Assignment 9

Sachin Shubham

4/1/2021

Task 1

A manufacturer of air conditioning ducts is concerned about the tensile strength of the sheet metal among the many supplies of this material. Four samples of sheet metal from four randomly chosen supplies are tested for tensile strength. The data are given.

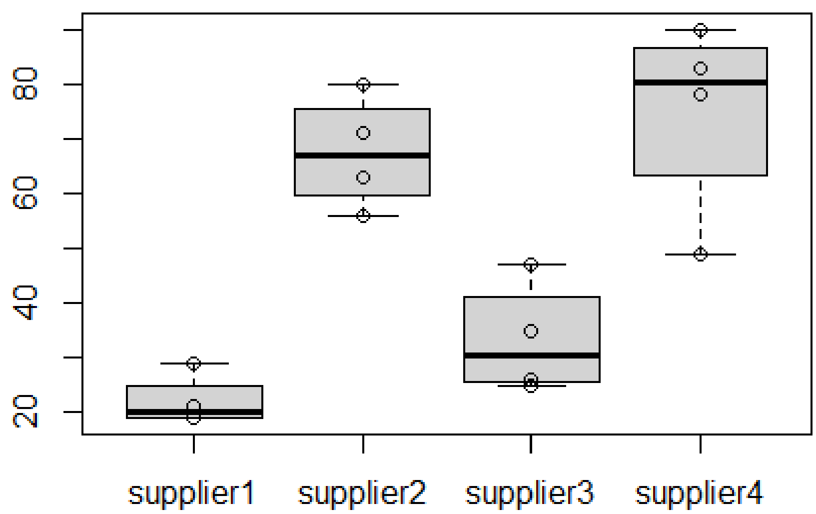
(a) Perform the appropriate analysis to ascertain whether there is excessive variation among suppliers.
(b) Estimate the appropriate variance components.

Solution:

```
library(lme4)

## Warning: package 'lme4' was built under R version 4.0.4

## Loading required package: Matrix

library(lattice)
library(nlme)

##
## Attaching package: 'nlme'

## The following object is masked from 'package:lme4':
##
##     lmList

supplier<-
c("supplier1","supplier1","supplier1","supplier1","supplier2","supplier2",

"supplier2","supplier2","supplier3","supplier3","supplier3","supplier3",
           "supplier4","supplier4","supplier4","supplier4")

strength<-c(19,21,19,29,80,71,63,56,47,26,25,35,90,49,83,78)

df_task1<-data.frame(supplier,strength)
df_task1

##     supplier strength
## 1  supplier1       19
## 2  supplier1       21
## 3  supplier1       19
```

```
## 4   supplier1       29
## 5   supplier2       80
## 6   supplier2       71
## 7   supplier2       63
## 8   supplier2       56
## 9   supplier3       47
## 10  supplier3       26
## 11  supplier3       25
## 12  supplier3       35
## 13  supplier4       90
## 14  supplier4       49
## 15  supplier4       83
## 16  supplier4       78
```

```
supplier_factor<-as.factor(df_task1$supplier)
boxplot(split(df_task1$strength,supplier_factor))
points(df_task1)
```




```
#(a)
aov1 <- lm(strength ~ supplier,data=df_task1)
anova(aov1)
```

```
## Analysis of Variance Table
##
## Response: strength
##           Df Sum Sq Mean Sq F value    Pr(>F)
## supplier   3 7978.2 2659.40  19.044 7.401e-05 ***
```

