# STA207 homework5

*Juanjuan Hu, Zhen Zhang*

*February 18, 2016*

## 26.4

### (a)

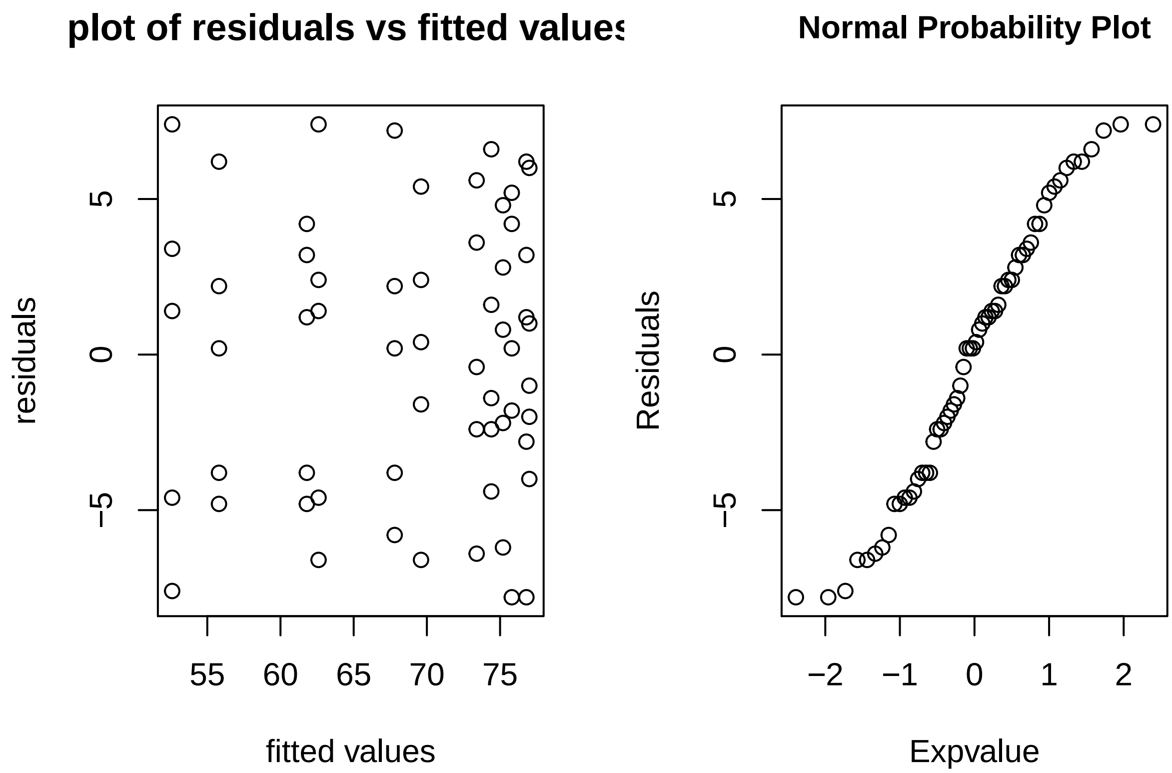
The residuals are:

```
a = data.frame(plant%>%group_by(Machine, Operator)%>%summarise(mean = mean(Cases)))
plant$Cases_bar = rep(a$mean, each = 5)
plant$residual = plant$Cases-plant$Cases_bar
plant
```

```
##    Cases Machine Operator Day Cases_bar residual
## 1     65       1        1   1      61.8      3.2
## 2     58       1        1   2      61.8     -3.8
## 3     63       1        1   3      61.8      1.2
## 4     57       1        1   4      61.8     -4.8
## 5     66       1        1   5      61.8      4.2
## 6     68       1        2   1      67.8      0.2
## 7     62       1        2   2      67.8     -5.8
## 8     75       1        2   3      67.8      7.2
## 9     64       1        2   4      67.8     -3.8
## 10    70       1        2   5      67.8      2.2
## 11    56       1        3   1      62.6     -6.6
## 12    65       1        3   2      62.6      2.4
## 13    58       1        3   3      62.6     -4.6
## 14    70       1        3   4      62.6      7.4
## 15    64       1        3   5      62.6      1.4
## 16    45       1        4   1      52.6     -7.6
## 17    56       1        4   2      52.6      3.4
## 18    54       1        4   3      52.6      1.4
## 19    48       1        4   4      52.6     -4.6
## 20    60       1        4   5      52.6      7.4
## 21    74       2        1   1      75.8     -1.8
## 22    81       2        1   2      75.8      5.2
## 23    76       2        1   3      75.8      0.2
## 24    80       2        1   4      75.8      4.2
## 25    68       2        1   5      75.8     -7.8
## 26    69       2        2   1      75.2     -6.2
## 27    76       2        2   2      75.2      0.8
## 28    80       2        2   3      75.2      4.8
## 29    78       2        2   4      75.2      2.8
## 30    73       2        2   5      75.2     -2.2
## 31    52       2        3   1      55.8     -3.8
## 32    56       2        3   2      55.8      0.2
## 33    62       2        3   3      55.8      6.2
```

```
## 34     58        2         3   4       55.8       2.2
## 35     51        2         3   5       55.8      -4.8
## 36     73        2         4   1       77.0      -4.0
## 37     78        2         4   2       77.0       1.0
## 38     83        2         4   3       77.0       6.0
## 39     75        2         4   4       77.0      -2.0
## 40     76        2         4   5       77.0      -1.0
## 41     69        3         1   1       76.8      -7.8
## 42     83        3         1   2       76.8       6.2
## 43     74        3         1   3       76.8      -2.8
## 44     78        3         1   4       76.8       1.2
## 45     80        3         1   5       76.8       3.2
## 46     63        3         2   1       69.6      -6.6
## 47     70        3         2   2       69.6       0.4
## 48     72        3         2   3       69.6       2.4
## 49     68        3         2   4       69.6      -1.6
## 50     75        3         2   5       69.6       5.4
## 51     81        3         3   1       74.4       6.6
## 52     72        3         3   2       74.4      -2.4
## 53     73        3         3   3       74.4      -1.4
## 54     76        3         3   4       74.4       1.6
## 55     70        3         3   5       74.4      -4.4
## 56     67        3         4   1       73.4      -6.4
## 57     79        3         4   2       73.4       5.6
## 58     73        3         4   3       73.4      -0.4
## 59     77        3         4   4       73.4       3.6
## 60     71        3         4   5       73.4      -2.4
```

```r
par(mfrow=c(1,2))
plot(plant$Cases_bar, plant$residual, xlab = "fitted values",
     ylab = "residuals", main = "plot of residuals vs fitted values")
qqplot = qqnorm(plant$residual, main = "Normal Probability Plot",
                xlab = "Expvalue", ylab = "Residuals", cex.main = 1)
```

