

Ch-04 R Codes

Ping-Yang Chen

2024-03-22

Textbook: Montgomery, D. C. (2012). *Design and analysis of experiments*, 8th Edition. John Wiley & Sons.

Online handouts: https://github.com/PingYangChen/ANOVA_Course_R_Code

Chapter 4

Randomized Complete Block Designs

Read the csv file `4_VascularGraft.csv` in R. Make sure that in the `data.frame` the variables `ExtPressure` and `ResinBatchNo` are the type of factor. If not sure, apply `as.factor()` on those variables after reading the dataset.

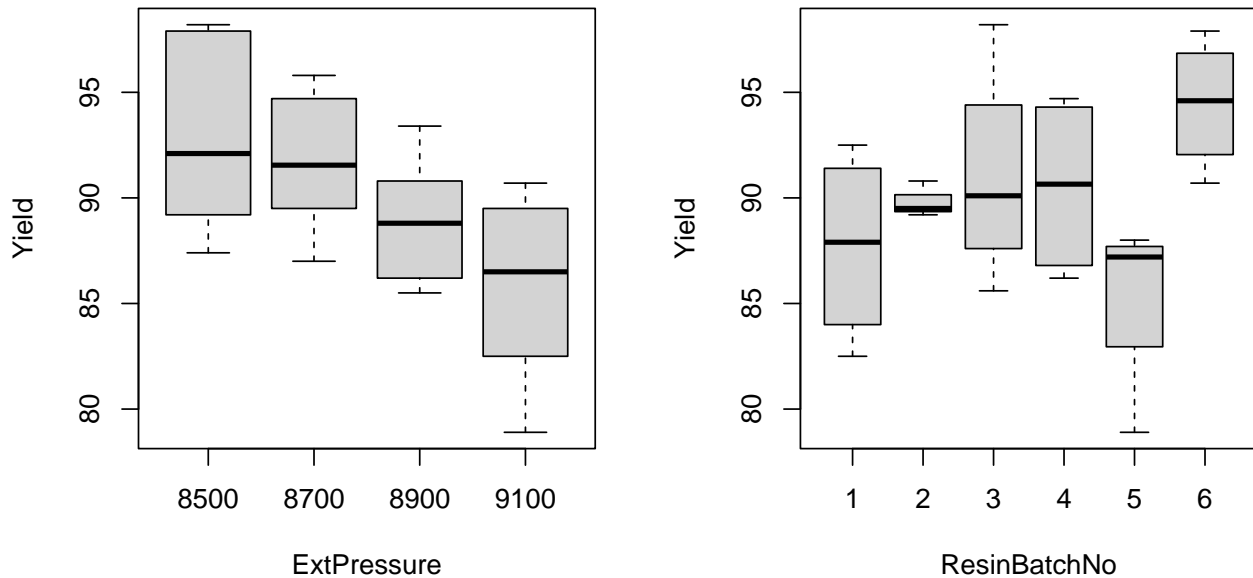
```
df1 <- read.csv(file.path("data", "4_VascularGraft.csv"))
df1$ExtPressure <- as.factor(df1$ExtPressure)
df1$ResinBatchNo <- as.factor(df1$ResinBatchNo)
```

```
tapply(df1$Yield, df1$ExtPressure, summary)
```

```
## $'8500'
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      87.40  89.47   92.10   92.82   96.90   98.20
##
## $'8700'
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      87.00  89.78   91.55   91.68   94.15   95.80
##
## $'8900'
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      85.50  86.65   88.80   88.92   90.50   93.40
##
## $'9100'
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      78.90  83.28   86.50   85.77   88.97   90.70
```

Use boxplots to observe the differences of `Yield` among four treatments of `ExtPressure`, and, the differences of `Yield` among six resin batches `ResinBatchNo`. The average `Yield` tends to be lower for higher `ExtPressure` and varies in different `ResinBatchNo`. Thus, to investigate the `ExtPressure` effect on the `Yield`, we must eliminate the effect caused by `ResinBatchNo`. That is, we treat `ResinBatchNo` as the block effect.

```
# Draw the grouped boxplot
par(mfrow = c(1, 2))
boxplot(Yield ~ ExtPressure, data = df1)
boxplot(Yield ~ ResinBatchNo, data = df1)
```



