STA322 Project 2

Sherry Huang, Eddie Yang

April 2019

Introduction

This study aims to examine the factors in desgining a paper helicopter that affect how long it can fly in the air. Specifically, the study focuses on four factors: rotor length, leg length, leg width, and whether or not there is a paper clip at the bottom of the leg. This report includes details of the full factorial experiment design, procedures of data collection, and analysis of estimation results using regression modelling.

Study Design and Data Collection

The four binary factors examined for paper helicopter fly time in this study are shown in the table below.

Table 1: Paper Helicopter Factors

Factor	Low	High
Rotor Length (a)	7.5cm	8.5cm
Leg Length (b)	7.5cm	12.0cm
Leg Width (c)	3.2cm	5.0cm
Paper Clip (d)	No	Yes

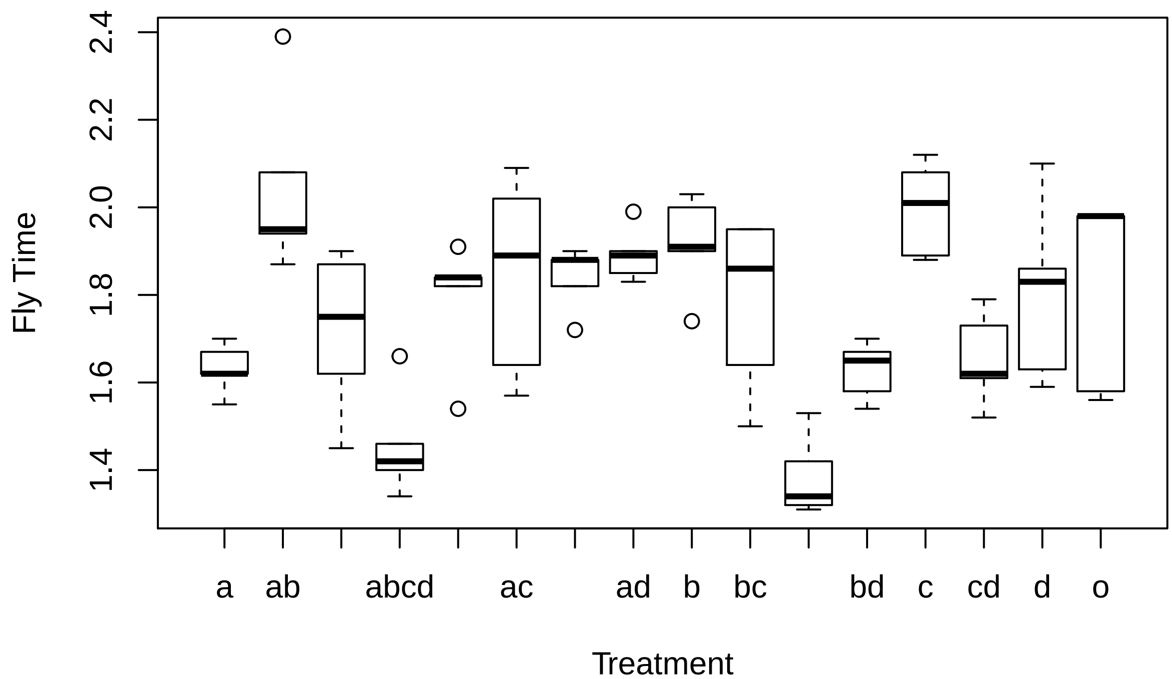
Because each factor has two levels, there are a total of $2^4 = 16$ total possible treatment assignments. Since we have no prior knowledge of how these factors interact with each other and thus cannot assume their interaction effects are negligible, we decided to use a full factorial design for this experiment, which allows us to examine all main effects and interaction effects. 8 paper helicopters are made following the guidelines posted on Sakai by the same person to ensure consistency, each of which corresponds to a possible treatment assignment. Adding the paper clip to each of 8 helicopters gives us the other 8 possible treatment combinations. A standard silver paper clip is added to the bottom of the leg of each helicopter so that the lower side of the clip alligns with the bottom edge of the leg. 5 trials are conducted for each of the 16 assignment, which gives us $16 * 5 = 80$ trials in total. The trials are completed in a completely randomized order. During each trial, we drop the paper helicopter at the height of exactly 2 meters, and the flight time is measured in seconds from the time the helicopter is dropped to the time it hits the floor. The data is recorded in the accompanying data file "helidata.csv". The 16 treatments are denoted with respective letter combinations. For example, a treatment of "o" indicates that all four factors are at their low level, while "ab" indicates that rotor length and leg length are at their high values, leg width is at its low level, and there is no paper clip on the leg. The table below shows the first 5 rows of the data frame.

