STAT 210
Applied Statistics and Data Analysis:
Homework 6

Due on Oct. 30/2022

Question 1

The table below has the results of an experiment run to determine the effect of four different oven temperatures in °C on the density of a certain type of ceramics.

Table 1: Question 1: Density of Ceramics

Temperature	Density				
100	21.8	21.9	21.7	21.6	21.7
125	21.5	21.4	21.5	21.4	21.6
150	21.7	21.8	21.9	21.8	21.9
175	21.9	21.7	21.8	21.6	21.7

The data can be loaded by copying the commands below.

```r
density <- c(21.8, 21.9, 21.7, 21.6, 21.7,
             21.5, 21.4, 21.5, 21.4, 21.6,
             21.7, 21.8, 21.9, 21.8, 21.9,
             21.9, 21.7, 21.8, 21.6, 21.7)
temp <- factor(rep(c(100,125,150,175), each = 5))
Q1data <- data.frame(temp,density)
```

Do a complete analysis of variance for this set. Plot the data. Write the equation for the model. Determine whether the treatments have an effect on the density of ceramics by means of a hypothesis test. Plot the diagnostic charts and comment on them. Use also Levene's test and Shapiro-Wilk. Use Tukey's HSD procedure to make pairwise comparisons and comment on the results.

Solution

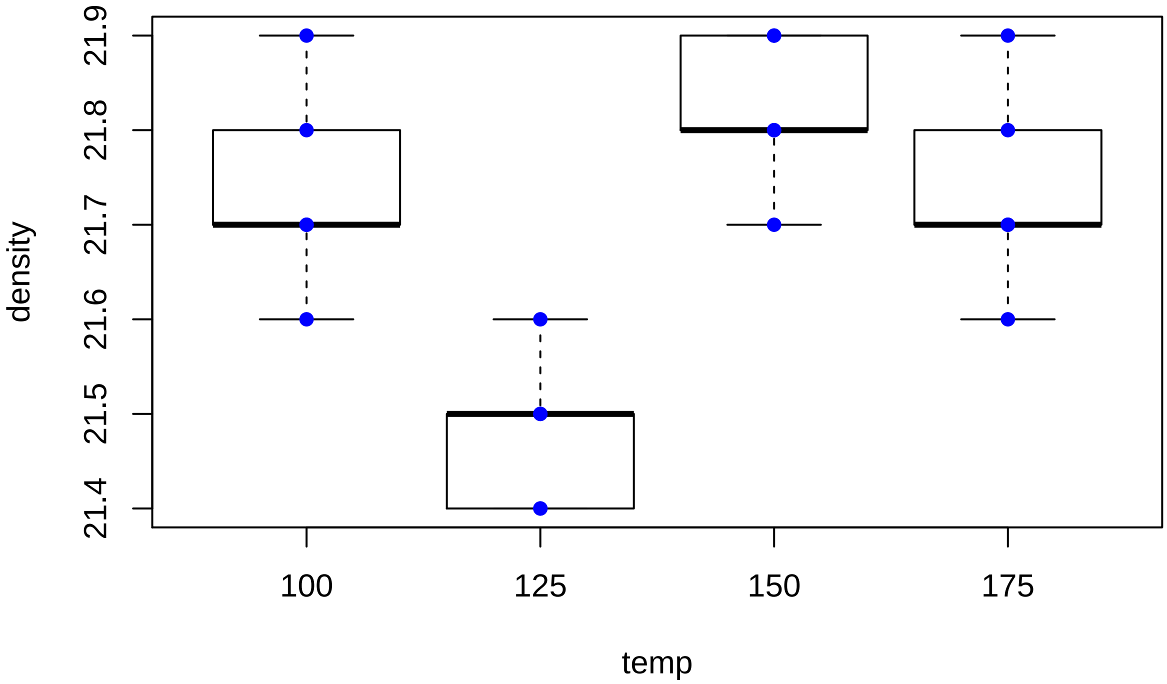
Read the data

```r
density <- c(21.8, 21.9, 21.7, 21.6, 21.7,
             21.5, 21.4, 21.5, 21.4, 21.6,
             21.7, 21.8, 21.9, 21.8, 21.9,
             21.9, 21.7, 21.8, 21.6, 21.7)
temp <- factor(rep(c(100,125,150,175), each = 5))
Q1data <- data.frame(temp,density)
str(Q1data)
```

```
## 'data.frame':    20 obs. of  2 variables:
##  $ temp   : Factor w/ 4 levels "100","125","150",..: 1 1 1 1 1 2 2 2 2 2 ...
##  $ density: num  21.8 21.9 21.7 21.6 21.7 21.5 21.4 21.5 21.4 21.6 ...
```

Plot the data:

```r
plot(density ~ temp, data = Q1data)
points(density ~ temp, data = Q1data, pch = 16, col = 'blue')
```



One observes that the boxes have similar sizes, which is an indication that the variances are equal. The density for a temperature of 125° seems to be lower than for the other values of `temp`.