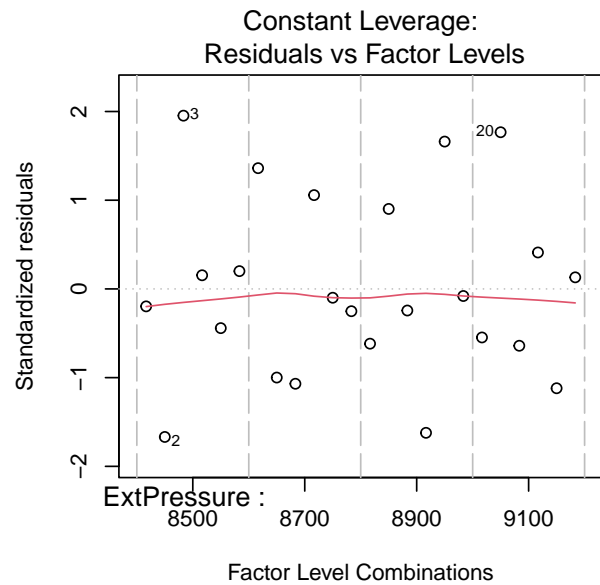
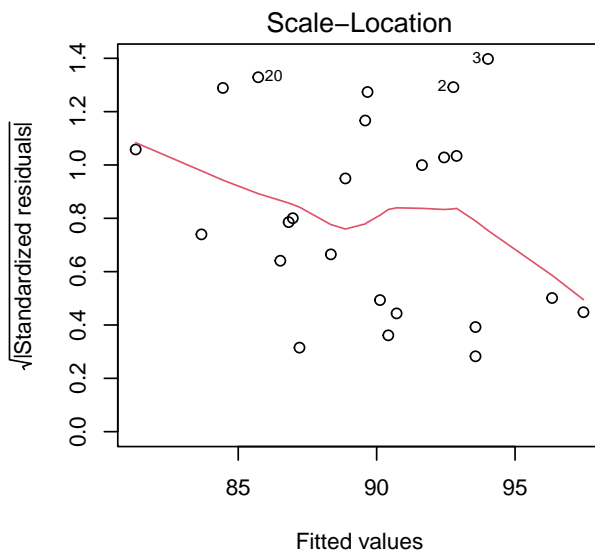
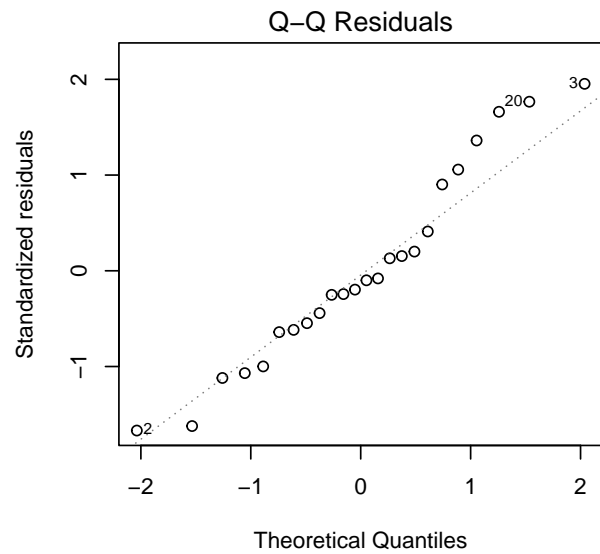
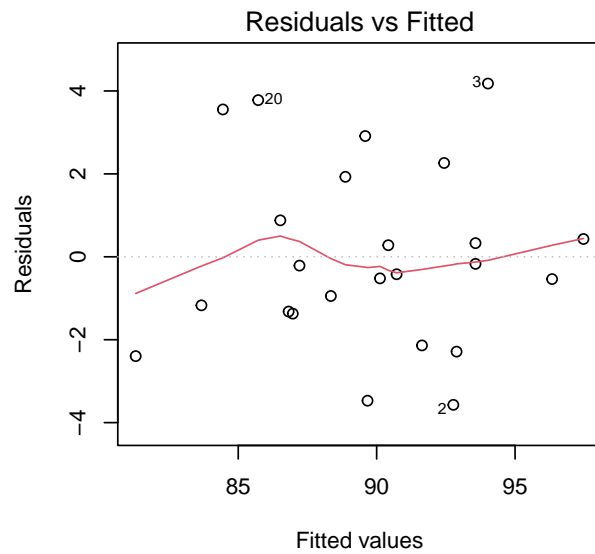
The function `aov()` fits the ANOVA model, and the ANOVA table is obtained by calling `summary()`. For Randomized Complete Block Design (RCBD), the R model formula includes two items on its right-hand-side: `ExtPressure + ResinBatchNo`. It's important to note that the order of placing the factor and block effects is somewhat arbitrary for RCBD. However, this order matters if the design is incomplete, as some sums of squares are adjusted accordingly.

```
fit1 <- aov(Yield ~ ExtPressure + ResinBatchNo, data = df1)
summary(fit1)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## ExtPressure   3  178.2    59.39   8.107 0.00192 **
## ResinBatchNo  5  192.2    38.45   5.249 0.00553 **
## Residuals    15   109.9     7.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

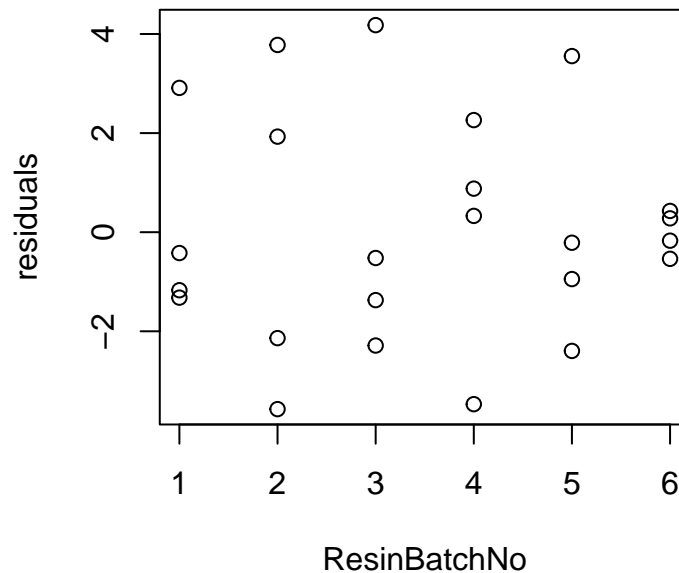
The procedure of diagnosing the residual is similar to that for the one-way ANOVA model. Please refer to the handout of R codes in Chapter 3 for more details of interpreting the residual plots.