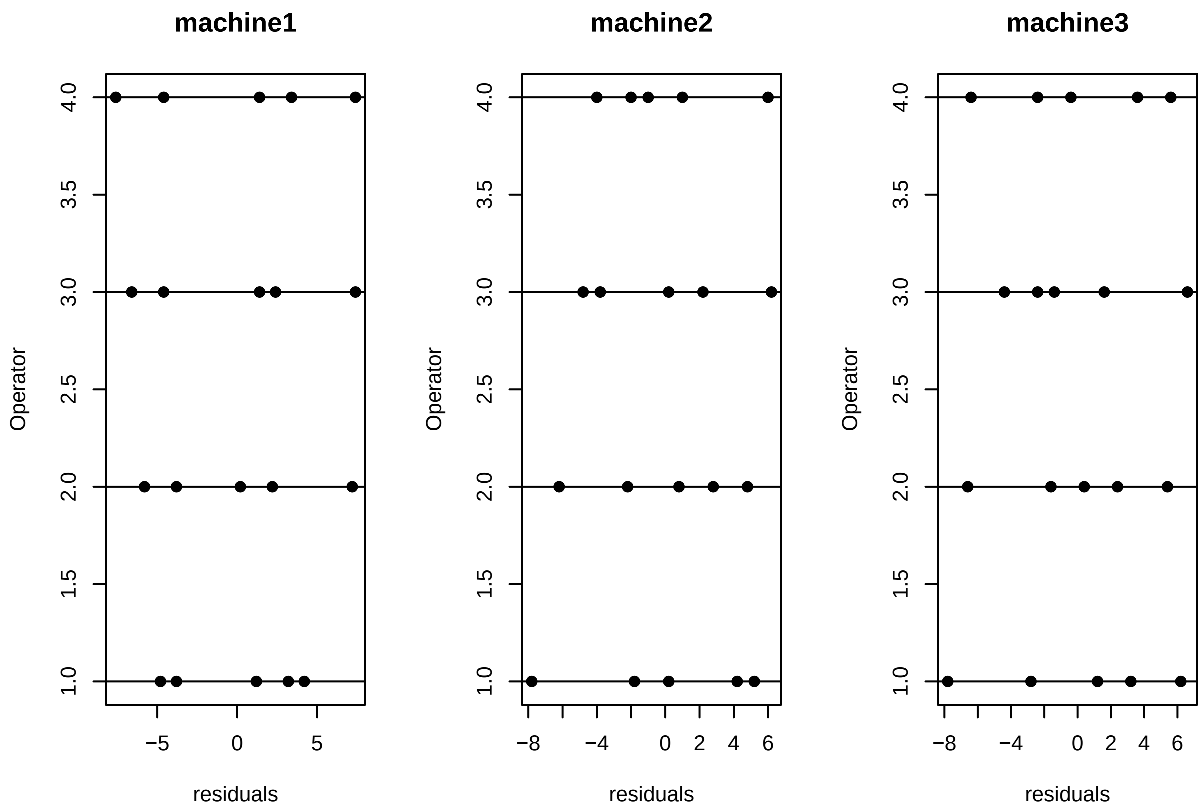



```r
cor(qqplot$x, qqplot$y)
```

```
## [1] 0.9850887
```

There is no obvious pattern in the residuals plot and the the qq plot does not show departure from normatlity. Thus, we consdier the model (26.7) is appropriate.

## (b)



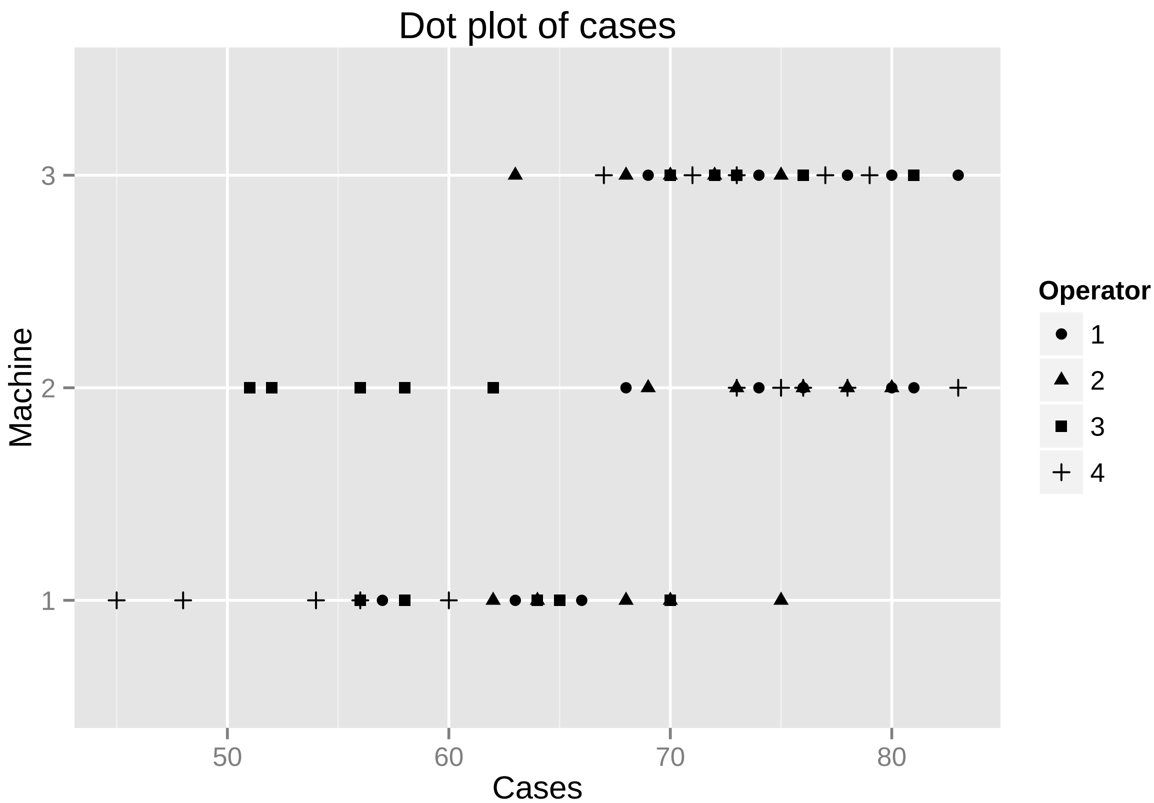

The aligned residual dot plots by machines support the assupmtion of constancy of the error variance.

# 26.5

## (a)

The operator effects cannot be distinguished from the effects of shifts in this study.

## (b)




It seems that for machine 1 and machine 2, the factor effects are present. However, we need further test on this.

