Hw6-2022

2022-10-25

Question1

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

1. Do a complete analysis of variance for this set. Plot the data.
2. Write the equation for the model.
3. Determine whether the treatments have an effect on the amount of hemoglobin in blood by means of a hypothesis test.
4. 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.

```
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, 21.8, 21.6, 21.5,
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)
```

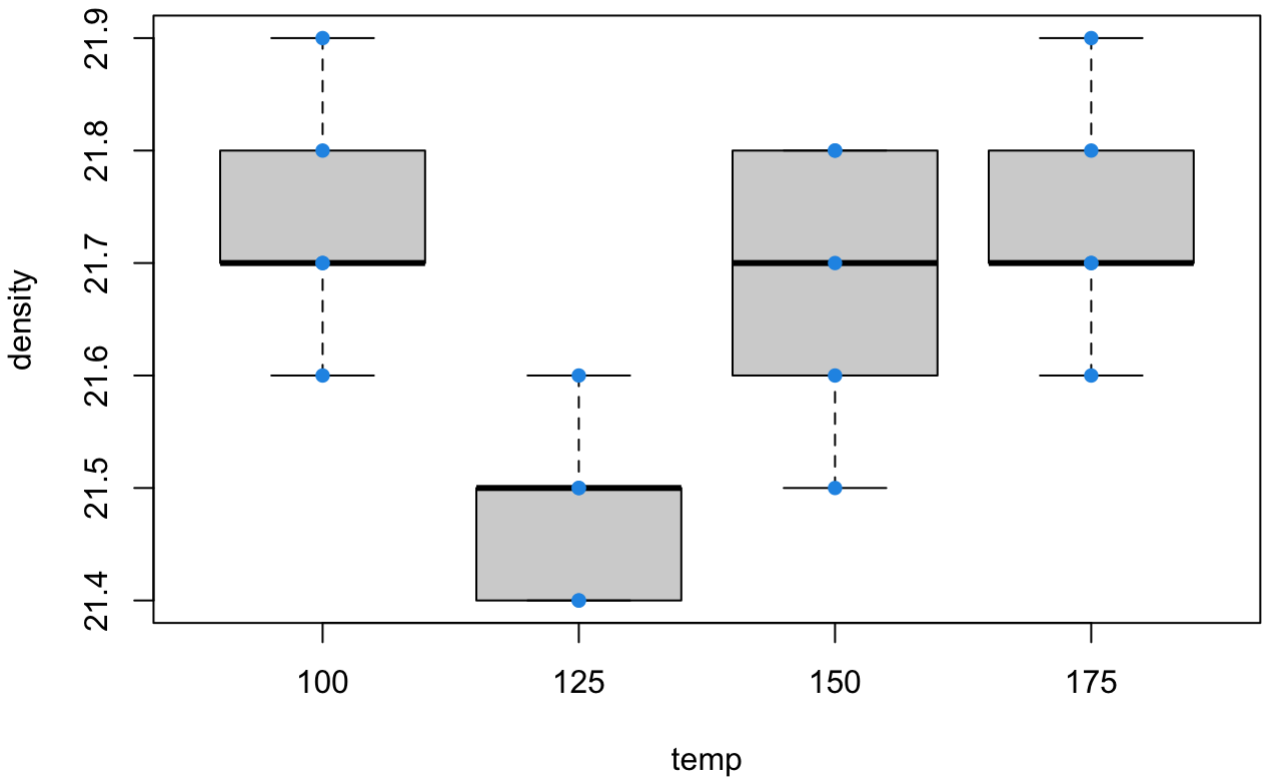
```
## 'data.frame':    20 obs. of  2 variables:
##  $ temp   : Factor w/ 4 levels "100","125","150",..: 1 1 1 1 1 2 2 2 2 2 ...
##  $ density: num  21.8 21.9 21.7 21.6 21.7 21.5 21.4 21.5 21.4 21.6 ...
```

From the formula for SST , we can obtain the total sum of squares by finding the differences between the data and the overall mean:

```
sum((density-mean(density))^2)
```

```
## [1] 0.428
```

```
boxplot(density~temp, data = Q1data)
points(density~temp, data = Q1data, pch=16,col=4)
```




```
mod1 = aov(density~temp, data = Q1data)
summary(mod1)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## temp         3  0.228  0.0760    6.08 0.0058 **
## Residuals   16  0.200  0.0125
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
print(mod1$coefficients)
```