```
par(mfrow = c(2, 2))
plot(fit1)
```



One additional plot is to draw the scatter plot of the residual against the levels of the block. A lack of any visually obvious pattern in the dots on the plot is desired.

```
plot(
  as.numeric(df1$ResinBatchNo), fit1$residuals,
  xlab = "ResinBatchNo", ylab = "residuals"
)
```



Multiple comparison is performed for the treatment effect. The following codes demonstrate the use of Tukey's test and Fisher's LSD method. For Tukey's test, add the input argument `which = "ExtPressure"` to only show the test results of comparing differences among the `ExtPressure` levels. For Fisher's LSD method, specify `trt = "ExtPressure"` as the input argument to the `LSD.test()` function to show the comparison results among the `ExtPressure` levels. For information of interpreting the results, please refer to the handout of R codes in Chapter 3.

```
TukeyHSD(fit1, which = "ExtPressure")
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Yield ~ ExtPressure + ResinBatchNo, data = df1)
##
## $ExtPressure
##          diff          lwr          upr      p adj
## 8700-8500 -1.133333 -5.637161  3.370495 0.8854831
## 8900-8500 -3.900000 -8.403828  0.603828 0.1013084
## 9100-8500 -7.050000 -11.553828 -2.546172 0.0020883
## 8900-8700 -2.766667 -7.270495  1.737161 0.3245644
## 9100-8700 -5.916667 -10.420495 -1.412839 0.0086667
## 9100-8900 -3.150000 -7.653828  1.353828 0.2257674
```

```
if (!("agricolae" %in% rownames(installed.packages()))) {
  install.packages("agricolae")
}
library(agricolae)
out <- LSD.test(fit1, trt = "ExtPressure", p.adj = "bonferroni")
out$group
```