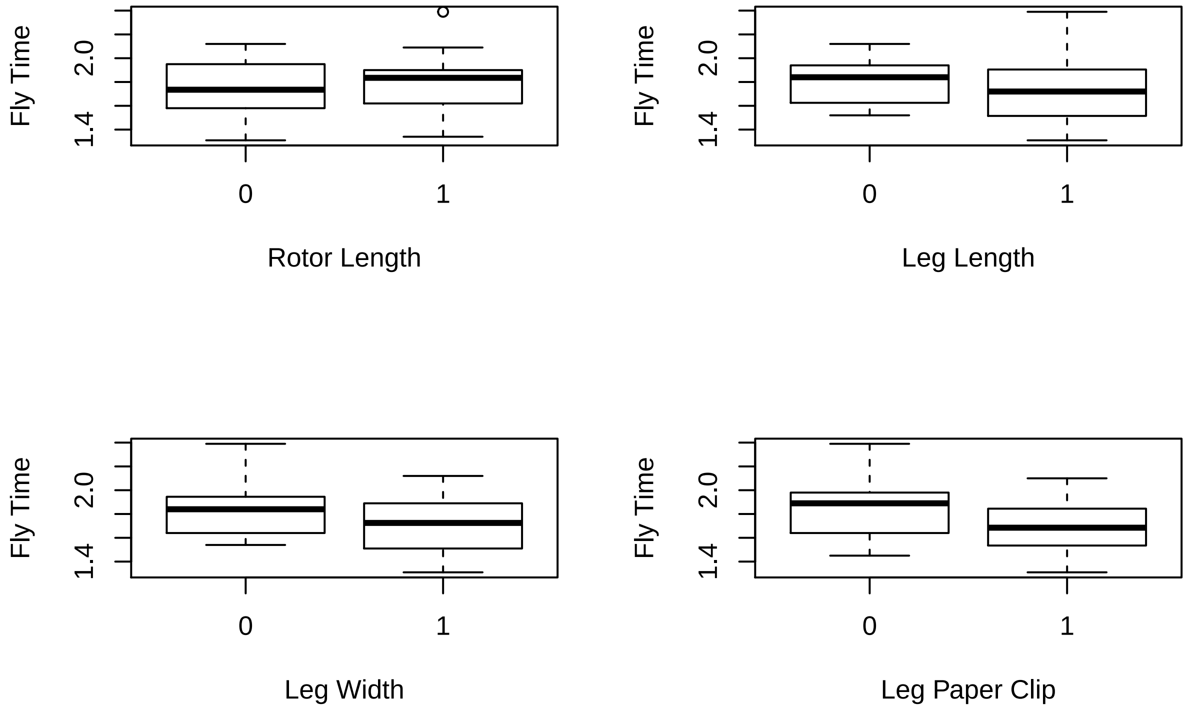
Table 2: First 5 Rows of Collected Data

treatments	a	b	c	d	time
ad	1	0	0	1	1.89
d	0	0	0	1	2.10
b	0	1	0	0	1.90
acd	1	0	1	1	1.88
o	0	0	0	0	1.56

Data Analysis and Results

As exploratory data analysis, we make boxplots of paper helicopter fly time against all treatment combinations and the four main factors. It is shown that a few treatment assignments yield a significantly higher or lower fly time than others, but there are great variabilities across trials. For each of the four main factors, it seems like there is a noticable difference in median fly times between low and high level of each factor. From the boxplots, it seems like a paper helicopter with a longer rotor, a narrower and shorter leg, and no paper clip will fly longer in the air.





We fit a linear regression model with all main effects, all two-way, three-way, and four-way interaction effects on the collected data. The coefficients and p-values for the full factorial regression are shown in the table below.