The model for the experiment is

$$y_{ik} = \mu + \tau_i + \epsilon_{ik}$$

with $1 \leq i \leq 4$, $1 \leq k \leq 5$. μ is the overall mean, τ_i represents the effect of temperature i, and ϵ_{ik} are the experimental errors, which are assumed to be independent Gaussian random variables with mean 0 and common variance σ^2. We fit this model with

```r
model1 <- aov(density ~ temp, data = Q1data)
```

We produce an anova table.

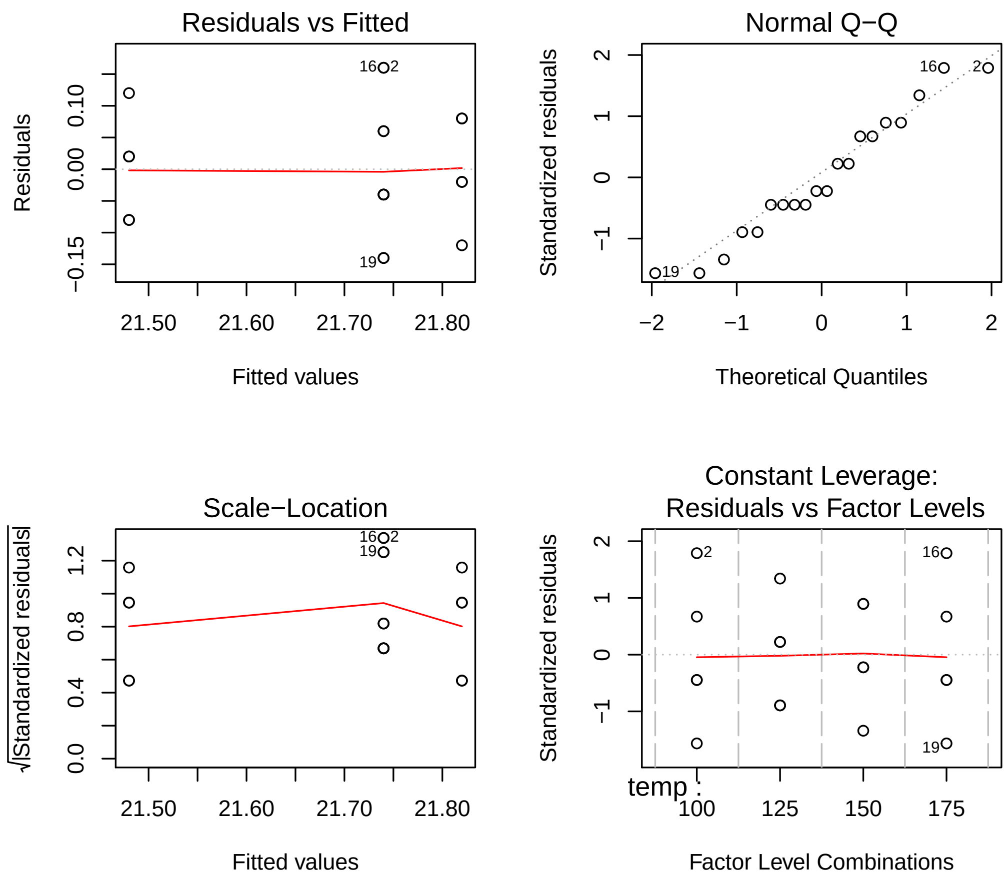
```r
anova(model1)
```

```
## Analysis of Variance Table
##
## Response: density
##           Df Sum Sq Mean Sq F value    Pr(>F)
## temp       3 0.3295 0.10983  10.983 0.0003663 ***
## Residuals 16 0.1600 0.01000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value is small and we reject the null hypothesis of no effect.

Diagnostic charts