```
##          Yield groups
## 8500 92.81667      a
## 8700 91.68333      a
## 8900 88.91667     ab
## 9100 85.76667      b
```

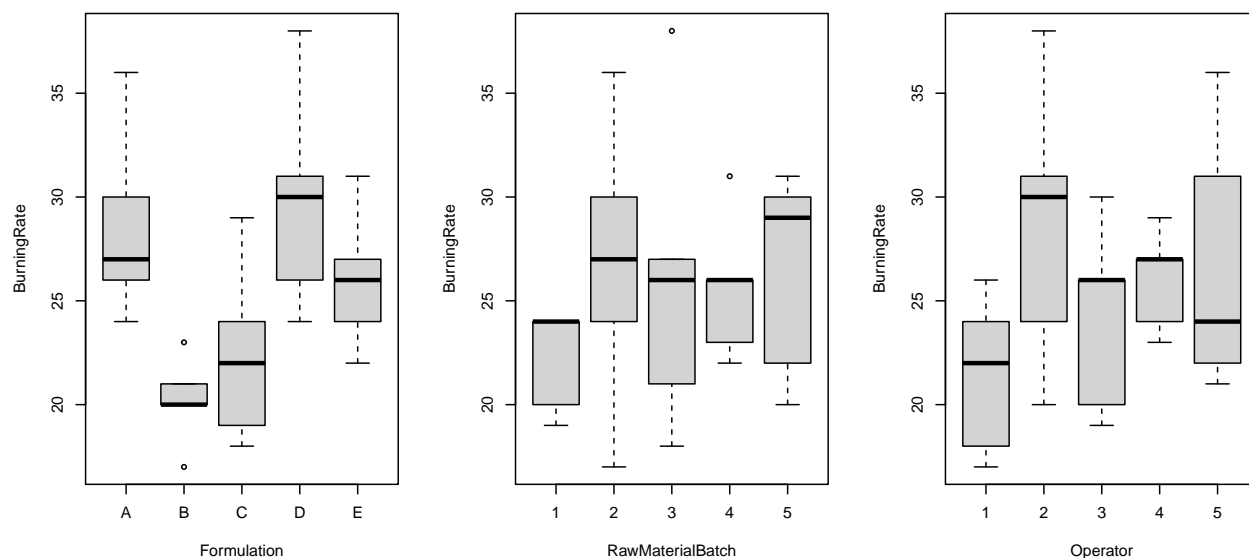
Latin Square Designs

Read the csv file `4_RocketPropellant.csv` in R. Make sure that in the `data.frame` the variables `RawMaterialBatch`, `Operator` and `Formulation` are the type of factor. If not sure, apply `as.factor()` on those variables after reading the dataset.

```
df2 <- read.csv(file.path("data", "4_RocketPropellant.csv"))
df2$RawMaterialBatch <- as.factor(df2$RawMaterialBatch)
df2$Operator <- as.factor(df2$Operator)
df2$Formulation <- as.factor(df2$Formulation)
```

The boxplots shows the distributions of `BurningRate` under subgroups of (left) five treatments of `Formulation`; (middle) five levels of the first block `RawMaterialBatch`; (right) five levels of the second block `Operator`. The left plot suggests that there may be differences in the `BurningRate` among the different `Formulations` used. The middle and right plots, show the `BurningRate` behaves differently within each level of blocks. Thus, the AONVA model must take the block effects into account.