Table 3: Full Factorial Regression Results

	X2.5	X97.5	coefficients	pvalues
(Intercept)	1.6820199	1.9499801	1.816	<2e-16
a1	-0.3734765	0.0054765	-0.184	0.0568
b1	-0.0894765	0.2894765	0.100	0.2957
c1	-0.0094765	0.3694765	0.180	0.0622
d1	-0.2034765	0.1754765	-0.014	0.8831
a1:b1	0.0460397	0.5819603	0.314	0.0224
a1:c1	-0.2379603	0.2979603	0.030	0.8237
b1:c1	-0.5839603	-0.0480397	-0.316	0.0216
a1:d1	0.0060397	0.5419603	0.274	0.0452
b1:d1	-0.5419603	-0.0060397	-0.274	0.0452
c1:d1	-0.5959603	-0.0600397	-0.328	0.0172
a1:b1:c1	-0.6009530	0.1569530	-0.222	0.2462
a1:b1:d1	-0.6209530	0.1369530	-0.242	0.2067
a1:c1:d1	-0.3129530	0.4449530	0.066	0.729
b1:c1:d1	-0.1589530	0.5989530	0.220	0.2504
a1:b1:c1:d1	-0.4999205	0.5719205	0.036	0.8937

To find out the most important factor that affect helicopter fly time, nested F-tests are conducted one at a time between the full model and models excluding each factor. Specific code and output can be found at the end of this report. The test done for the model without the paper clip factor yields the highest F score (7.7634) and the lowest p-value (3.224e-07), indicating that whether or not the helicopter has a paper clip at the bottom of its leg is the most important factor for fly time. The regression coefficients show that, although the main effect of paper clip is not statistically significant, it has significant interactione effects with all the other factors. When all other factors are at their lower levels, the fly time of the paper helicopter decreases 0.014s (95% CI: -0.20348, 0.17548) on average if there is a paper clip added to the leg. When leg length and width are short, adding a paper clip increases the impact of having a longer roter length by 0.274s (95% CI: 0.00604, 0.54196); when rotor length and leg width are low, adding a paper clip decreases the impact of having a longer leg length by 0.274s (95% CI: -0.54196, -0.00604); similarly, when rotor length and leg length are low, adding a paper clip decreases the impact of a wider leg by 0.328s (95% CI: -0.59596, -0.06004). Since the interaction effect between leg width (c) and rotor length (a) has a close to zero coefficient and a large p-value of 0.8831, there is not enough evidence that the effect of rotor length differs by leg width. Its 95% confidence interval, (-0.23796, 0.29796), further suggests that the interaction effect is not significant, if any. Using the coefficients from the full regression model, we can also calculate the estimated average fly time for all possible combination of assignments. Among the combinations, a rotor length of 7.5cm, a leg length of 12.0cm, a leg width of 5.0cm, and a paper clip at the bottom of the leg (bcd) will make the helicopter fly the longest, for 2.302s on average. Given that all three-way and four-way interactions have large p-values, we attempt a smaller model with only main effects and two-way interactions. However, nested F-test against the full model shows that excluding three-way and four-way interactions does not make a difference to model fit. The adjusted R-squared value of the smaller model is also lower than that of the full model. Therefore, we decide to keep the full model for interpretation and analysis.