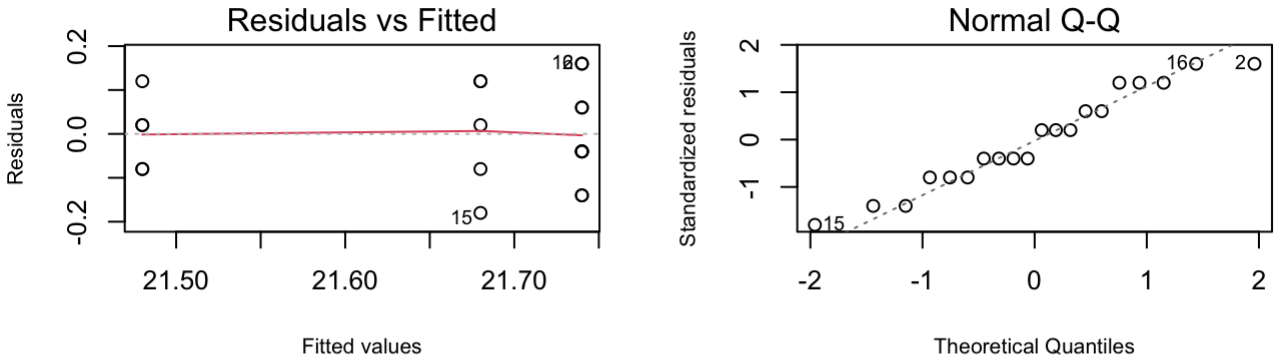
```
##   (Intercept)       temp125       temp150       temp175
##  2.174000e+01 -2.600000e-01 -6.000000e-02  2.641965e-15
```

p = 0.0058, alpha = 0.01, p < alpha, reject null hypothesis. We choose alternative that at least two of means are different. The treatments have an effect.

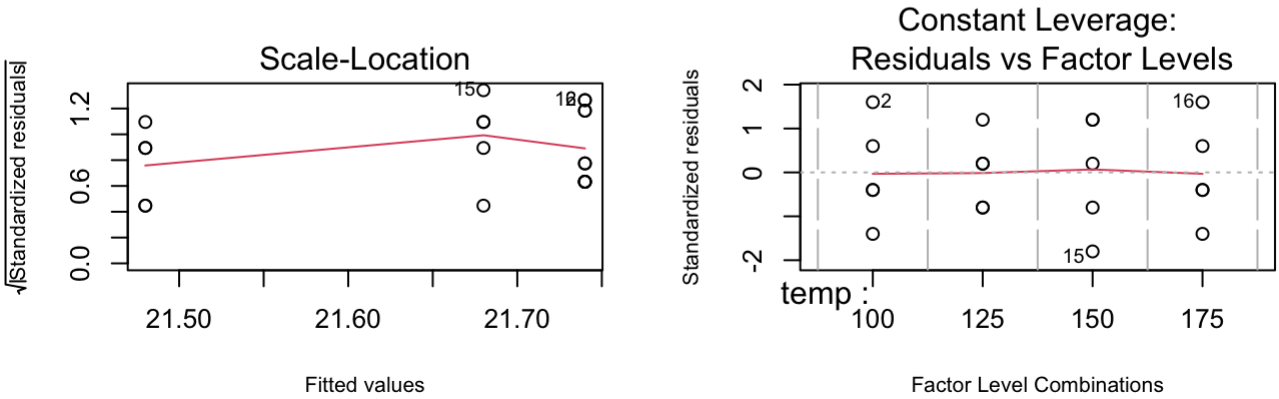
2. write the equation μ̂ + τ̂1 = 2.174000e+01 ; μ̂+τ̂2= 2.174000e+01 -2.600000e-01 = 1.914 μ̂+τ̂3 = 2.174000e+01 μ̂+τ̂4 = 2.174000e+01

Plot the diagnostic charts and comment on them.

```
par(mfrow=c(2,2))
plot(mod1,  cex.lab=0.8, cex.sub=0.8)
```



Since we



have 5 replications for each treatment level and there are only 5 x-values, the points appear vertically aligned at these values. We look in this graphs for constant variance. We see that values in some cases appear to be more spread than in others, and this may be a sign of non-constant variance. However, we only have a few points and this is difficult to determine.

