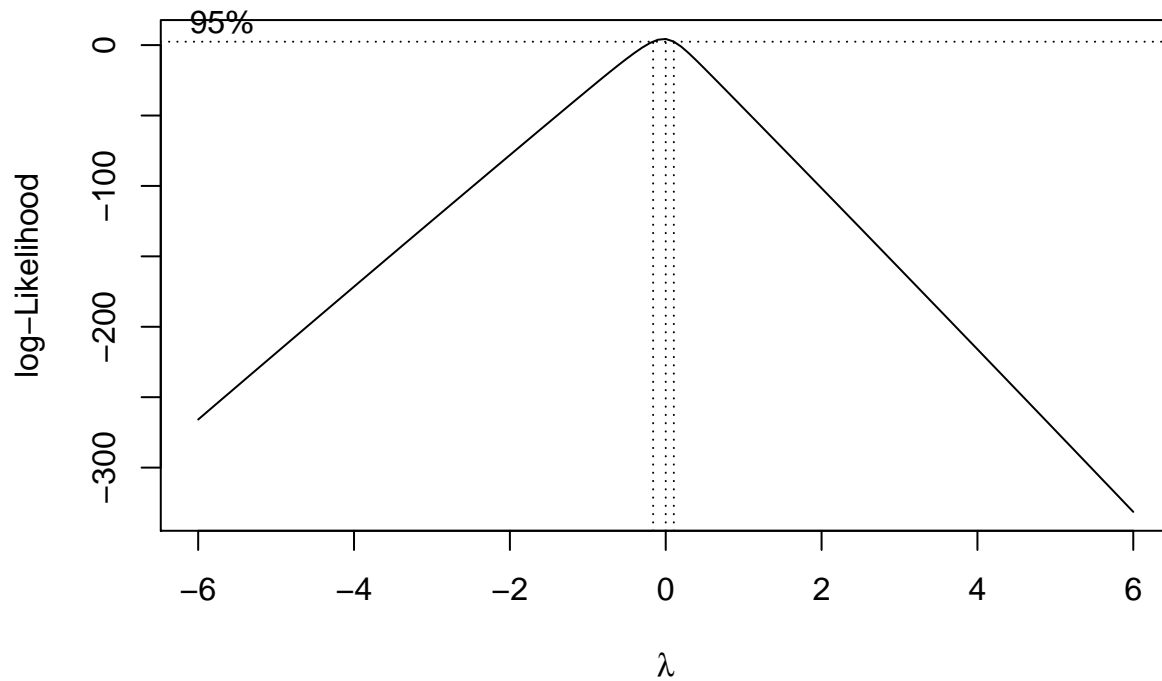
Also ratio of max\_var of residuals is > min\_var of residuals by 7 (thumbrule):

```
##
## max_var/min_var of residuals 1190083
```



Mean variance relationship is linear - another assumption violated.

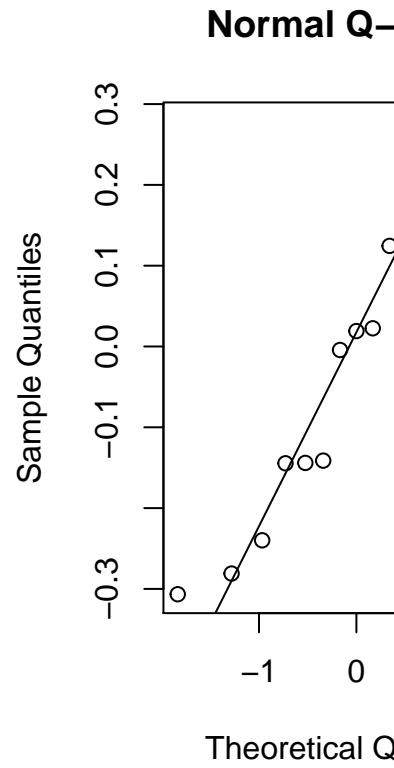
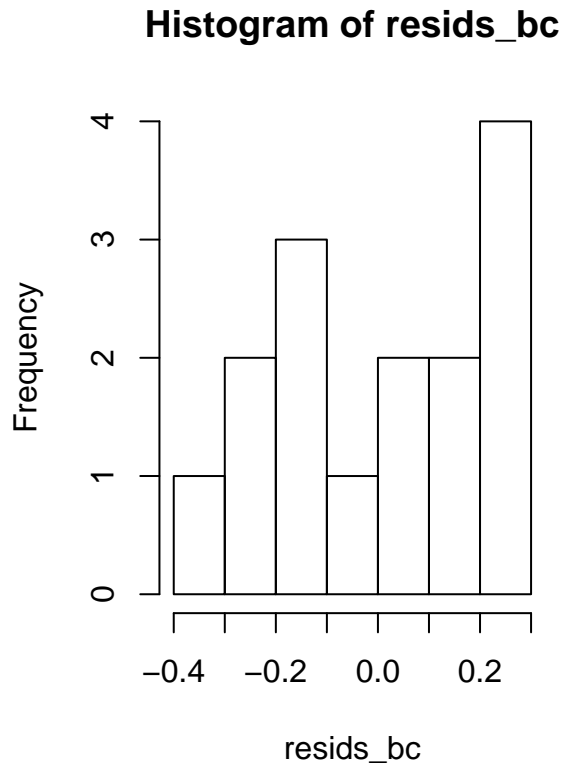
### c) Box-cox transform



```
##      box.x      box.y
## 61      0 4.349758
##
## optimal lambda 0
Lambda eq 0 is equivalent to log transform
dat$responses_box = log(dat$responses)
```