```r
par(mfrow=c(2,2))
plot(model1)
```




```r
par(mfrow = c(1,1))
```

There are no concerns with these plots. They all look good. The quantile plot indicates that the assumption of normality is satisfied. This is confirmed with a Shapiro-Wilk test:

```r
shapiro.test(resid(model1))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(model1)
## W = 0.95319, p-value = 0.4181
```

The p-value is large and we do not reject the hypothesis of normality. To test for homoscedasticity we use Levene's test:

```r
library(car)
leveneTest(model1)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  3  0.1333 0.9388
##       16
```

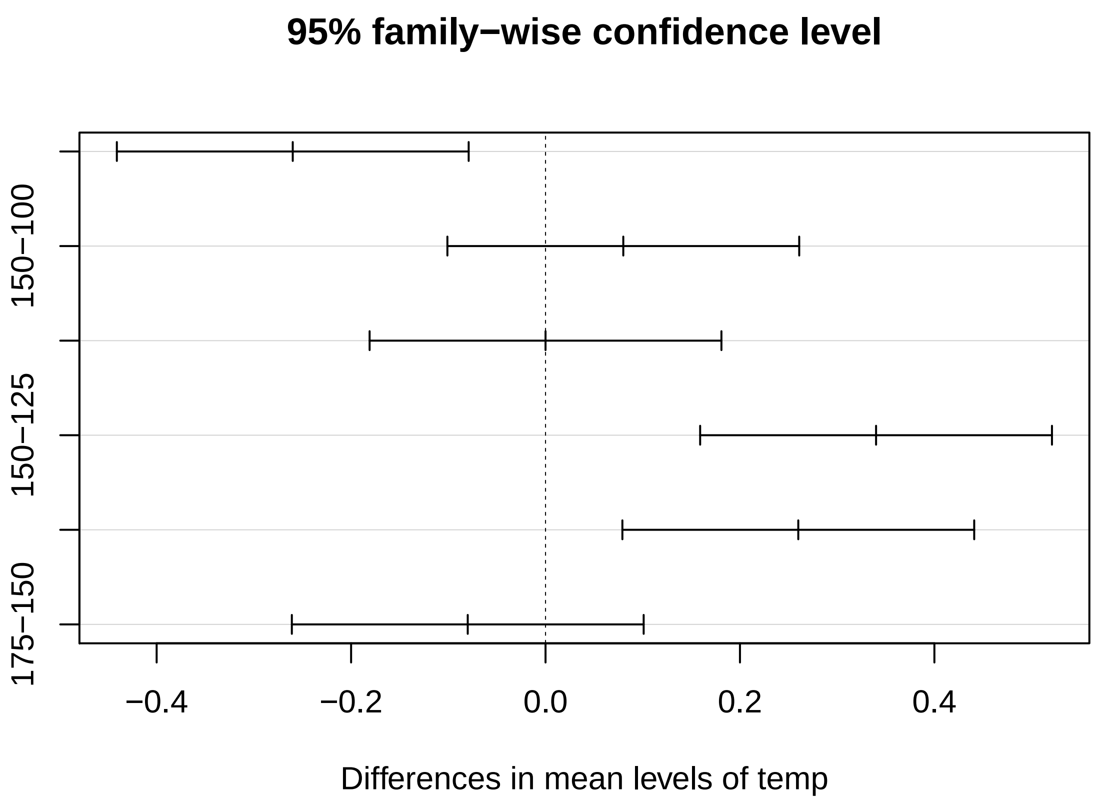
The p-value is large and homoscedasticity is not rejected. We conclude that the assumptions for the model are satisfied.

We can do pairwise comparisons using `TukeyHSD`. We will use a confidence level of 5%.

```r
(model1.tukey <- TukeyHSD(model1, conf.level = 0.95))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = density ~ temp, data = Q1data)
##
## $temp
##                  diff         lwr         upr     p adj
## 125-100 -2.600000e-01 -0.44094678 -0.07905322 0.0040869
## 150-100  8.000000e-02 -0.10094678  0.26094678 0.5968856
## 175-100  3.552714e-15 -0.18094678  0.18094678 1.0000000
## 150-125  3.400000e-01  0.15905322  0.52094678 0.0003259
## 175-125  2.600000e-01  0.07905322  0.44094678 0.0040869
## 175-150 -8.000000e-02 -0.26094678  0.10094678 0.5968856
```