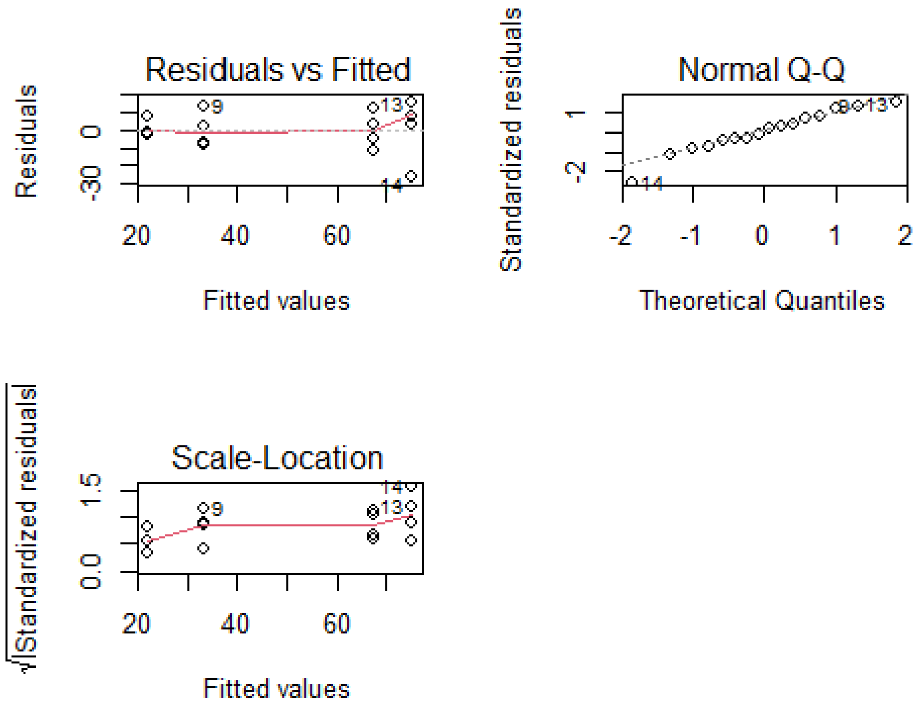
```
## Residuals 12 1675.8  139.65
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov1)
```

```
##
## Call:
## lm(formula = strength ~ supplier, data = df_task1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.000  -5.188   0.375   7.250  15.000
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)           22.000      5.909   3.723 0.002909 **
## suppliersupplier2     45.500      8.356   5.445 0.000149 ***
## suppliersupplier3     11.250      8.356   1.346 0.203075
## suppliersupplier4     53.000      8.356   6.343  3.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.82 on 12 degrees of freedom
## Multiple R-squared:  0.8264, Adjusted R-squared:  0.783
## F-statistic: 19.04 on 3 and 12 DF,  p-value: 7.401e-05
```

```
par(mfrow=c(2, 2))
plot(aov1)
```

```
## hat values (leverages) are all = 0.25
##  and there are no factor predictors; no plot no. 5
```




```r
#(b)
lme2 <- lmer(strength ~ 1 + (1|supplier),data=df_task1)
summary(lme2)

## Linear mixed model fit by REML ['lmerMod']
## Formula: strength ~ 1 + (1 | supplier)
##    Data: df_task1
##
## REML criterion at convergence: 128.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.08660 -0.45320 -0.06519  0.55047  1.38293
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  supplier (Intercept) 629.9    25.10
##  Residual             139.6    11.82
## Number of obs: 16, groups:  supplier, 4
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)    49.44      12.89   3.835
```

(a) Our study find that the F-statistic is 19.04 and supplier1, supplier2 and supplier4 looks very significant with p-value is quite less than 0.05

(b) The estimated variance components are 629.9 and 139.6

Task 2

A local bank has three branch offices. The bank has a liberal sick leave policy, and a vice-president was concerned about employees taking advantage of this policy. She thought that the tendency to take advantage depended on the branch at which the employee took for sick leave, she asked each branch manager to sample employees randomly and record the number of days of sick leave during 2008. Ten employees were chosen, and the data are listed

(a) Do the data indicate a difference in branches? Use a level of significance of 0.05.

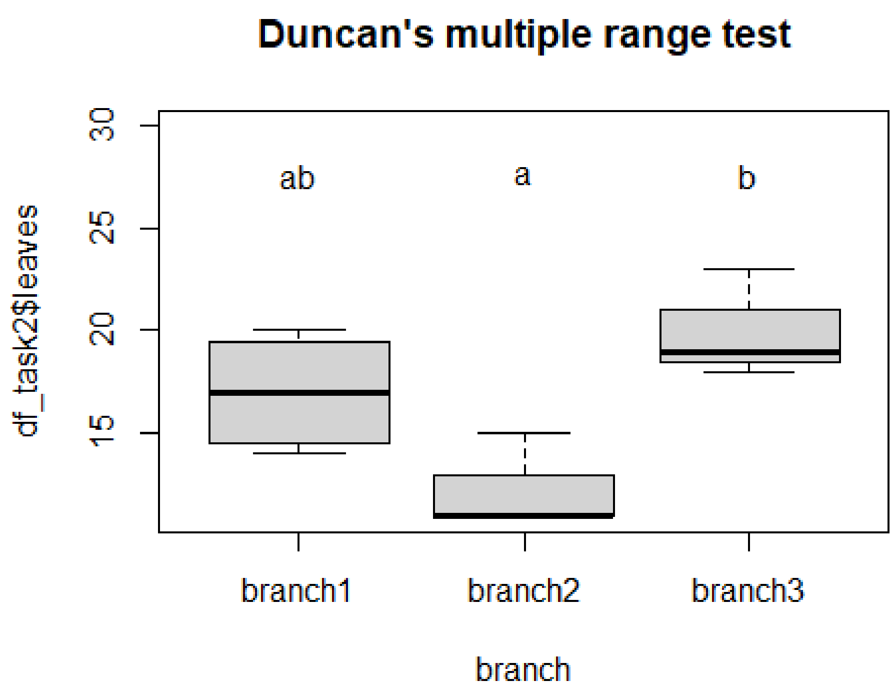
(b) Use Duncan's multi-range test to determine which branches differ. Explain your results with a summary plot