## (c)

```r
result = aov(Cases ~ Machine + Machine/Operator, data = plant)
summary(result)
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## Machine           2   1696   847.8   35.92 2.90e-10 ***
## Machine:Operator  9   2272   252.5   10.70 6.99e-09 ***
## Residuals        48   1133    23.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA table is:

```r
kable(anovaTable)
```

| Source | df | SS | MS |
|---|---|---|---|
| Machine | 2 | 1696 | 848.00000 |

| Source | df | SS | MS |
| --- | --- | --- | --- |
| Operator | 9 | 2272 | 252.44444 |
| Error | 48 | 1133 | 23.60417 |
| Total | 59 | 5101 | NA |

### (d)

```
p_val = 1-pf(35.3, 2, 48)
p_val
```

```
## [1] 3.730415e-10
```

$H_0$: all $\alpha_0 = 0$

$H_1$: not all $\alpha_0$ equal to zero

$F^* = \frac{848}{24} = 35.3$

$F(0.99, 2, 48) = 5.075$

The decision rule is: if $F^*$ is greater thatn 5.075, then reject $H_0$, otherwise, accept $H_1$. Here, $35.3 > 4.0517$, so we reject $H_0$, concluding that the mean outputs differ for the three machine models. The p value is almost zero, leading to the same conclusion.

### (e)

```
p_val = 1-pf(10.5, 9, 48)
p_val
```

```
## [1] 9.23519e-09
```

$H_0$: all $\beta_{j(i)} = 0$

$H_1$: not all $\beta_{j(i)} = 0$ equal to zero

$F^* = \frac{252}{24} = 10.5$

$F(0.99, 9, 48) = 2.802$

The decision rule is: if $F^*$ is greater than 2.802, then reject $H_0$, otherwise, accept $H_1$. Here, $10.5 > 2.802$, so we reject $H_0$, concluding that the mean outputs differ for the operators assigned to each machine. The p value is almost zero, leading to the same conclusion.

### (f)

```
b = data.frame(plant%>%group_by(Machine)%>%summarise(mean = mean(Cases)))
plant$Yj_bar_bar = rep(b$mean, each = 20)
plant$operatorEffect = plant$Cases_bar-rep(b$mean, each = 20)
SSB_1 = sum(subset(plant, Machine ==1)$operatorEffect^2)
SSB_2 = sum(subset(plant, Machine ==2)$operatorEffect^2)
SSB_3 = sum(subset(plant, Machine ==3)$operatorEffect^2)
```

From the R output, we know that: $SSB(A_1) = 599$, $SSB(A_2) = 1539$, $SSB(A_3) = 135$ **Test for $\beta_{j(1)}$:** $H_0$: all $\beta_{j(1)} = 0$

$H_1$: not all $\beta_{j(1)} = 0$ equal to zero

$F^* = \frac{(599/3)}{24} = 8.32$

$F(0.99, 3, 48) = 4.22$

The decision rule is: if $F^*$ is greater than 4.22, then reject $H_0$, otherwise, accept $H_1$. Here, $8.32 > 4.22$, so we reject $H_0$, concluding that the mean outputs differ for the operators assigned to the first machine.

**Test for $\beta_{j(2)}$:** $H_0$: all $\beta_{j(2)} = 0$

$H_1$: not all $\beta_{j(2)} = 0$ equal to zero

$F^* = \frac{(1539/3)}{24} = 21.4$

$F(0.99, 3, 48) = 4.22$

The decision rule is: if $F^*$ is greater than 4.22, then reject $H_0$, otherwise, accept $H_1$. Here, $21.4 > 4.22$, so we reject $H_0$, concluding that the mean outputs differ for the operators assigned to the second machine.

**Test for $\beta_{j(3)}$:** $H_0$: all $\beta_{j(3)} = 0$

$H_1$: not all $\beta_{j(3)} = 0$ equal to zero

$F^* = \frac{(135/3)}{24} = 1.88$

$F(0.99, 3, 48) = 4.22$

The decision rule is: if $F^*$ is greater thatn 4.22, then reject $H_0$, otherwise, accept $H_1$. Here, $1.88 < 4.22$, so we cannot reject $H_0$, concluding that the mean outputs do not differ for the operators assigned to the third machine.

### (g)

```r
1-(0.99)^5
```

```
## [1] 0.04900995
```

Therefore, the family level of significante for the combined tests is 0.05. In summary, the mean outputs differ for the machines and for the operators. For each machine seperately, the mean outputs differ for operators assigned to machine 1 and machine2, but not for machine 3.

## 26.6

### (a)

```r
unique(plant$Yj_bar_bar)
```

```
## [1] 61.20 70.95 73.55
```

$\bar{Y}_{1..} = 61.2$, $\bar{Y}_{2..} = 71$, $\bar{Y}_{3..} = 73.5$

$\hat{L}_1 = \bar{Y}_{1..} - \bar{Y}_{2..} = 61.2 - 71 = -9.8$, $\hat{L}_2 = \bar{Y}_{1..} - \bar{Y}_{3..} = 61.2 - 73.5 = -12.3$, $\hat{L}_3 = \bar{Y}_{2..} - \bar{Y}_{3..} = 71 - 73.5 = -2.5$

$s^2(\hat{L}_i) = 2 * s^2(\bar{Y}_{1..}) = 2 * MSE/bn = 2 * 24/(4 * 5) = 2.4$

$s(\hat{L}_i) = 1.55$

$qtukey(0.95, 3, 48) = 3.42$, $T = 3.43/1.414 = 2.42$

$\hat{L}_1 + T * s(\hat{L}_i) = -9.8 + 2.42 * 1.55 = -6.05$, $\hat{L}_1 - T * s(\hat{L}_i) = -9.8 - 2.42 * 1.55 = -13.6$

$\hat{L}_2 + T * s(\hat{L}_i) = -12.3 + 2.42 * 1.55 = -8.55$, $\hat{L}_1 - T * s(\hat{L}_i) = -12.3 - 2.42 * 1.55 = -16.1$

$\hat{L}_3 + T * s(\hat{L}_i) = -2.5 + 2.42 * 1.55 = 1.25$, $\hat{L}_1 - T * s(\hat{L}_i) = -2.5 - 2.42 * 1.55 = -6.25$

Therefore, the confidence intervals for the pairwise comparisons are: [-13.6, -6.05], [-16.1, -8.55], [-6.25, 1.25]. We find that the first two CIs do not contain zero, and the last one contains zero and conclude that mean outputs differ from machine 1 to machine 2, and from machine 1 to machine 3.

## (b)