### d) Perform anova on log-transformed data and analyse residuals:

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treatment    2 119.77   59.88   1283 1.02e-14 ***
## Residuals   12   0.56    0.05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

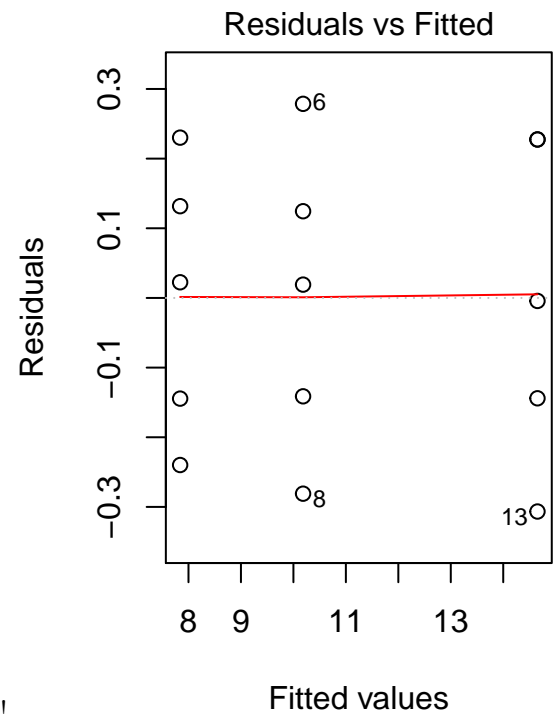


Residuals are normal based on qq plot:

Also ratio of max\_var of residuals is < min\_var of residuals by 4 (thumbrule)

##

## max\_var/min\_var of residuals 1.462541



Let's take a look at the mean variance relationship: There is no relationship!

Conclusion - difference in responses due to treatments is observed as: 1) F-ratio is high (>1) and 2) significant