Solution:

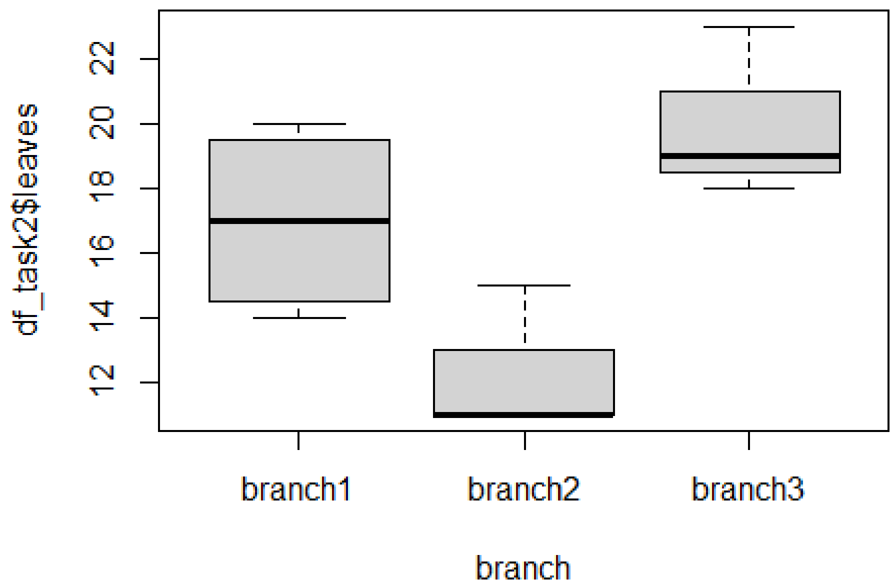
```
library(PMCMRplus)

## Warning: package 'PMCMRplus' was built under R version 4.0.4

leaves<-c(15,20,19,14,11,15,11,18,19,23)
branch<-
c("branch1","branch1","branch1","branch1","branch2","branch2","branch2","bran
ch3","branch3","branch3")
df_task2<-data.frame(branch,leaves)
df_task2

##     branch leaves
## 1  branch1     15
## 2  branch1     20
## 3  branch1     19
## 4  branch1     14
## 5  branch2     11
## 6  branch2     15
## 7  branch2     11
## 8  branch3     18
## 9  branch3     19
## 10 branch3     23

branch <- as.factor(df_task2$branch)
#Model
fit<-aov(df_task2$leaves~branch)
#Anova
anova(fit)

## Analysis of Variance Table
##
## Response: df_task2$leaves
##           Df Sum Sq Mean Sq F value  Pr(>F)
## branch     2 89.833  44.917  6.2056 0.02816 *
## Residuals  7 50.667   7.238
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#Duncan Test
res <- duncanTest(fit)
#Plot
plot(res)
```




```r
boxplot((df_task2$leaves~branch))
```




```
summary(res)

##
##  Pairwise comparisons using Duncan's multiple range test

## data: df_task2$leaves by branch

## alternative hypothesis: two.sided

## P value adjustment method: duncan

## H0

##                           q value  Pr(>|q|)
## branch2 - branch1 == 0   3.138 0.0619793  .
## branch3 - branch1 == 0   2.017 0.1967854
## branch3 - branch2 == 0   5.155 0.0099042 **

## ---

## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summaryGroup(res)

##
##  Pairwise comparisons using Duncan's multiple range test

## data: df_task2$leaves by branch
```

```
## alternative hypothesis: two.sided

## P value adjustment method: duncan

## Different letters indicate significant differences Pr(>|q|) < 0.05

##            mean    sd n Sig. group
## branch1 17.000 2.944 4         ab
## branch2 12.333 2.309 3          a
## branch3 20.000 2.646 3          b
```

(a) Our study find that F = 6.2056, p=-value = 0.02816 (<0.05). Yes the data indicates a significant difference of branches.