```
# Draw the grouped boxplot
par(mfrow = c(1, 3))
boxplot(BurningRate ~ Formulation, data = df2)
boxplot(BurningRate ~ RawMaterialBatch, data = df2)
boxplot(BurningRate ~ Operator, data = df2)
```



Again, functions `aov()` fits the ANOVA model followed by calling `summary()` to obtain the ANOVA table. For Latin Square Designs, the R model formula includes three items on its right-hand-side: `Formulation + RawMaterialBatch + Operator`. Similar to RCBD, the order of placing the factor and block effects is arbitrary here.

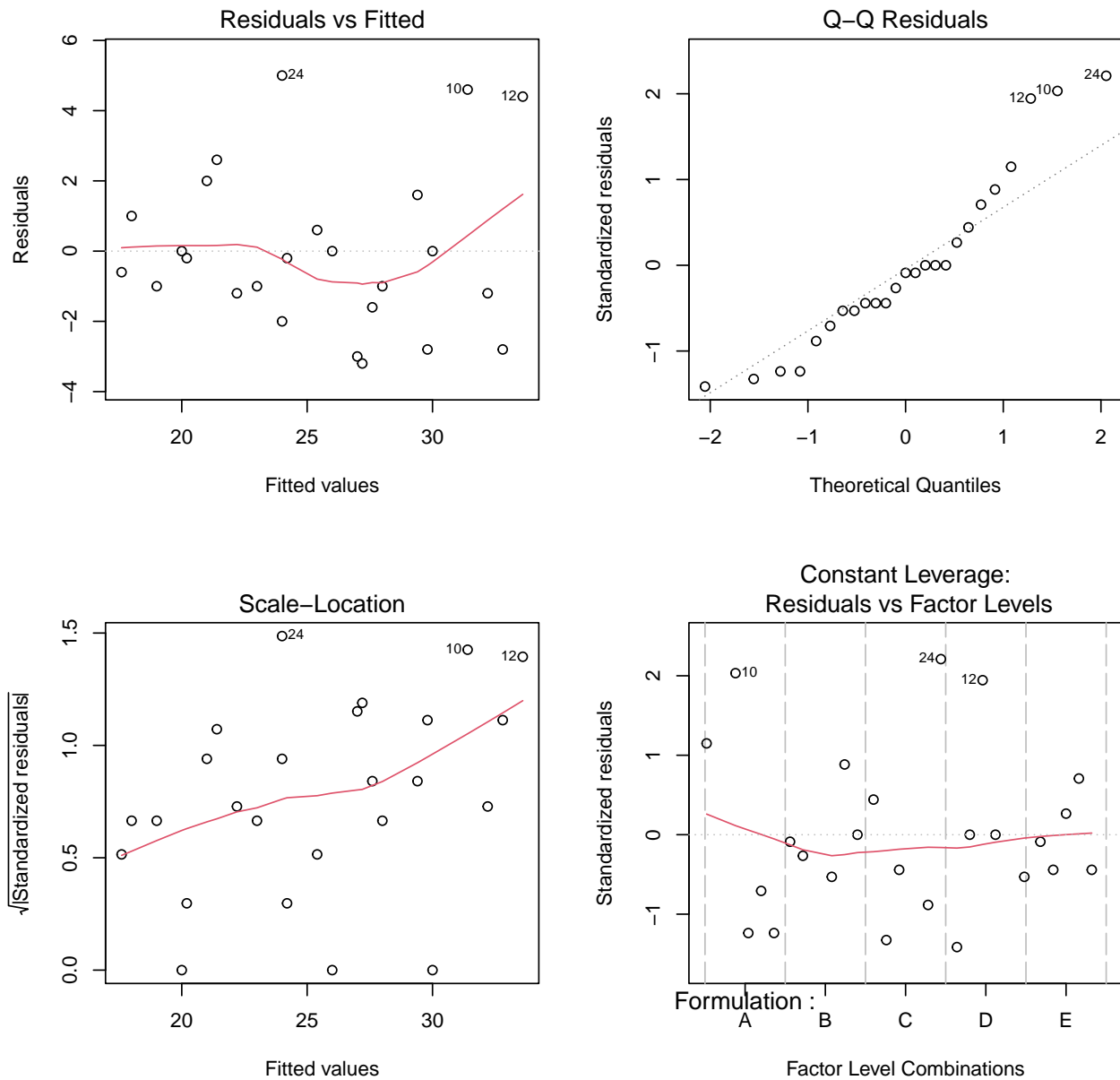
```
fit2 <- aov(BurningRate ~ Formulation + RawMaterialBatch + Operator, data = df2)
summary(fit2)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## Formulation    4      330    82.50   7.734 0.00254 **
```

```
## RawMaterialBatch 4      68    17.00    1.594 0.23906
## Operator          4     150    37.50    3.516 0.04037 *
## Residuals        12    128    10.67
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The procedures of diagnosing the residual and multiple comparison are similar to those for the RCBD. Please refer to the previous section for more details.

```
par(mfrow = c(2, 2))
plot(fit2)
```