##

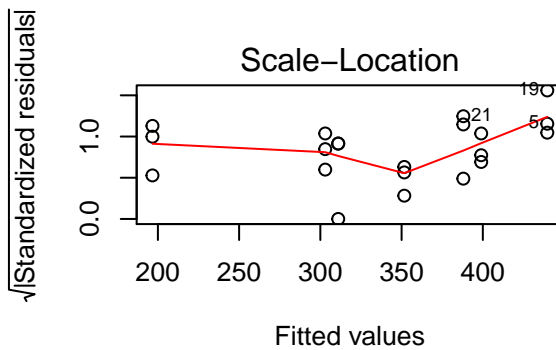
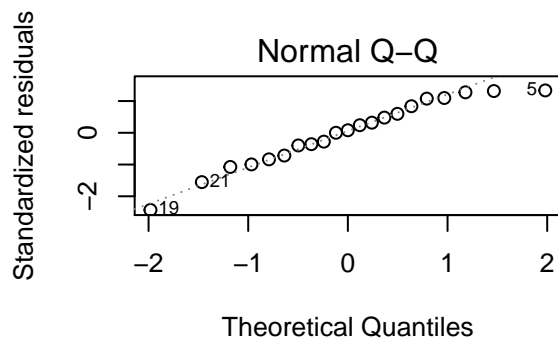
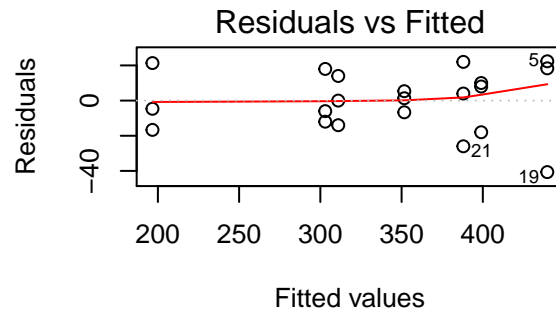
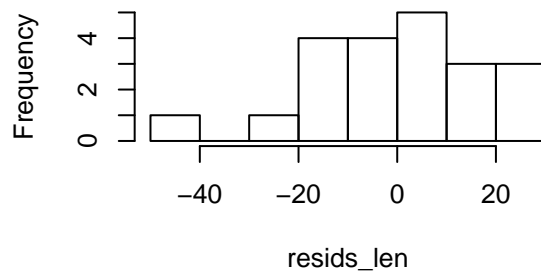
Df	Sum Sq	Mean Sq	F value	Pr(>F)
1	1.46	1.46	1.46	0.25
13	1.46	0.11	0.11	0.74
1	1.46	1.46	1.46	0.25

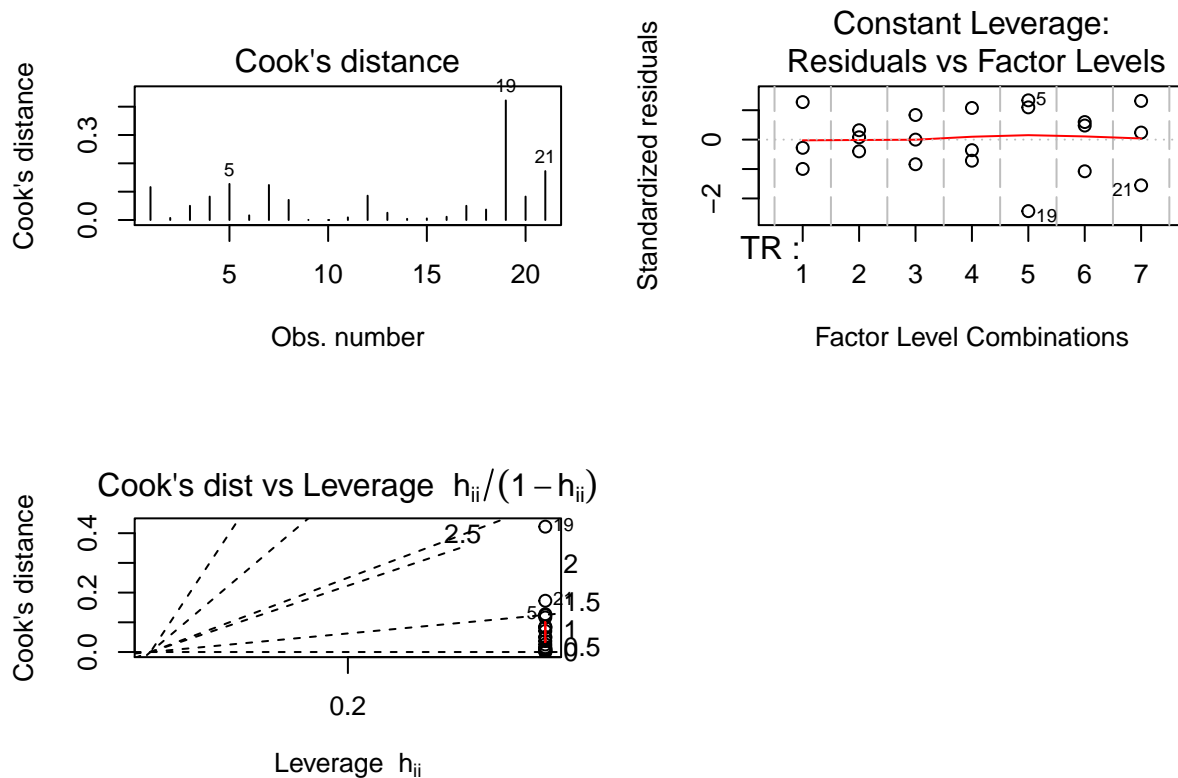
```
## treatment      2 119.77    59.88    1283 1.02e-14 ***
## Residuals     12   0.56     0.05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Q2 a)