(b) As per Duncan's multi-range test branch3 - branch2 looks significantly different (p value = 0.0099), on other hand branch2 - branch1 and branch3 - branch1 with p value 0.0619793 and 0.1967854 respectively, are not significant.

Task 3

Brunye et al. (2008) examined the accuracy with which people could understand spatial representations from descriptions that were presented either in surveyperspective form or in route-perspective form. They used regression to examine whether the time spent reading the description (in seconds) would predict the response time (in milliseconds) to questions about the description. They state: There was strong evidence that increases in route description reading times predicted… response times [$\beta = -.03$, $t(18) = -2.11$, $p < .05$] Note that the degrees of freedom for the t test statistic are shown within parentheses. (a) What is the implication of the negative slope? (b) Give a 95% confidence interval for the expected change in response time, if reading time increases by 20 seconds. (c) Calculate r^2. Is reading time an accurate predictor of individual response times?

Solution:

(a) More time spent reading description led to lower response time to questions Less time spent reading description led to longer response time to questions.

(b) Since, the t-statistic is the ratio of the departure of the estimated value of a parameter from its hypothesized value to its standard error.

SE = Beta/t-stat

SE = -0.03/-2.11

therefore standard error is =0.0142

$20 * (-0.03 \pm 2.11 * 0.0142) = -1.196$ to -0.034

Therefore, With 95% confidence, if reading time increases by 20 seconds, the mean response time will be decrease between 0.034 ms and 1.196 ms

(c) As, t^2=F

```r
F = (-2.11)^2
df = 20-2
r2 = F/(df+F)
r2

## [1] 0.1982933
```

No, reading time only accounts for 19.82% of the variability in the individual response time

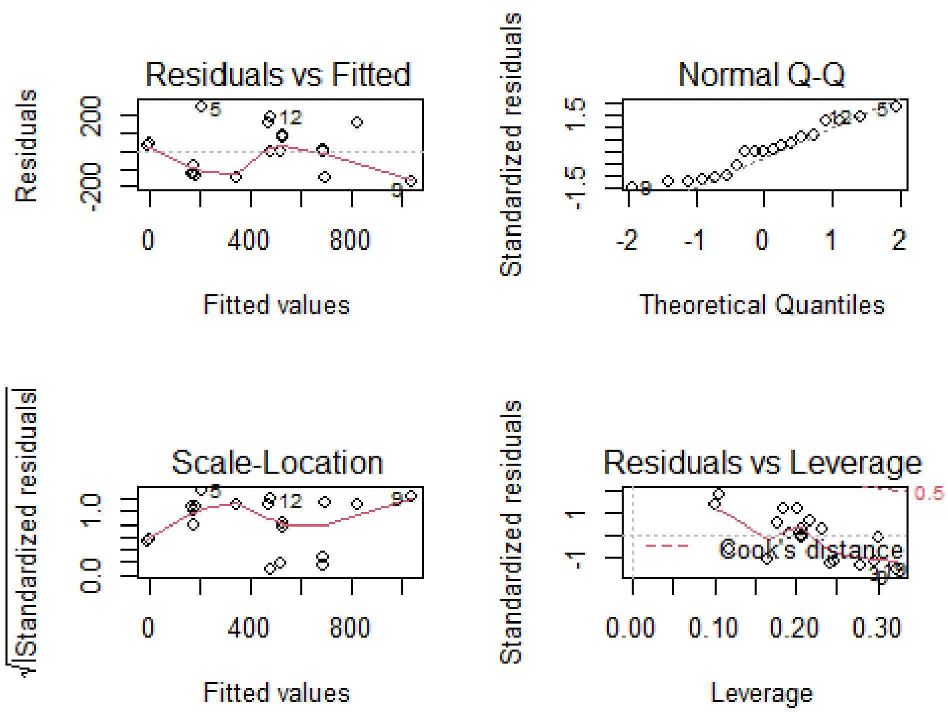
Task 4

The data in Table 3 represent the result of a test for the strength of an asphalt concrete mix. The test consisted of applying a compressive force on the top of different sample specimens. Two responses occurred: the stress and strain at which a sample specimen failed. The factors relate to mixture proportions, rates of speed at which the force was applied, and ambient temperature. Higher values of the response variables indicate stronger materials. The variables are: X1: percent binder (the amount of asphalt in the mixture) X2: loading rate (the speed at which the force was applied) X3: ambient temperature Y1: the stress at which the sample specimen failed Y2: the strain at which the specimen failed. Perform separate regressions to relate stress and strain to the factors of the experiment. Check the residuals for possible specification errors. Interpret all results.