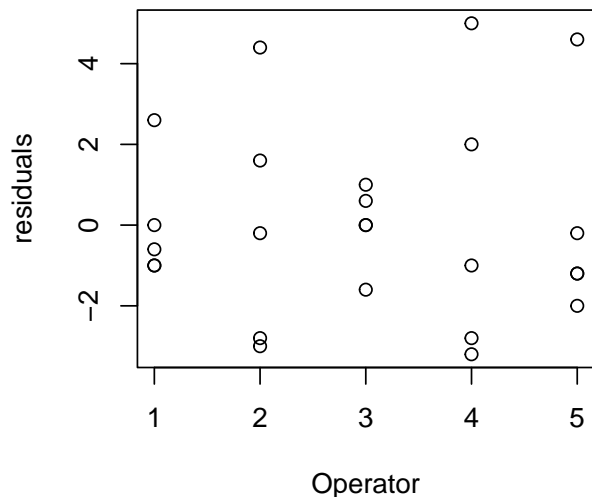
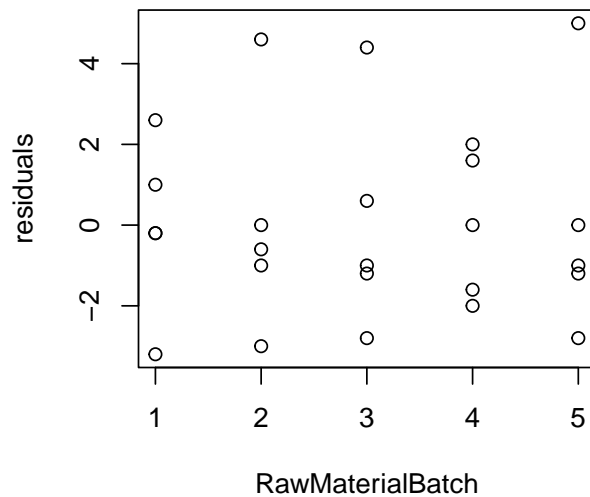


```
par(mfrow = c(1, 2))
plot(
```

```

as.numeric(df2$RawMaterialBatch), fit2$residuals,
xlab = "RawMaterialBatch", ylab = "residuals"
)
plot(
as.numeric(df2$Operator), fit2$residuals,
xlab = "Operator", ylab = "residuals"
)

```



```

if (!("agricolae" %in% rownames(installed.packages()))) {
  install.packages("agricolae")
}
library(agricolae)
out <- LSD.test(fit2, "Formulation", p.adj = "bonferroni")
out$group

```

```

##   BurningRate groups
## D      29.8      a
## A      28.6     ab
## E      26.0     abc
## C      22.4     bc
## B      20.2     c

```

Balanced Incomplete Block Design

Read the csv file 4_Catalysts.csv in R. Make sure that in the data.frame the variables RawMaterialBatch and Catalysts are the type of factor. If not sure, apply as.factor() on those variables after reading the dataset.

```

df3 <- read.csv(file.path("data", "4_Catalysts.csv"))
df3$RawMaterialBatch <- as.factor(df3$RawMaterialBatch)
df3$Catalysts <- as.factor(df3$Catalysts)

```

This burning rate experiment data is obtained by the balanced incomplete block design (BIBD) with $a = 4$ Catalysts treatment levels and $b = 4$ RawMaterialBatch block levels. Each treatment levels are performed

in $r = 3$ blocks, and in each block, $k = 3$ treatment levels are performed. The effect model is

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, \quad i = 1, \dots, 4, \quad j = 1, \dots, 4$$

where τ_i denotes the treatment effect of types of `Catalysts` and β_j denotes the block effect of the `RawMaterialBatch`.

If the design is BIBD, the adjusted treatment sum of squares is

$$SS_{Treatment(adj)} = \frac{k \sum_{i=1}^b Q_i^2}{\lambda a} = \frac{3 \sum_{i=1}^4 Q_i^2}{4\lambda}$$

where $\lambda = \frac{r(k-1)}{a-1} = \frac{3 \cdot 2}{3} = 2$ and

$$Q_i = y_{i\cdot} - \frac{1}{k} \sum_{j=1}^b n_{ij} y_{\cdot j} = y_{i\cdot} - \frac{1}{3} \sum_{j=1}^4 n_{ij} y_{\cdot j}$$

for $i = 1, \dots, 4$, and, $n_{ij} = 1$ if the i th treatment level appears in block j and $n_{ij} = 0$ otherwise. Let's compute $SS_{Treatment(adj)}$ by ourselves in R first.

```
appears <- tapply(df3$RawMaterialBatch, df3$Catalyst, table)
y_i_dot <- tapply(df3$ReactionTime, df3$Catalysts, sum)
y_dot_j <- tapply(df3$ReactionTime, df3$RawMaterialBatch, sum)
q <- numeric(4)
for (i in 1:4) {
  n_i_j <- appears[[i]]
  q[i] <- y_i_dot[i] - (1/3)*sum(n_i_j*y_dot_j)
}
SS_Treatment_adj <- 3*sum(q^2)/(2*4)
SS_Treatment_adj
```

```
## [1] 22.75
```