```r
plot(model1.tukey)
```



There are only three significant differences in this test and all involve the 125° temperature: 100° vs 125°, 100° vs 150°, and 125° vs 175°. Therefore, a reasonable conclusion is that temperatures of 100°, 150°, and 175° produce ceramics with similar densities, while temperature 125° produces ceramics with lower densities.

Question 2

In an experiment to study the effect of fertilizers on the spear elongation in asparagus, four different fertilizers and a control group (no fertilizer) were tested and five asparagus spears were measured for each treatment. The treatments are coded `trmt1`, `trmt2`, `trmt3`, `trmt4`, and the control `Ctrl`. The measurements (`length`) represent the length in mm of the asparagus spear. The data is in the file `spear`.

Do a complete analysis of variance for this set. Plot the data. Determine whether the treatments have an effect of the length of the asparagus spear by means of a hypothesis test. Plot the diagnostic charts and comment on them. Use also Levene's test and Shapiro-Wilk. Use Tukey's HSD procedure to make pairwise comparisons and comment on the results.

Solution

Read the data

```
spear <- read.table('spear', header = T)
str(spear)
```

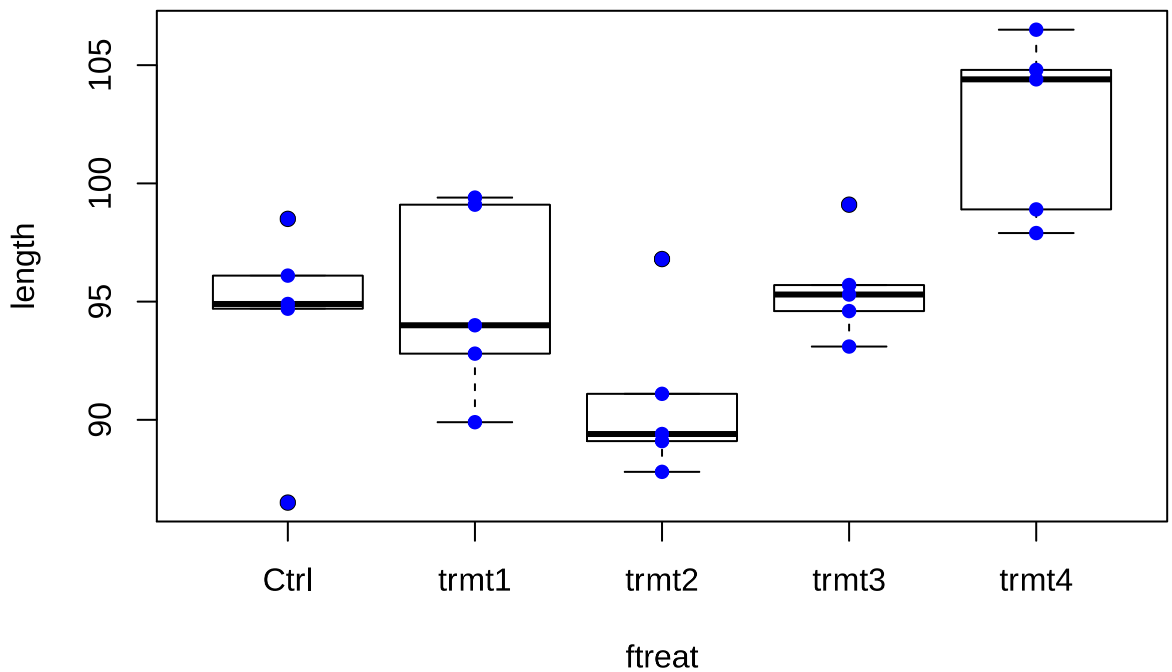
```
## 'data.frame':    25 obs. of  2 variables:
##  $ treat : Factor w/ 5 levels "Ctrl","trmt1",..: 1 1 1 1 1 2 2 2 2 2 ...
##  $ length: num  94.7 96.1 86.5 98.5 94.9 89.9 94 99.1 92.8 99.4 ...
```

We add a factor `ftreat` with the information in the variable `treat`.