```r
unique(subset(plant, Machine == 1)$Cases_bar)
```

```
## [1] 61.8 67.8 62.6 52.6
```

$\bar{Y}_{11.} = 61.8$, $\bar{Y}_{12.} = 67.8$, $\bar{Y}_{13.} = 62.5$, $\bar{Y}_{14.} = 52.6$

$\hat{L}_1 = \bar{Y}_{11.} - \bar{Y}_{12.} = 61.8 - 67.8 = -6$, $\hat{L}_2 = \bar{Y}_{11.} - \bar{Y}_{13.} = 61.8 - 62.5 = -0.7$, $\hat{L}_3 = \bar{Y}_{11.} - \bar{Y}_{14.} = 61.8 - 52.6 = 9.2$, $\hat{L}_4 = \bar{Y}_{12.} - \bar{Y}_{13.} = 67.8 - 62.5 = 5.3$, $\hat{L}_5 = \bar{Y}_{12.} - \bar{Y}_{14.} = 67.8 - 52.6 = 15.2$,$\hat{L}_6 = \bar{Y}_{13.} - \bar{Y}_{14.} = 62.5 - 52.6 = 9.9$

$s^2(\hat{L}_i) = 2 * MSE/n = 2 * 24/5 = 9.6$

$s(\hat{L}_i) = 3.1$

$B = t(1 - 0.05/12, 48) = 2.753$

$\hat{L}_1 + B * s(\hat{L}_i) = -6 + 2.753 * 3.1 = 2.53$, $\hat{L}_1 - B * s(\hat{L}_i) = -6 - 2.753 * 3.1 = -14.5$

$\hat{L}_2 + B * s(\hat{L}_i) = -0.7 + 2.753 * 3.1 = 7.83$, $\hat{L}_1 - B * s(\hat{L}_i) = -0.7 - 2.753 * 3.1 = -9.23$

$\hat{L}_3 + B * s(\hat{L}_i) = 9.2 + 2.753 * 3.1 = 17.7$, $\hat{L}_1 - B * s(\hat{L}_i) = 9.2 - 2.753 * 3.1 = 0.67$

$\hat{L}_4 + B * s(\hat{L}_i) = 5.3 + 2.753 * 3.1 = 13.8$, $\hat{L}_1 - B * s(\hat{L}_i) = 5.3 - 2.753 * 3.1 = -3.23$

$\hat{L}_5 + B * s(\hat{L}_i) = 15.2 + 2.753 * 3.1 = 23.7$, $\hat{L}_1 - B * s(\hat{L}_i) = 15.2 - 2.753 * 3.1 = 6.67$

$\hat{L}_6 + B * s(\hat{L}_i) = 9.9 + 2.753 * 3.1 = 18.4$, $\hat{L}_1 - B * s(\hat{L}_i) = 9.9 - 2.753 * 3.1 = 1.37$

Therefore, the pairwise comparisons CIs are: [-14.5, 2.53], [-9.23, 7.83], [0.67, 17.7], [-3.23, 13.8], [6.67, 23.7], [1.37, 18.4]. We conclude that the mean output differ from operator 1 to operator 4, from operator 2 to operator 4, and from operator 3 to operator 4.

## (c)

$\hat{L} = (1/3) * (\bar{Y}_{11.} + \bar{Y}_{12.} + \bar{Y}_{13.}) - \bar{Y}_{14.} = (61.8 + 67.8 + 62.5)/3 - 52.6 = 11.4$

$s^2(\hat{L}) = (3 * (1/3)^2 + 1) * MSE/5 = (3 * (1/3)^2 + 1) * 24/5 = 6.4$, $s(\hat{L}) = 2.53$

$T = t(0.995, 48) = 2.682$

$\hat{L}_6 + T * s(\hat{L}) = 11.4 + 2.682 * 2.53 = 18.2$, $\hat{L}_1 - T * s(\hat{L}) = 11.4 - 2.682 * 2.53 = 4.61$

Therefore, the CI is [4.61, 18.2]. Since the confidence interval does not contain zero, we conclude that that the mean output of operator 4 is significantly less than the average outputs from other three operators.

## 26.19 Plant acid levels

```
plant = read.table('CH26PR19.txt')
names(plant) = c('Y', 'A', 'B', 'K');
plant$A = as.factor(plant$A)
plant$B = as.factor(plant$B)
plant$K = as.factor(plant$K)
(a = length(unique(plant$A)))
```

```
## [1] 4
```

```
(b = length(unique(plant$B)))
```

```
## [1] 3
```

```
(k = length(unique(plant$K)))
```

```
## [1] 3
```

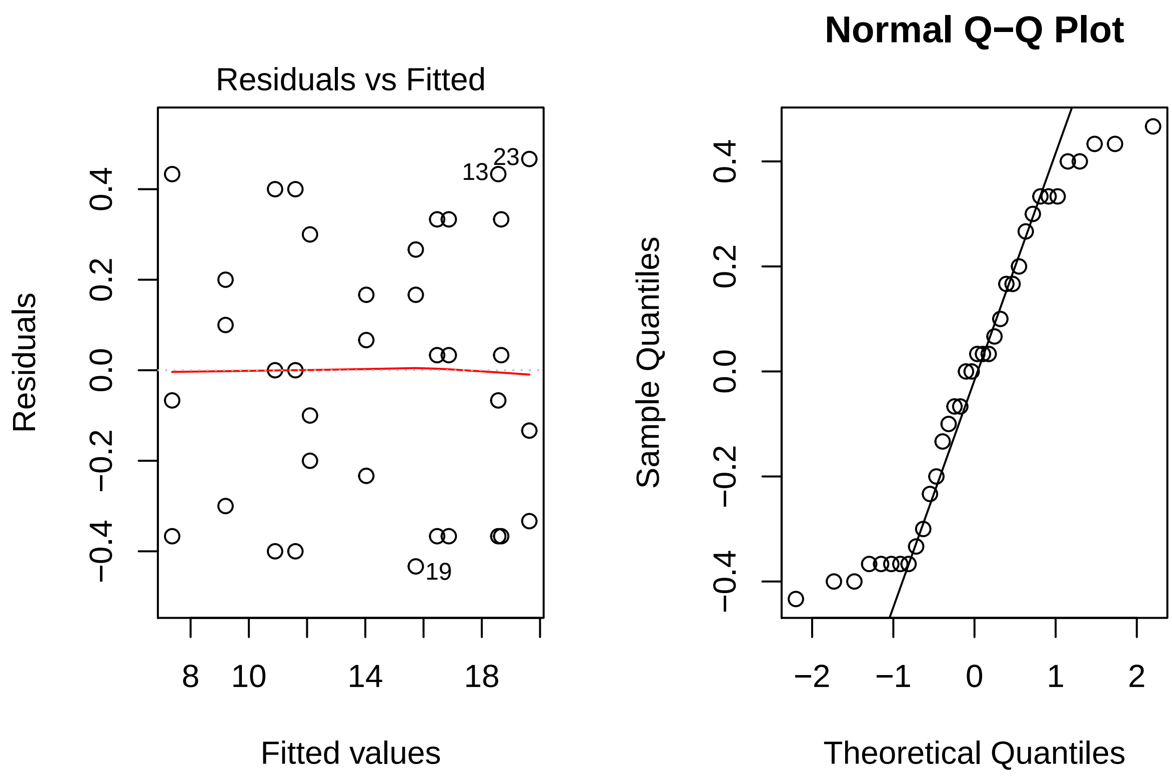
```
fit = aov(Y ~ A/B, data = plant)
(fit.aov = summary(fit))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## A            3  343.2  114.39   905.1 <2e-16 ***
## A:B          8  187.5   23.43   185.4 <2e-16 ***
## Residuals   24    3.0    0.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