To obtain the ANOVA table with the adjusted treatment sum of squares for BIBD in R, keep the block effect in the first place of the R model formula for the `aov` function.

```
fit3_treatment <- aov(ReactionTime ~ RawMaterialBatch + Catalysts, data = df3)
summary(fit3_treatment)
```

```
##               Df Sum Sq Mean Sq F value    Pr(>F)
## RawMaterialBatch  3  55.00  18.333    28.20 0.00147 **
## Catalysts         3   22.75   7.583    11.67 0.01074 *
## Residuals        5   3.25   0.650
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Likewise, if one is interested to test the significance of the block effect, then the block sum of squares should be adjusted. In R, the adjusted block sum of squares is obtained by placing the “treatment effect” in the first place of the R model formula for the `aov` function.

```
fit3_block <- aov(ReactionTime ~ Catalysts + RawMaterialBatch, data = df3)
summary(fit3_block)
```



```
##               Df Sum Sq Mean Sq F value    Pr(>F)
## Catalysts      3  11.67    3.889    5.983 0.041463 *
## RawMaterialBatch 3  66.08   22.028   33.889 0.000953 ***
## Residuals      5    3.25    0.650
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

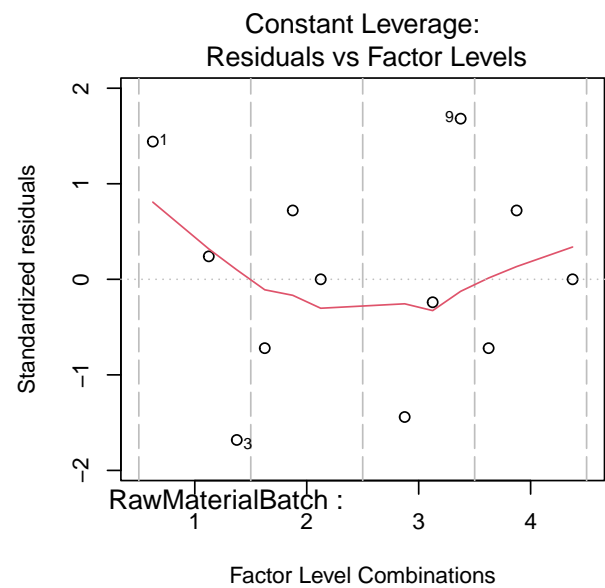
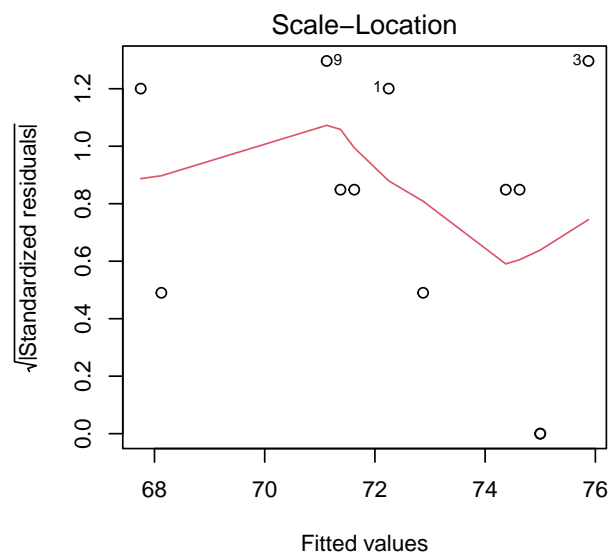
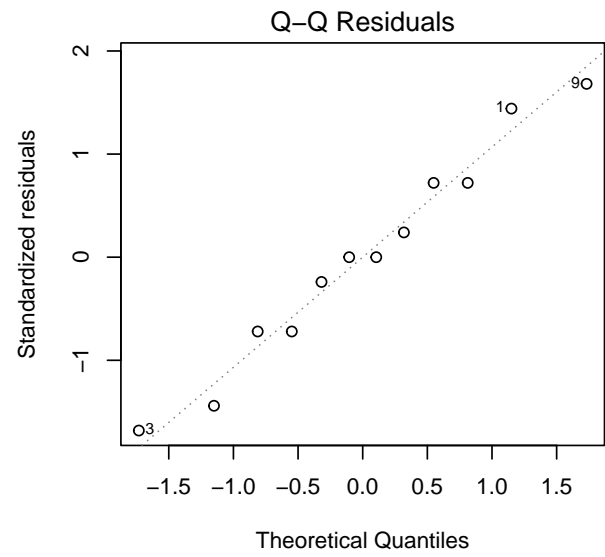
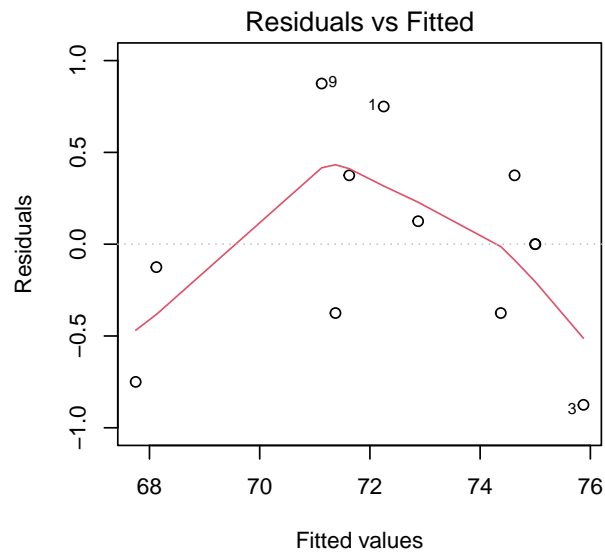
The R package **ibd** provides the function `aov.ibd` to compute the adjusted treatment and block sum of squares at the same time.

```
if (!("ibd" %in% rownames(installed.packages()))) {
  install.packages("ibd")
}
library(ibd)
fit3 <- aov.ibd(ReactionTime ~ Catalysts + RawMaterialBatch, data = df3)
fit3
```