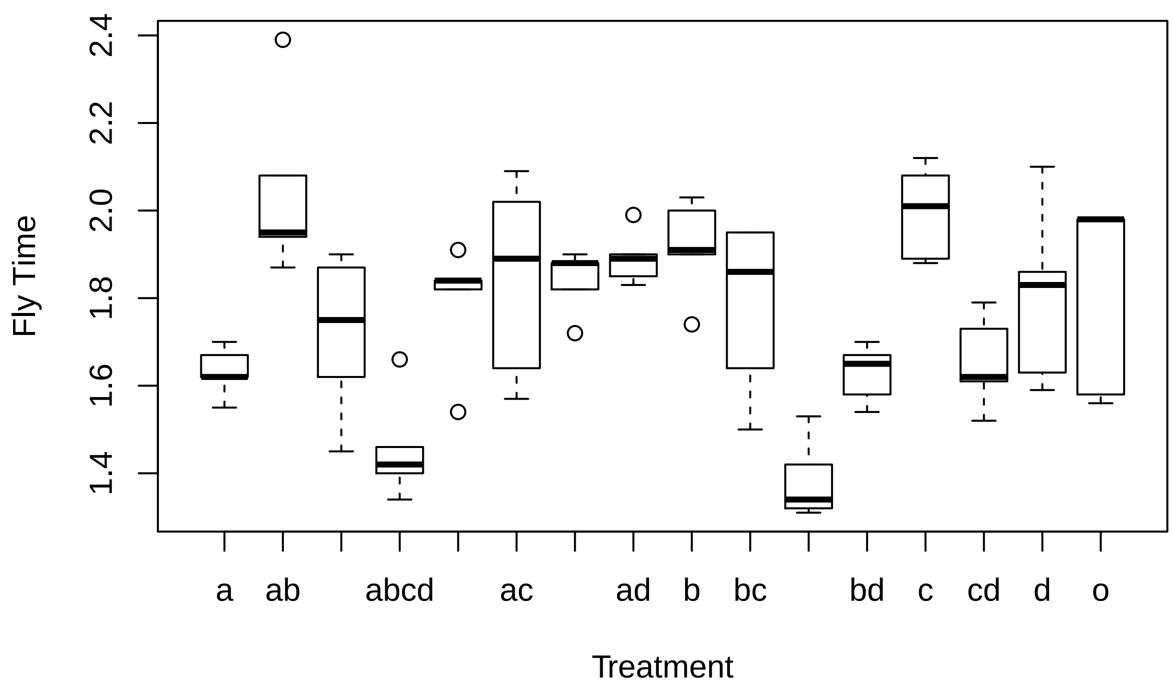
Discussion

There are a few limitations and possible sources of error in our experiment design and data collection process. First, the helicopters are dropped by hand. The person who drops them might apply some force on the helicopter upon releasing, resulting in different initial velocities. In some cases, in order not to let helicopters hit the wall, the person stood away from any surfaces and may make dropping heights vary due to lack of marking. Second, because the time is recorded by hand, inaccuracy is inevitable. It is especially difficult to start and stop the timer exactly at the moment the helicopter is dropped and when it hits the ground respectively. Eyeballing for both timing and dropping heights can lead to noises and inaccuracies in the data. Increasing repetitions for each assignment may be helpful. In addition, we notice that in some trials the helicopter spins vertically and stably while in others it randomly floats around. The helicopter can move in the air for a similar amount of time, but some designs may allow it to fly more "properly" than others, and therefore the fly time may not be the most optimal outcome variable upon which the design factors are evaluated. For the narrower leg width, since two pieces of paper are folded to the center according to instruction, this may add more stability to the helicopter than the case with a wider leg width where it is just a piece of paper. It can be a confounding factor in the analysis.