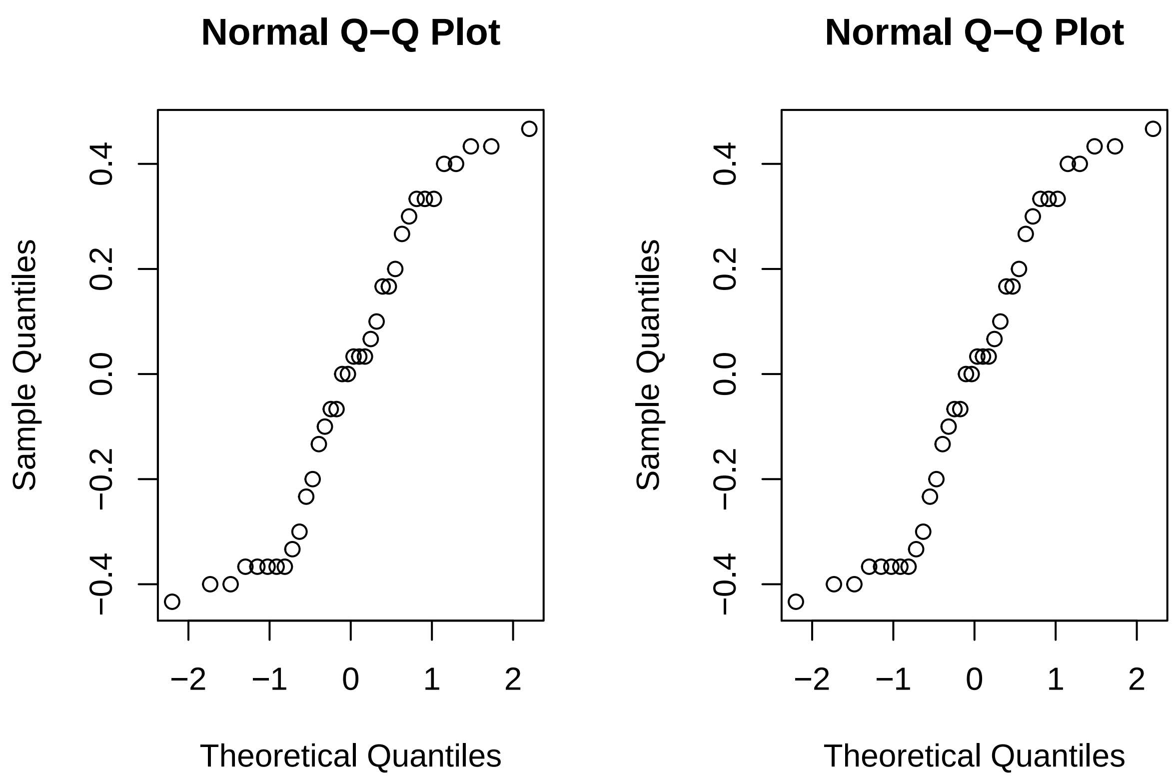
```
(res.fit = residuals(fit))
```

```
##             1             2             3             4             5
## -4.000000e-01  4.586609e-15  4.000000e-01  3.333333e-02  3.333333e-01
##             6             7             8             9            10
## -3.666667e-01 -3.666667e-01  3.333333e-02  3.333333e-01  6.666667e-02
##            11            12            13            14            15
## -2.333333e-01  1.666667e-01  4.333333e-01 -6.666667e-02 -3.666667e-01
##            16            17            18            19            20
## -2.000000e-01  3.000000e-01 -1.000000e-01 -4.333333e-01  1.666667e-01
##            21            22            23            24            25
##  2.666667e-01 -1.333333e-01  4.666667e-01 -3.333333e-01 -3.666667e-01
##            26            27            28            29            30
##  3.333333e-01  3.333333e-02 -6.666667e-02  4.333333e-01 -3.666667e-01
##            31            32            33            34            35
## -3.000000e-01  2.000000e-01  1.000000e-01  4.000000e-01 -4.857226e-17
##            36
## -4.000000e-01
```

```r
par(mfrow=c(1, 2))
# residuals against fitted value plot
plot(fit, which = 1)
# Normal probability plot
qqnorm(res.fit); qqline(res.fit)
```




```r
cor(qqnorm(res.fit)$x, qqnorm(res.fit)$y)
```




```
## [1] 0.9659539
```

This model is appropriate.

## 26.20 Plant acid levels

### (a)

```
fit.aov
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)    
## A            3  343.2  114.39   905.1 <2e-16 ***
## A:B          8  187.5   23.43   185.4 <2e-16 ***
## Residuals   24    3.0    0.13                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### (b)

```r
MSEE = 114.393
MSOE = 23.432
(F_E = MSEE/MSOE)
```

```
## [1] 4.881914
```

```r
qf(0.95, a-1, a*(b-1))
```

```
## [1] 4.066181
```

```r
1-pf(F_E, a-1, a*(b-1))
```

```
## [1] 0.03243786
```

Null hypothesis ($H_0$): There are no variations in mean concentration levels between plants

Alternatives: There are variations in mean concentration levels between plants

Decision Rule: $F^* = \frac{MSEE}{MSOE}$, and reject $H_0$ if $F^* > F_{0.95,a-1,a(b-1)}$

Conclusion: since $F^* = 4.881914 > 4.066181 = F_{0.95,a-1,a(b-1)}$, then reject $H_0$. The p-value is 0.03243786

## (c)

```r
MSE = 0.126
(F_EE = MSOE/MSE)
```

```
## [1] 185.9683
```

```r
qf(0.95, a*(b-1), a*b*(k-1))
```

```
## [1] 2.355081
```

```r
1-pf(F_EE, a*(b-1), a*b*(k-1))
```

```
## [1] 0
```

Null hypothesis ($H_0$): There are no variations in mean concentration levels between leaves of the same plant

Alternatives: There are variations in mean concentration levels between leaves of the same plant

Decision Rule: $F^* = \frac{MSOE}{MSE}$, and reject $H_0$ if $F^* > F_{0.95,a(b-1),ab(c-1)}$

Conclusion: since $F^* = 185.9683 > 2.355081 = F_{0.95,a(b-1),ab(c-1)}$, then reject $H_0$. The p-value is 0.0000

## (d)