```
spear$ftreat <- factor(spear$treat)
```

Plot the data:

```
plot(length ~ ftreat, data = spear)
points(length ~ ftreat, data = spear, pch = 16, col = 'blue')
```



Looking at the boxes, one may think that homoscedasticity may be a problem here. However, looking at the points, we see that the dispersion is similar for all treatments. We will test for this using Leven's test.

The model for the experiment is

$$y_{ik} = \mu + \tau_i + \epsilon_{ik}$$

with $1 \leq i \leq 5$, $1 \leq k \leq 5$. μ is the overall mean, τ_i represents the effect of treatment i, and ϵ_{ik} are the experimental errors, which are assumed to be independent Gaussian random variables with mean 0 and common variance σ^2. We fit this model with

```
model2 <- aov(length ~ ftreat, data = spear)
```

We produce an anova table.

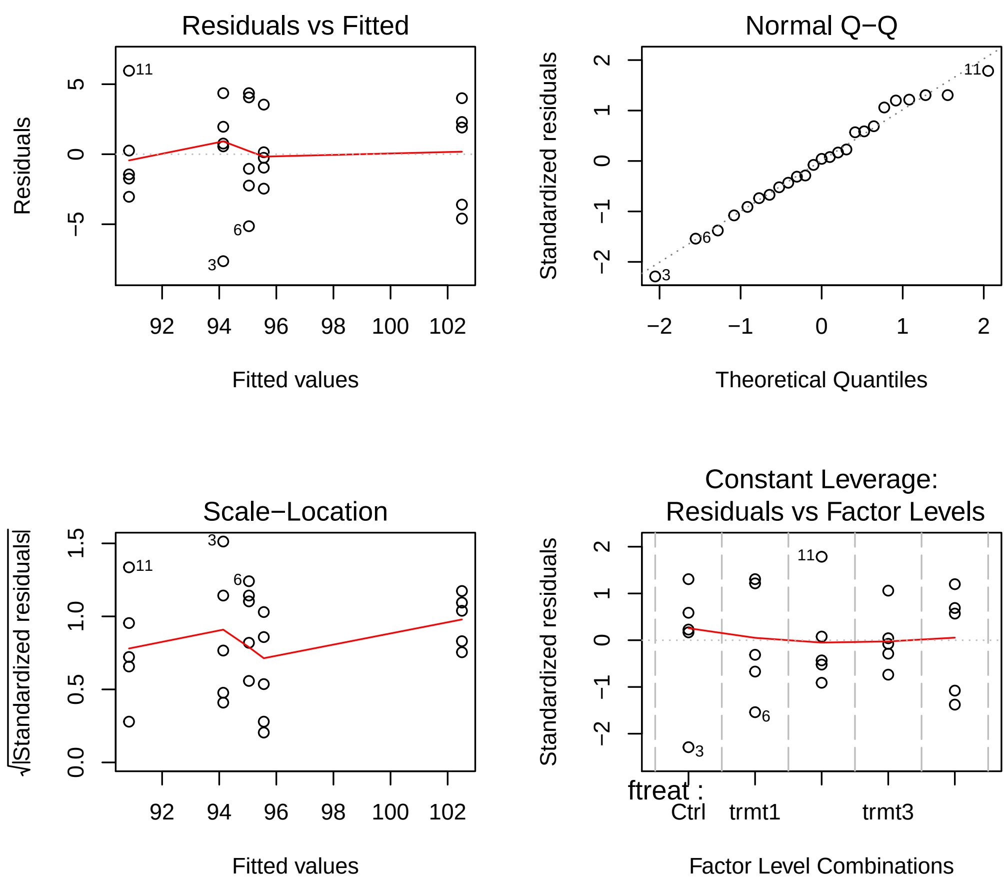
```r
anova(model2)
```

```
## Analysis of Variance Table
##
## Response: length
##           Df Sum Sq Mean Sq F value   Pr(>F)
## ftreat     4 363.57  90.891  6.5233 0.001576 **
## Residuals 20 278.67  13.933
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value is small and we reject the null hypothesis of no effect.

Diagnostic charts

```r
par(mfrow=c(2,2))
plot(model2)
```




```r
par(mfrow = c(1,1))
```

Again, the only concern here is whether the residuals are homoscedastic. The quantile plot indicates that the assumption of normality is satisfied. This is confirmed with a Shapiro-Wilk test:

```r
shapiro.test(resid(model2))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(model2)
```

```
## W = 0.98028, p-value = 0.8906
```

The p-value is large and we doo not reject the hypothesis of normality. To test for homoscedasticity we use Levene's test:

```
library(car)
leveneTest(model2)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  4  0.3062 0.8704
##       20
```