Solution:

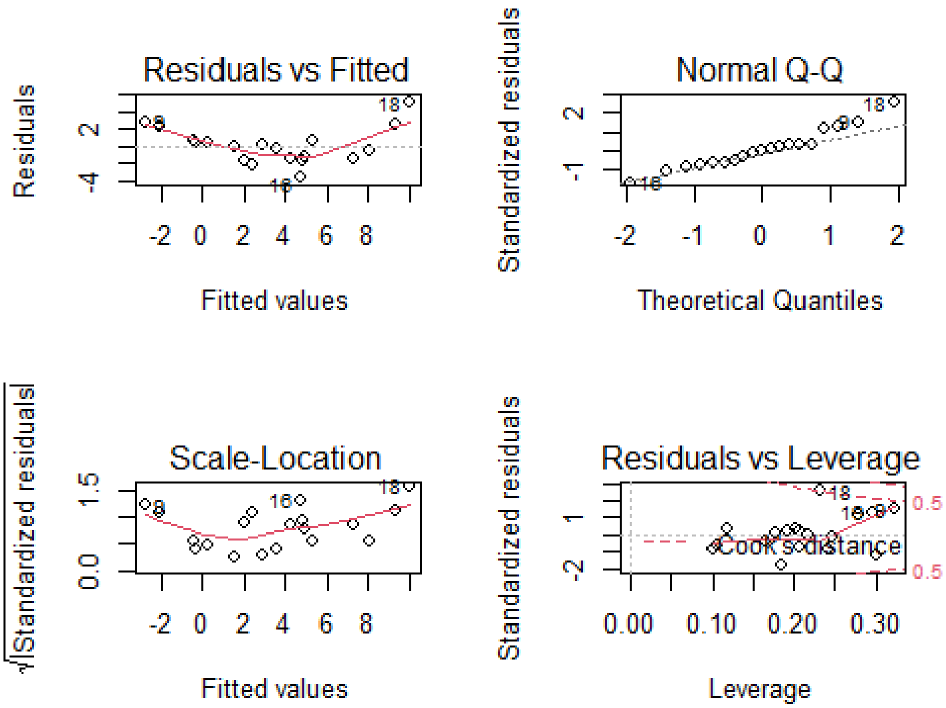
```r
x1 = c(5.3, 5.3, 5.3, 6.0, 7.8, 8.0, 8.0, 8.0, 8.0, 8.0, 8.0, 8.0, 8.0,
       10.0, 12.0, 12.0, 12.0, 12.0, 14.0)
x2 = c(0.02, 0.02, 0.02, 2.00, 0.20, 2.00, 2.00, 2.00, 2.00, 0.02, 0.02,
0.02, 0.02, 2.00, 0.02, 0.02, 0.02, 0.02, 0.02)
x3 = c(77, 32, 0, 77, 77, 104, 77, 32, 0, 104, 77, 32, 0, 77, 77, 32, 0, 104,
77)
Y1 = c(42, 481, 543, 609, 444, 194, 593, 977, 872, 35, 96, 663, 702, 518, 40,
627, 683, 22, 35)
Y2 = c(3.20, 0.73, 0.16, 1.44, 3.68, 3.11, 3.07, 0.19, 0.00, 5.86, 5.97,
0.29, 0.04, 2.72, 7.35, 1.17, 0.14, 15.00, 11.80)

df_task4 <- data.frame(Y1, Y2, x1, x2, x3)
df_task4

##     Y1    Y2   x1   x2  x3
## 1   42  3.20  5.3 0.02  77
## 2  481  0.73  5.3 0.02  32
## 3  543  0.16  5.3 0.02   0
## 4  609  1.44  6.0 2.00  77
## 5  444  3.68  7.8 0.20  77
## 6  194  3.11  8.0 2.00 104
## 7  593  3.07  8.0 2.00  77
## 8  977  0.19  8.0 2.00  32
## 9  872  0.00  8.0 2.00   0
## 10  35  5.86  8.0 0.02 104
## 11  96  5.97  8.0 0.02  77
```

```
## 12 663  0.29  8.0 0.02  32
## 13 702  0.04  8.0 0.02   0
## 14 518  2.72 10.0 2.00  77
## 15  40  7.35 12.0 0.02  77
## 16 627  1.17 12.0 0.02  32
## 17 683  0.14 12.0 0.02   0
## 18  22 15.00 12.0 0.02 104
## 19  35 11.80 14.0 0.02  77

#Model for Y1
model1=lm(Y1~x1+x2+x3, data=df_task4)
summary(model1)

##
## Call:
## lm(formula = Y1 ~ x1 + x2 + x3, data = df_task4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -168.380 -131.124   -0.743   74.773  235.765
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 700.6180   125.8722   5.566 5.40e-05 ***
## x1           -1.5257    13.0242  -0.117 0.908302
## x2          175.9839    35.6550   4.936 0.000179 ***
## x3           -6.6971     0.8847  -7.570 1.69e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 137.9 on 15 degrees of freedom
## Multiple R-squared:  0.8376, Adjusted R-squared:  0.8051
## F-statistic: 25.79 on 3 and 15 DF,  p-value: 3.599e-06

#Plot for Y1
par(mfrow=c(2, 2))
plot(model1)
```