```r
(Y_hat = mean(plant$Y))
```

```
## [1] 14.26111
```

```r
(t = qt(0.975, a-1))
```

```
## [1] 3.182446
```

```r
(s = sqrt(MSEE/(a*b*k)))
```

```
## [1] 1.782578
```

```r
c(Y_hat, Y_hat-t*s, Y_hat+t*s)
```

```
## [1] 14.261111  8.588153 19.934069
```

So the confidence interval is (8.588153, 19.934069)

### (e)

```r
(delta = MSE)
```

```
## [1] 0.126
```

```r
(delta_a = (MSEE-MSOE)/(k*b))
```

```
## [1] 10.10678
```

```r
(delta_ba = (MSOE-MSE)/k)
```

```
## [1] 7.768667
```

```r
(delta_Y = delta + delta_a + delta_ba)
```

```
## [1] 18.00144
```

so $\hat{\sigma}_r^2$ seems to be the most important

## 27.6

### (a)

```r
sale = read.table('CH27PR06.txt')
names(sale) = c('y', 'A', 'B')
sale$id <- 1:24
sale$A = as.factor(sale$A)
sale$B = as.factor(sale$B)
(s = nlevels(sale$A))
```

```
## [1] 8
```

```r
(r = nlevels(sale$B))
```

```
## [1] 3
```

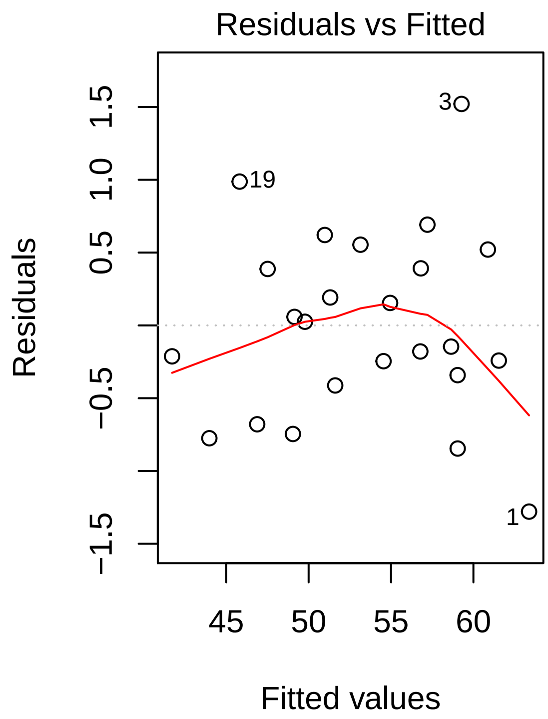
```r
model = aov(y ~ A + B, data = sale)
(model.summary = summary(model))
```

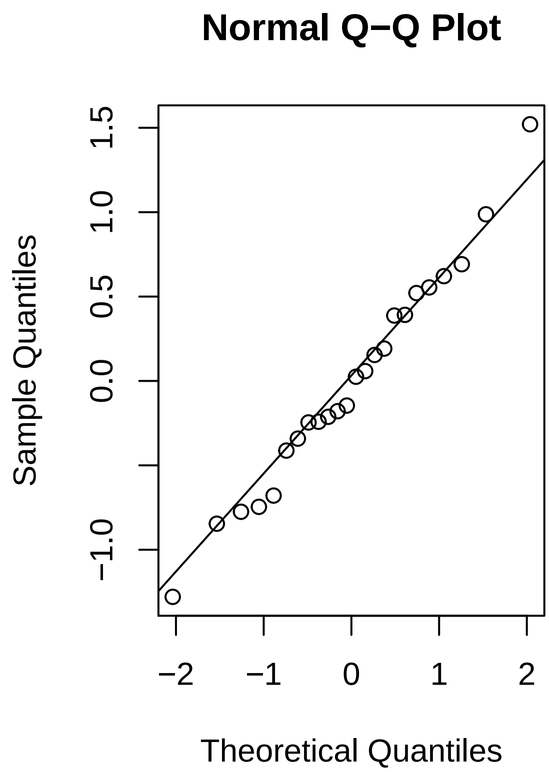
```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## A            7  745.2  106.46  155.69 3.47e-12 ***
## B            2   67.5   33.74   49.35 4.57e-07 ***
## Residuals   14    9.6    0.68                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
(model.resid = residuals(model))
```

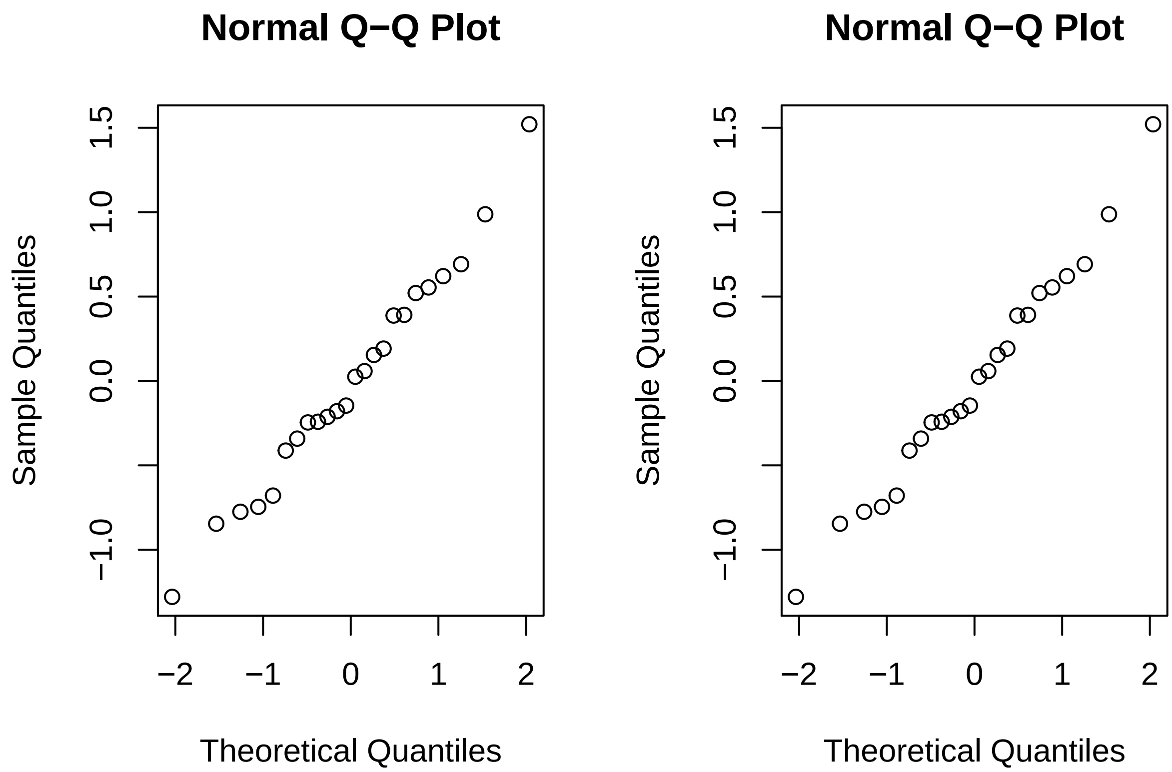
```
##           1           2           3           4           5           6 
## -1.27916667 -0.24166667  1.52083333 -0.84583333  0.69166667  0.15416667 
##           7           8           9          10          11          12 
##  0.62083333  0.05833333 -0.67916667  0.55416667  0.19166667 -0.74583333 
##          13          14          15          16          17          18 
##  0.52083333 -0.34166667 -0.17916667 -0.14583333  0.39166667 -0.24583333 
##          19          20          21          22          23          24 
##  0.98750000 -0.77500000 -0.21250000 -0.41250000  0.02500000  0.38750000
```