Next, we do the Shapiro-Wilk and Levene tests for the residuals.

```
shapiro.test(resid(mod1))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(mod1)
## W = 0.94977, p-value = 0.3636
```

This shows that at the 5% level (or lower levels) we cannot reject the null hypothesis of Gaussianity.

```
library(car)
```

```
## Loading required package: carData
```

```
leveneTest(mod1)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  3  0.2424 0.8655
##       16
```

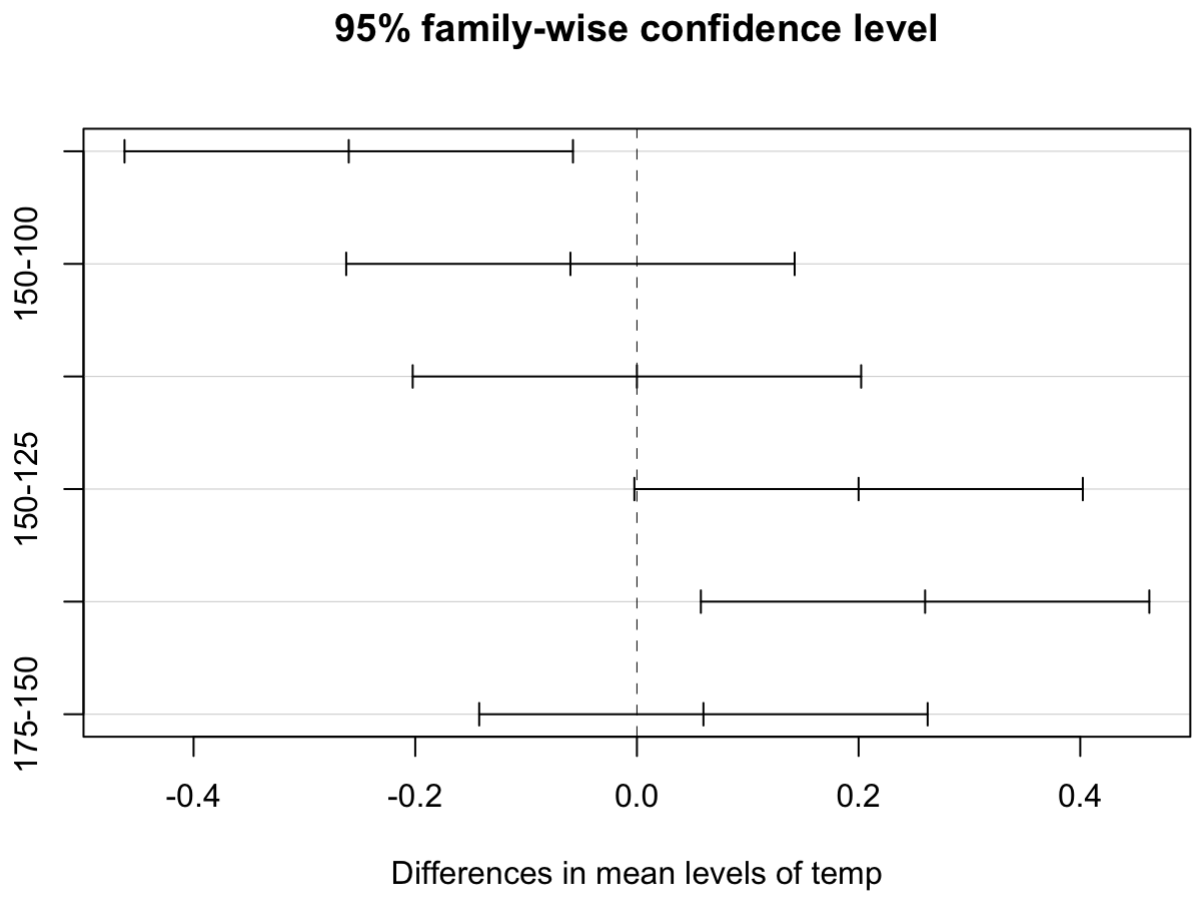
This test has a large p-value, saying that hypothesis of homoscedasticy is not rejected.

Use Tukey's HSD procedure to make pairwise comparisons and comment on the results.

```
(mod1.tky = TukeyHSD(mod1))
```

```
##   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 -0.462304652 -0.05769535 0.0098844
## 150-100 -6.000000e-02 -0.262304652  0.14230465 0.8306158
## 175-100  3.552714e-15 -0.202304652  0.20230465 1.0000000
## 150-125  2.000000e-01 -0.002304652  0.40230465 0.0532168
## 175-125  2.600000e-01  0.057695348  0.46230465 0.0098844
## 175-150  6.000000e-02 -0.142304652  0.26230465 0.8306158
```

```
plot(mod1.tky)
```



We get a table with six comparisons. we can see the value of difference. The 150-100, 175-100, 175-150,150-125 p value higher than 5%, so that they don't have significant difference. 125-100, 175-125 p value is smaller than 1%, they have significant difference.

If the intervals include zero, then the difference is not significant. so that 125-100, 175-125 have significant difference.