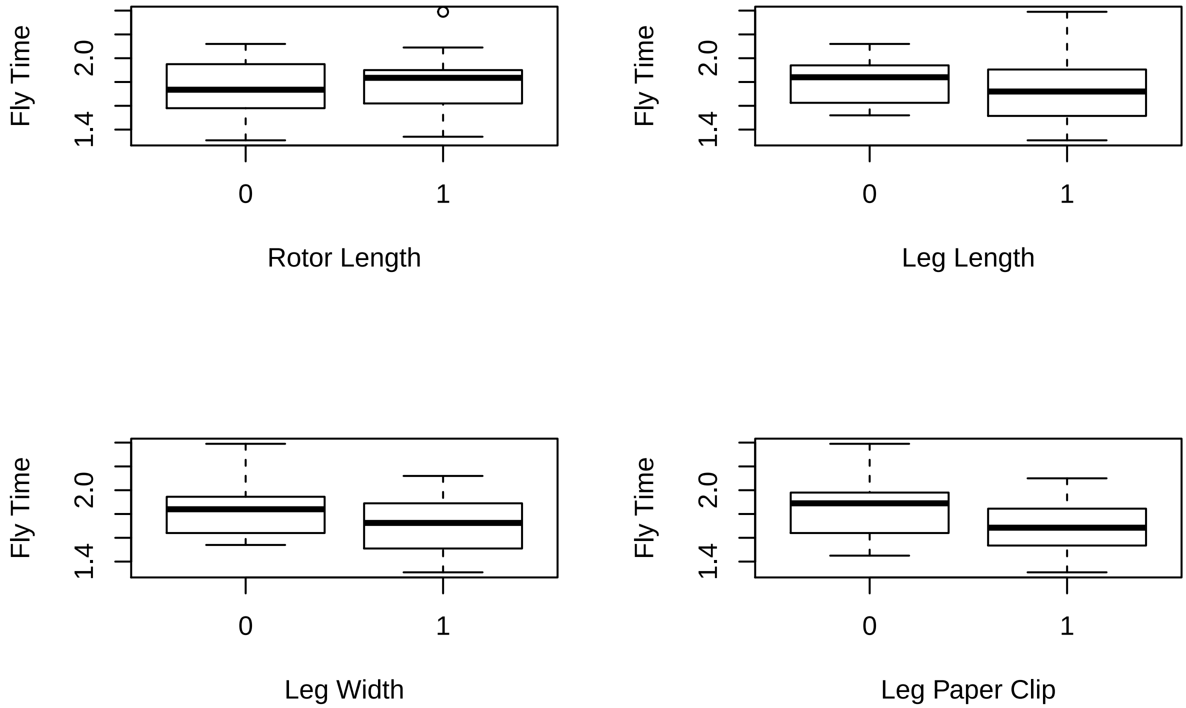
Code

```r
#Randomize treatment assignments
# assignments = c(rep("o",5), rep("a",5), rep("b",5), rep("c",5), rep("d",5),
#                 rep("ab",5), rep("ac",5),rep("ad",5), rep("bc",5), rep("bd",5),
#                 rep("cd",5), rep("abc",5), rep("abd",5), rep("acd",5),
#                 rep("bcd",5), rep("abcd",5))
# set.seed(322)
# assignments = sample(assignments)
# helidata = data.frame(treatments = assignments,
#                       a = NA, b = NA, c = NA, d = NA, time = NA)
# write.csv(helidata, file = "helidata.csv")

#Collect data, enter manually, and read file again
```

```r
heliData = read.csv("helidata.csv", header = TRUE)
#Convert treatments to factors
heliData$treatments = as.factor(heliData$treatments)
heliData$a = as.factor(heliData$a)
heliData$b = as.factor(heliData$b)
heliData$c = as.factor(heliData$c)
heliData$d = as.factor(heliData$d)

#Exploratory Data Analysis
boxplot(heliData$time ~ heliData$treatments, xlab = "Treatment", ylab = "Fly Time")
```




```r
par(mfrow=c(2,2))
boxplot(heliData$time ~ heliData$a, xlab = "Rotor Length", ylab = "Fly Time")
boxplot(heliData$time ~ heliData$b, xlab = "Leg Length", ylab = "Fly Time")
boxplot(heliData$time ~ heliData$c, xlab = "Leg Width", ylab = "Fly Time")
boxplot(heliData$time ~ heliData$d, xlab = "Leg Paper Clip", ylab = "Fly Time")
```




```r
#Regression
regHeli = lm(time ~ a*b*c*d, data = heliData)
summary(regHeli)
```

```
##
## Call:
## lm(formula = time ~ a * b * c * d, data = heliData)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -0.280 -0.090  0.011  0.081  0.344
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.81600    0.06707  27.078   <2e-16 ***
## a1          -0.18400    0.09485  -1.940   0.0568 .
## b1           0.10000    0.09485   1.054   0.2957
## c1           0.18000    0.09485   1.898   0.0622 .
## d1          -0.01400    0.09485  -0.148   0.8831
## a1:b1        0.31400    0.13413   2.341   0.0224 *
## a1:c1        0.03000    0.13413   0.224   0.8237
## b1:c1       -0.31600    0.13413  -2.356   0.0216 *
## a1:d1        0.27400    0.13413   2.043   0.0452 *
## b1:d1       -0.27400    0.13413  -2.043   0.0452 *
## c1:d1       -0.32800    0.13413  -2.445   0.0172 *
## a1:b1:c1    -0.22200    0.18969  -1.170   0.2462
```

```
## a1:b1:d1    -0.24200    0.18969  -1.276   0.2067
## a1:c1:d1     0.06600    0.18969   0.348   0.7290
## b1:c1:d1     0.22000    0.18969   1.160   0.2504
## a1:b1:c1:d1  0.03600    0.26826   0.134   0.8937
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.15 on 64 degrees of freedom
## Multiple R-squared:  0.6247, Adjusted R-squared:  0.5367
## F-statistic: 7.101 on 15 and 64 DF,  p-value: 8.443e-09
```

```
confint(regHeli)
```

```
##                    2.5 %       97.5 %
## (Intercept)  1.682019866  1.949980134
## a1          -0.373476522  0.005476522
## b1          -0.089476522  0.289476522
## c1          -0.009476522  0.369476522
## d1          -0.203476522  0.175476522
## a1:b1        0.046039732  0.581960268
## a1:c1       -0.237960268  0.297960268
## b1:c1       -0.583960268 -0.048039732
## a1:d1        0.006039732  0.541960268
## b1:d1       -0.541960268 -0.006039732
## c1:d1       -0.595960268 -0.060039732
## a1:b1:c1    -0.600953045  0.156953045
## a1:b1:d1    -0.620953045  0.136953045
## a1:c1:d1    -0.312953045  0.444953045
## b1:c1:d1    -0.158953045  0.598953045
## a1:b1:c1:d1 -0.499920535  0.571920535
```

```
#Print regression results
knitr::kable(data.frame(estimates = regHeli$coefficients,
                        pvalues = c("<2e-16",0.0568,0.2957,0.0622,0.8831,0.0224,0.8237,
                                    0.0216,0.0452,0.0452,0.0172,0.2462,0.2067,0.7290,
                                    0.2504,0.8937)),
             caption = "Full Factorial Regression Results")
```