```r
par(mfrow=c(1, 2))
# residuals against fitted value plot
plot(model, which = 1)
# Normal probability plot
qqnorm(model.resid); qqline(model.resid)
```




```
cor(qqnorm(model.resid)$x, qqnorm(model.resid)$y)
```

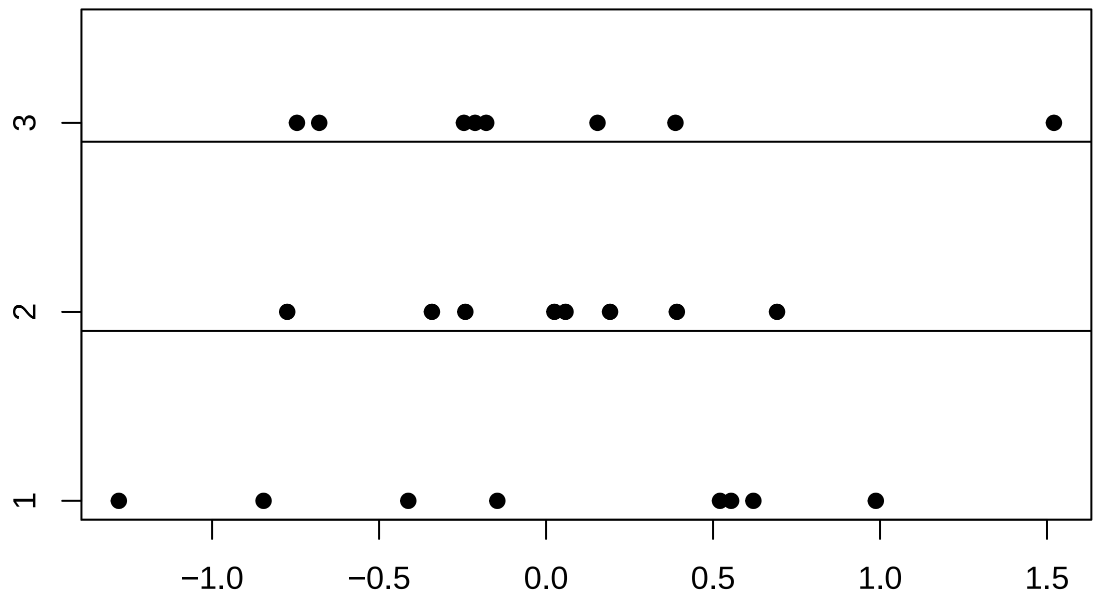



```
## [1] 0.9924551
```

It is appropriate

### (b)

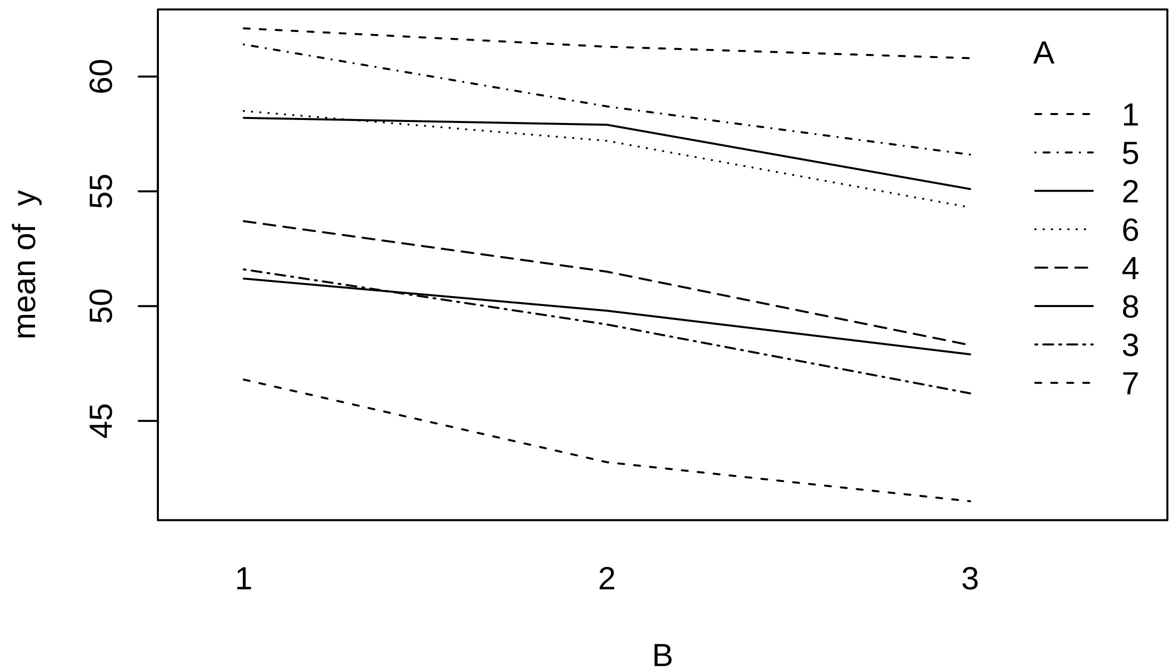
```
stripchart(split(model.resid, sale$B), method = 'stack', pch = 19)
abline(h = seq(1, 3) - 0.1)
```




The assumption of no interactions between subjects and treatments appear to be reasonable here

## (c)