```
## [1] "/Users/studentuser/Documents/STAT502/HW"
##           Df Sum Sq Mean Sq F value    Pr(>F)
## TR           6 115792   19299   45.97 2.03e-08 ***
## Residuals    14   5878     420
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

**Histogram of resid\_len**





We reject the null hypothesis at 0.002 level.

Residuals are approx. normally distributed and hence ANOVA assumptions are valid.

b)

- i. These contrasts are orthogonal as the production of any two of them are 0.
- ii. C1: Mean of group 1 vs the mean of rest.  
C2: Mean of groups 2, 3, 4 vs mean of groups 5, 6, 7.  
C3: Mean of groups 2, 5 vs mean of groups 3, 4, 6, 7.  
C4: Mean of groups 3, 6 vs mean of groups 4, 7.

c)

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: User-defined Contrasts
##
##
## Fit: lm(formula = Y ~ TR, data = d.len)
##
## Linear Hypotheses:
##           Estimate Std. Error t value Pr(>|t|)
## constant == 0    144.62      10.95  13.204 2.72e-09 ***
## c1 == 0          1012.33      76.67  13.204 2.72e-09 ***
```

```
## c2 == 0      261.00      28.98      9.007 3.36e-07 ***
## c3 == 0      564.33      47.32     11.926 1.01e-08 ***
## c4 == 0     -19.00      23.66     -0.803   0.435
## c5 == 0       3.00      40.98      0.073   0.943
## c6 == 0      -3.00      23.66     -0.127   0.901
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- none method)
```

$H_{0(1)}, H_{0(2)}, H_{0(3)}$  are rejected. Therefore:

- 1) There exists a difference in the mean of group 1 (the control group) and mean of other treatment groups.
- 2) There exists a difference in the mean of groups 2, 3, 4 vs mean of groups 5, 6, 7. The choice of fertilizer is a factor and can impact harvesting weight.
- 3) There exists a difference in the mean of groups 2, 5 vs the mean of groups 3, 4, 6, 7. The process of weeding is a factor and can impact harvesting weight.

### Q3 a)

Null hypothesis should be all types have the same expected response time, that  $\mu_1 = \mu_2 = \mu_3$ .

```
## Analysis of Variance Table
##
## Response: circ$Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## circ$Type  2  543.6    271.8   16.083 0.0004023 ***
## Residuals 12  202.8     16.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-value is 0.0004023, which is smaller than 0.01. Null hypothesis is rejected, so there is at least one circuit type that is different.

### b)

Compare Type 2 to Type 1 and 3 i.e constrast A

(constrast A):  $-\frac{1}{2}, 1, -\frac{1}{2}$

Compare Type 1 to Type 3 i.e constrast B

(constrast B):  $1, 0, -1$

Multiplying the corresponding coefficients of contrasts A and B:

$$-\frac{1}{2} \times 1 + 1 \times 0 - \frac{1}{2} \times (-1) = 0,$$

Hence A and B are orthogonal.



c)

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: User-defined Contrasts
##
##
## Fit: lm(formula = Y ~ Type, data = circ)
##
## Linear Hypotheses:
##           Estimate Std. Error t value Pr(>|t|)
## constant == 0      3.000      1.501   1.999 0.137667
## c1 == 0           12.600      2.252   5.596 0.000351 ***
## c2 == 0            2.400      2.600   0.923 0.374155
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- holm method)
```

P-value for the first contrast:  $p_{(1)} = 0.000351 < \frac{0.01}{2}$ , and  $p_{(2)} = 0.374155 > 0.01$ .  $j = 2$ .

Conclusion: 1st constrast null hypothesis  $H_{0(1)}$  is rejected, there is difference in the mean between Type 2 and Type1/3. On the other hand, we cannot reject the 2nd constrast null hypothesis  $H_{0(2)}$ , there is no difference in the mean of Type1 and Type3.