Table 4: Full Factorial Regression Results

	estimates	pvalues
(Intercept)	1.816	<2e-16
a1	-0.184	0.0568
b1	0.100	0.2957
c1	0.180	0.0622
d1	-0.014	0.8831
a1:b1	0.314	0.0224
a1:c1	0.030	0.8237
b1:c1	-0.316	0.0216
a1:d1	0.274	0.0452
b1:d1	-0.274	0.0452
c1:d1	-0.328	0.0172
a1:b1:c1	-0.222	0.2462
a1:b1:d1	-0.242	0.2067
a1:c1:d1	0.066	0.729

	estimates	pvalues
b1:c1:d1	0.220	0.2504
a1:b1:c1:d1	0.036	0.8937

```
#nested F-test to find the most important main factor
regHeli2 = lm(time ~ b*c*d, data = heliData)
regHeli3 = lm(time ~ a*c*d, data = heliData)
regHeli4 = lm(time ~ a*b*c, data = heliData)

anova(regHeli, regHeli2)
```

```
## Analysis of Variance Table
##
## Model 1: time ~ a * b * c * d
## Model 2: time ~ b * c * d
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     64 1.4393
## 2     72 1.8204 -8   -0.3811 2.1182 0.04664 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#F = 2.1182, p-value = 0.04664
anova(regHeli, regHeli3)
```

```
## Analysis of Variance Table
##
## Model 1: time ~ a * b * c * d
## Model 2: time ~ a * c * d
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     64 1.4393
## 2     72 2.7005 -8   -1.2612 7.0098 1.332e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#F = 7.0098, p-value = 1.332e-06
anova(regHeli, regHeli4)
```

```
## Analysis of Variance Table
##
## Model 1: time ~ a * b * c * d
## Model 2: time ~ a * b * c
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     64 1.4393
## 2     72 2.8361 -8   -1.3968 7.7634 3.224e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#F = 7.7634, p-value = 3.224e-07

#try a smaller model with three-way and four-way interactions
smallHeli = lm(time ~ a*b + a*c + a*d + b*c + b*d + c*d, data = heliData)
summary(smallHeli)
```