```
with(sale, interaction.plot(B, A, y))
```




### (d)

```r
library('additivityTests')
(additivity_sale = matrix(sale$y, byrow = TRUE, ncol = r))
```

```
##      [,1] [,2] [,3]
## [1,] 62.1 61.3 60.8
## [2,] 58.2 57.9 55.1
## [3,] 51.6 49.2 46.2
## [4,] 53.7 51.5 48.3
## [5,] 61.4 58.7 56.6
## [6,] 58.5 57.2 54.3
## [7,] 46.8 43.2 41.5
## [8,] 51.2 49.8 47.9
```

```r
tukey.test(additivity_sale, 0.01)
```

```
##
## Tukey test on 1% alpha-level:
##
## Test statistic: 5.765
## Critival value: 9.074
## The additivity hypothesis cannot be rejected.
```

```r
1-pf(5.765, 1, 13)
```

```
## [1] 0.03202252
```

Null hypothesis ($H_0$): there is additivity effect

Alternatives: there is no additivity effect

Decision Rule: $F^* = \frac{MSTR.S/1}{SSRem/13}$, and reject $H_0$ if $F^* > F_{0.95,1,13}$

Conclusion: since $F^* = 5.765 < 9.074 = F_{0.95,1,13}$, then cannot reject $H_0$. The p-value is 0.03202252

# 27.7

## (a)

```r
anova(model)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## A          7 745.19 106.455 155.693 3.473e-12 ***
## B          2  67.48  33.740  49.346 4.567e-07 ***
## Residuals 14   9.57   0.684
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## (b)

```r
MSTR = 33.740
MSTR.S = 0.684
(F_star = MSTR/MSTR.S)
```

```
## [1] 49.32749
```

```r
qf(0.95, r-1, (r-1) * (s-1))
```

```
## [1] 3.738892
```

```r
1-pf(F_star, r-1, (r-1) * (s-1))
```

```
## [1] 4.577662e-07
```

Null hypothesis ($H_0$): mean sales of grapefruits do not differ for the three price levels

Alternatives: mean sales of grapefruits differ for the three price levels

Decision Rule: $F^* = \frac{MSTR}{MSE}$, and reject $H_0$ if $F^* > F_{0.95,r-1,(r-1)(s-1)}$

Conclusion: since $F^* = 49.32749 > 3.738892 = F_{0.95,r-1,(r-1)(s-1)}$, then reject $H_0$. The p-value is 0.0000

### (c)

use tukey's method

```r
(mu = with(sale, tapply(y, B, mean)))
```

```
##       1       2       3
## 55.4375 53.6000 51.3375
```

```r
(mu1 = mu[1])
```

```
##       1
## 55.4375
```

```r
(mu2 = mu[2])
```

```
##    2
## 53.6
```

```r
(mu3 = mu[3])
```

```
##       3
## 51.3375
```

```r
(tukey_stat = 1/sqrt(2) * qtukey(0.95, 3, 14))
```

```
## [1] 2.61728
```

```r
(l1 = mu1-mu2)
```

```
##      1
## 1.8375
```

```r
(l2 = mu1-mu3)
```

```
##   1
## 4.1
```

```r
(l3 = mu2-mu3)
```

```
##      2
## 2.2625
```

```r
(sd_tukey = sqrt(MSTR.S * 2 / s))
```

```
## [1] 0.4135215
```

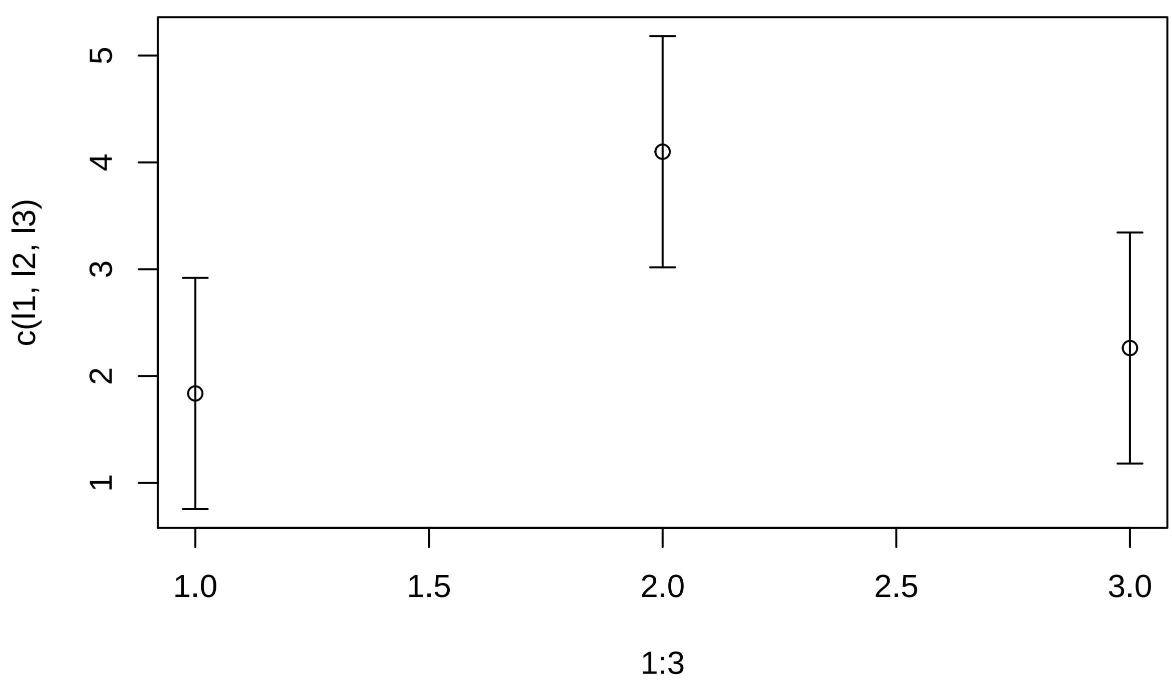
So the confidence intervals are:

$L_1 : 1.8375 \pm 0.4135215 * 2.61728 = (0.756, 2.919)$

$L_2 : 4.1 \pm 0.4135215 * 2.61728 = (3.018, 5.182)$

$L_3 : 2.2625 \pm 0.4135215 * 2.61728 = (1.181, 3.344)$

```
library(plotrix)
plotCI(1:3, y = c(l1, l2, l3), li = c(0.756, 3.018, 1.181), ui = c(2.919, 5.182, 3.344))
```




## (d)

```
MSS = 106.46
(E = ((s - 1) * MSS + s * (r - 1) * MSTR.S)/((s * r - 1) * MSTR.S))
```

```
## [1] 48.06534
```