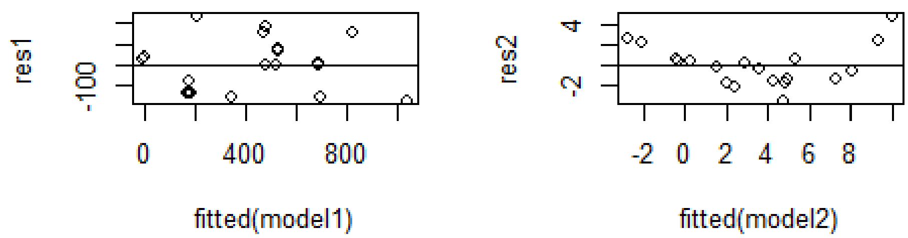



```r
#Model for Y2
model2=lm(Y2~x1+x2+x3, data=df_task4)
summary(model2)

##
## Call:
## lm(formula = Y2 ~ x1 + x2 + x3, data = df_task4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5466 -1.4827 -0.1190  0.6097  5.0135
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.61130    2.04575  -2.743 0.015100 *
## x1           0.66754    0.21168   3.154 0.006558 **
## x2          -1.23535    0.57949  -2.132 0.049966 *
## x3           0.07319    0.01438   5.090 0.000133 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.241 on 15 degrees of freedom
## Multiple R-squared:  0.7601, Adjusted R-squared:  0.7121
## F-statistic: 15.84 on 3 and 15 DF,  p-value: 6.447e-05
```

```r
#Plot for Y2
par(mfrow=c(2, 2))
plot(model2)
```




```r
#Residuals for Model1 and Model2
res1 = resid(model1)
res2 = resid(model2)
plot(fitted(model1), res1)
abline(0,0)
plot(fitted(model2), res2)
abline(0,0)
```



As per our study p-value: 6.447e-05(<<0.05),Multiple R-squared: 0.7601, Adjusted F-statistic: 15.84. and as per residual plot, we can conclude that it's multiplicative model i.e. at least one of the variables are significant in changing the outcomes of the maximum stress and strain of the material.

Task 5

Martinussen et al. (2007) studied burnout among Norwegian policemen. In a sample of n = 220, they regressed y = frequency of psychosomatic complaints on demographic variables gender (0 = man, 1 = woman) and age (m = 2). This regression has R2 = 0.05. They then added independent variables exhaustion burnout score, cynicism burnout score, and professional efficacy burnout score (m = 5). The regression had R2 = 0.34. Given that TSS = 33.7, is there significant evidence that at least one of the burnout scores is related to psychosomatic complaints, after controlling for gender and age? Use α = 0.05.

Solution:

```
# For all model and variables
n = 220
m_full=5
r2_full = 0.34
TSS = 33.7
SSR_full = r2_full*TSS
SSR_full

## [1] 11.458
```

```
MSE_full = 22.242/(n-m_full-1)
MSE_full

## [1] 0.1039346

#For the Age and gender variables only,
r2_rest = 0.05
m_rest = 2
TSS = 33.7
SSR = r2_rest*TSS
SSR

## [1] 1.685

SSE = 33.7 - 1.685
SSE

## [1] 32.015

F_dist_calc = ((r2_full - r2_rest)/(1-r2_full)) * (n-m_full-1)/(m_full-
m_rest)

F_dist_calc

## [1] 31.34343
```

As F_dist_calc > F(3,214) (31.34343 > 2.63), therefore there is a significant evidence that at least one of the burnout scores is related to psychosomatic complaints.