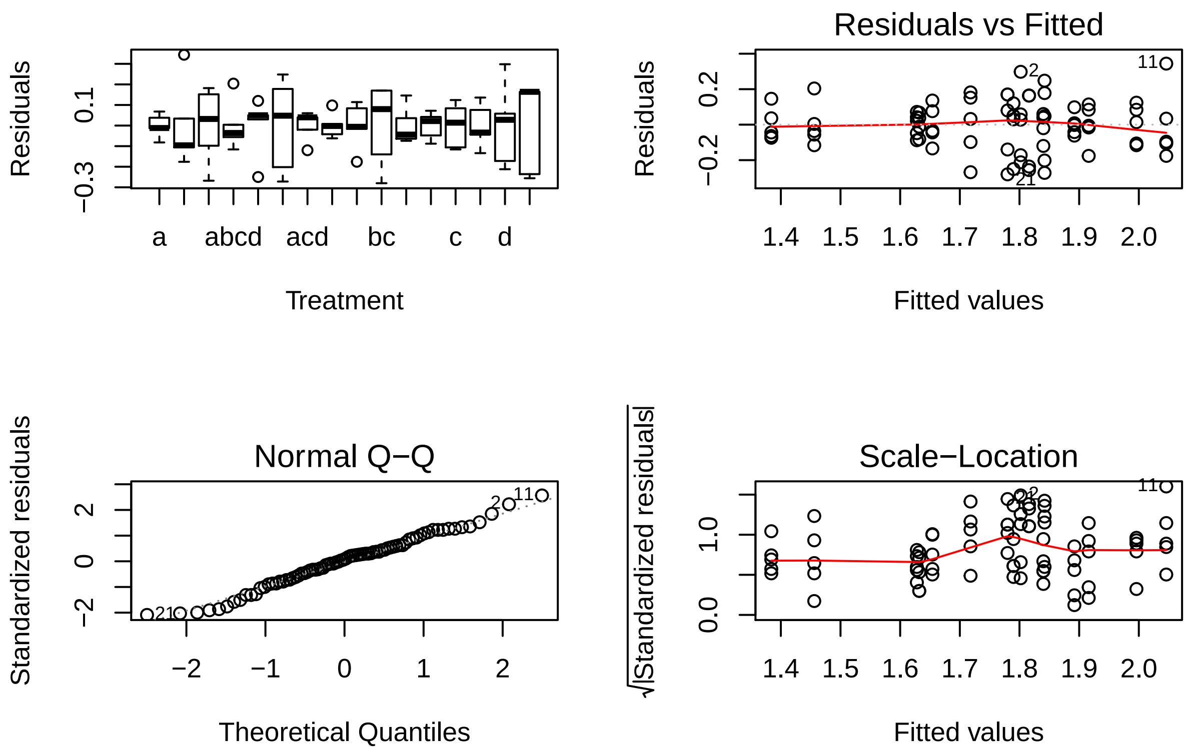
```
##
## Call:
```

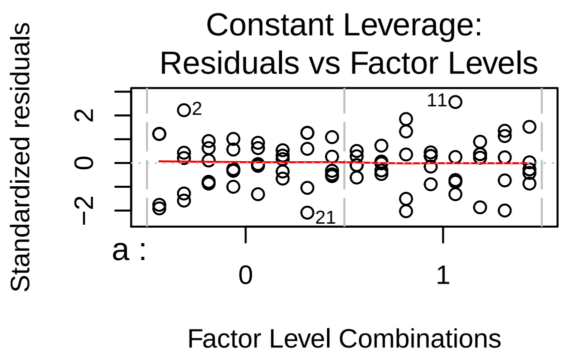
```
## lm(formula = time ~ a * b + a * c + a * d + b * c + b * d + c *
##     d, data = heliData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30800 -0.08700  0.00250  0.08112  0.44000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.80050    0.05706  31.552  < 2e-16 ***
## a1          -0.09350    0.06882  -1.359 0.178707
## b1           0.15200    0.06882   2.209 0.030528 *
## c1           0.15500    0.06882   2.252 0.027496 *
## d1          -0.03400    0.06882  -0.494 0.622856
## a1:b1        0.09100    0.06882   1.322 0.190452
## a1:c1       -0.03900    0.06882  -0.567 0.572772
## a1:d1        0.19500    0.06882   2.833 0.006034 **
## b1:c1       -0.30800    0.06882  -4.475 2.93e-05 ***
## b1:d1       -0.27600    0.06882  -4.010 0.000151 ***
## c1:d1       -0.17600    0.06882  -2.557 0.012750 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1539 on 69 degrees of freedom
## Multiple R-squared:  0.5739, Adjusted R-squared:  0.5121
## F-statistic: 9.293 on 10 and 69 DF,  p-value: 1.523e-09
```

```r
#R-squared and adjusted R-squared both decreased
anova(regHeli, smallHeli)
```

```
## Analysis of Variance Table
##
## Model 1: time ~ a * b * c * d
## Model 2: time ~ a * b + a * c + a * d + b * c + b * d + c * d
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     64 1.4393
## 2     69 1.6341 -5  -0.19477 1.7321   0.14
```

```r
#does not make a difference so we can keep the full model

#Residual Plot
boxplot(regHeli$residuals~heliData$treatments, xlab = "Treatment", ylab = "Residuals")
#Diagnostic Plots
plot(regHeli)
```


Residuals
0.1
−0.3
a
abcd
acd
bc
c
d
Treatment
Residuals vs Fitted
Residuals
0.2
−0.2
2
11
21
1.4
1.5
1.6
1.7
1.8
1.9
2.0
Fitted values
Normal Q−Q
Standardized residuals
2
0
−2
21
2
11
−2
−1
0
1
2
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
0.0
1.0
11
2
1.4
1.5
1.6
1.7
1.8
1.9
2.0
Fitted values


Constant Leverage:
Residuals vs Factor Levels
Standardized residuals
2
0
−2
2
11
21
a :
0
1
Factor Level Combinations