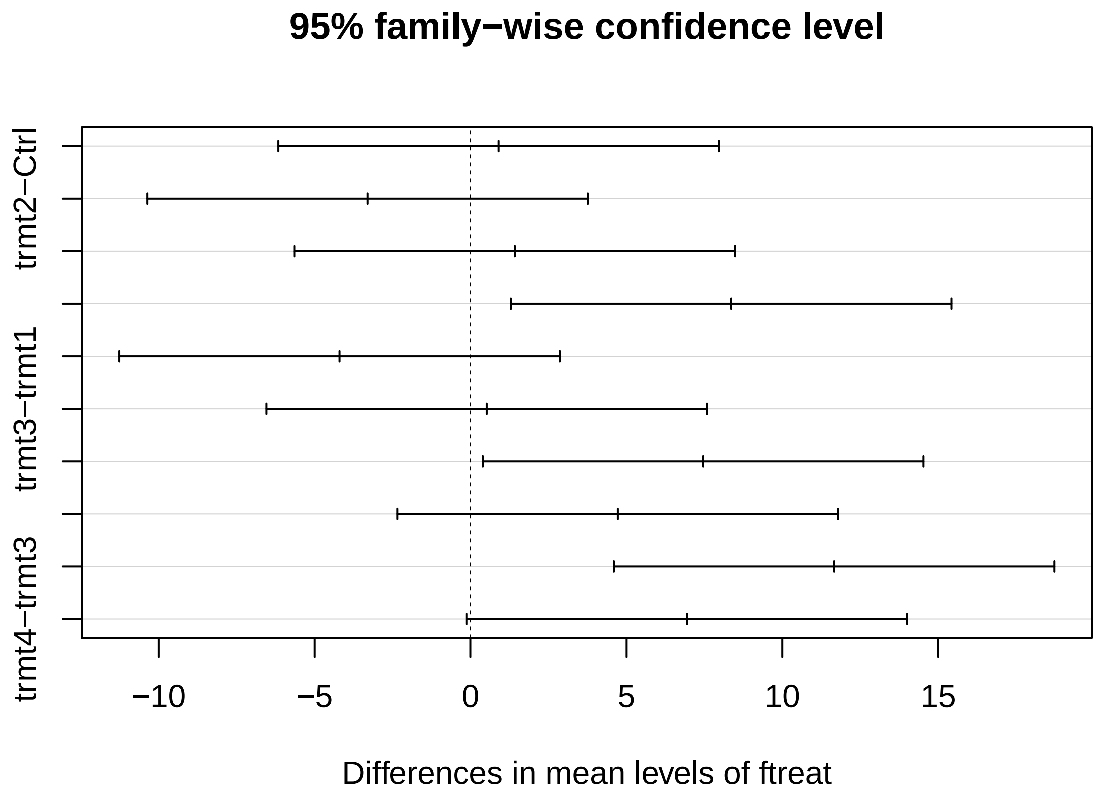
The p-value is large and homoscedasticity is not rejected. We conclude that the assumptions for the model are satisfied.

We can do pairwise comparisons using `TukeyHSD`. We will use a confidence level of 5%.

```
(model2.tukey <- TukeyHSD(model2, conf.level = 0.95))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = length ~ ftreat, data = spear)
##
## $ftreat
##              diff         lwr       upr     p adj
## trmt1-Ctrl   0.90  -6.1643873  7.964387 0.9951434
## trmt2-Ctrl  -3.30 -10.3643873  3.764387 0.6359591
## trmt3-Ctrl   1.42  -5.6443873  8.484387 0.9732262
## trmt4-Ctrl   8.36   1.2956127 15.424387 0.0155922
## trmt2-trmt1 -4.20 -11.2643873  2.864387 0.4121197
## trmt3-trmt1  0.52  -6.5443873  7.584387 0.9994290
## trmt4-trmt1  7.46   0.3956127 14.524387 0.0353277
## trmt3-trmt2  4.72  -2.3443873 11.784387 0.3019232
## trmt4-trmt2 11.66   4.5956127 18.724387 0.0006763
## trmt4-trmt3  6.94  -0.1243873 14.004387 0.0556639
```

```
plot(model2.tukey)
```



All the significant comparisons involve treatment 4, which is significantly different from the control and from treatments 1 and 2. The difference with treatment 3 has a p-value just above 0.05. From the graph we plotted at the beginning, we see that spear length is larger with treatment 4, and the tests support this conclusion, except for treatment 3.