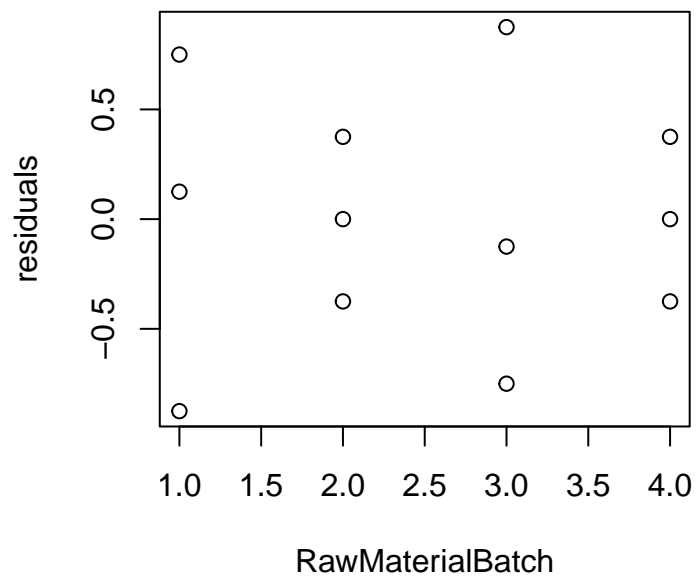
```
## $ANOVA.table
## Anova Table (Type III tests)
##
## Response: ReactionTime
##               Sum Sq Df    F value    Pr(>F)
## (Intercept)   8948.7  1 13767.198 8.528e-10 ***
## Catalysts      22.7   3    11.667 0.0107387 *
## RawMaterialBatch 66.1   3    33.889 0.0009528 ***
## Residuals      3.3   5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The procedures of diagnosing the residual and multiple comparison are similar to those for the RCBD. Please refer to the previous section for more details.

```
# Residual Checking:
# can use any aov model fit3_treatment or fit3_block
par(mfrow = c(2, 2))
plot(fit3_treatment)
```



```
plot(
  as.numeric(df3$RawMaterialBatch), fit3_treatment$residuals,
  xlab = "RawMaterialBatch", ylab = "residuals"
)
```



```
if (!("agricolae" %in% rownames(installed.packages()))) {
  install.packages("agricolae")
}
library(agricolae)
out <- LSD.test(fit3_treatment, "Catalysts", p.adj = "bonferroni")
out$group
```

```
## ReactionTime groups
## 4      74.00000      a
## 1      72.66667      a
## 3      72.00000      a
## 2      71.33333      a
```