Question2

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) is the length in mm of the asparagus spear. The data is in the file spear.

1. Do a complete analysis of variance for this set. Plot the data.
2. Write the equation for the model.
3. Determine whether the treatments have an effect of the length of the asparagus spear by means of a hypothesis test.
4. 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.

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

```
## 'data.frame':    25 obs. of  2 variables:
##  $ treat : chr  "Ctrl" "Ctrl" "Ctrl" "Ctrl" ...
##  $ length: num  94.7 96.1 86.5 98.5 94.9 89.9 94 99.1 92.8 99.4 ...
```

```
results$treat = as.factor(treat)
```

From the formula for SST , we can obtain the total sum of squares by finding the differences between the data and the overall mean:

```
sum((length-mean(length))^2)
```

```
## [1] 642.2336
```

```
model1 = aov(length~treat)
summary(model1)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## treat        4  363.6   90.89   6.523 0.00158 **
## Residuals   20  278.7   13.93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

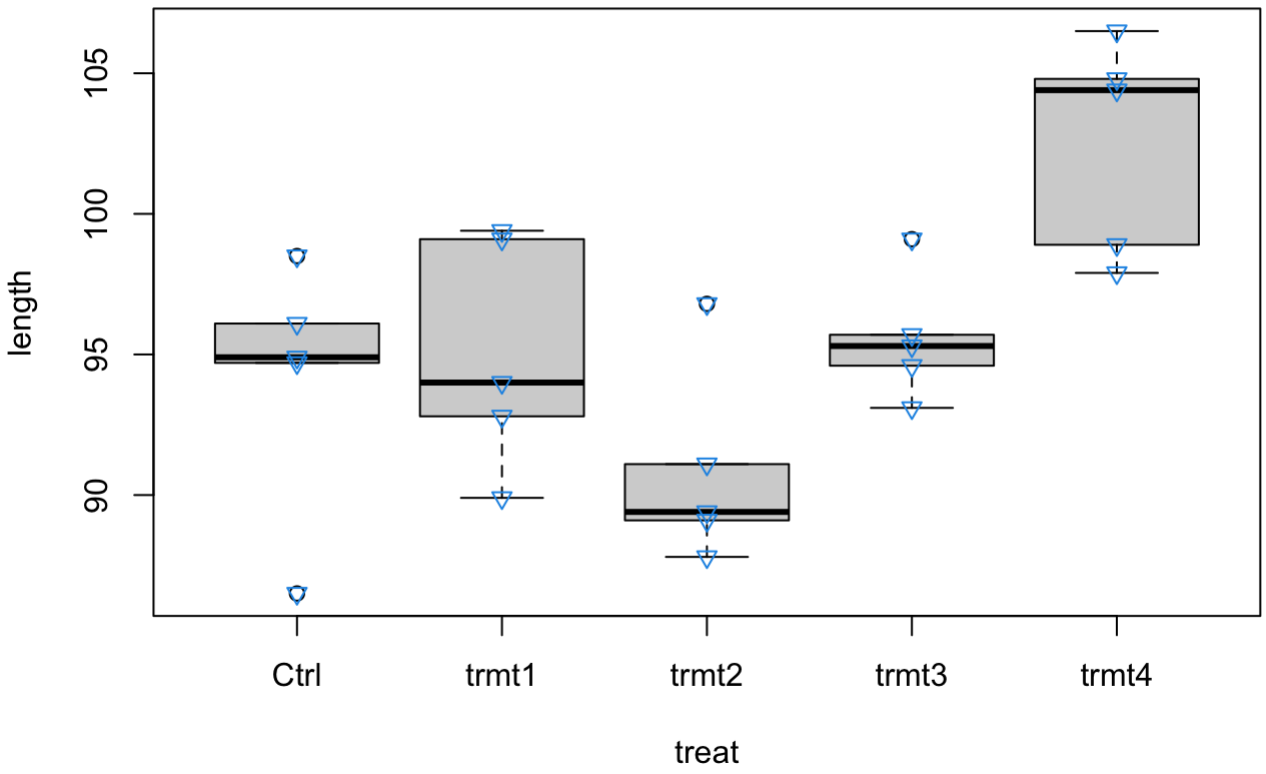
```
print(model1$coefficients)
```

```
## (Intercept)  treattrmt1  treattrmt2  treattrmt3  treattrmt4
##       94.14        0.90       -3.30        1.42        8.36
```

null hypothesis: the treatment means are not significantly different. p is smaller than 0.01, so we can reject the null hypothesis, the treatment means are significantly different.

Plot the data.

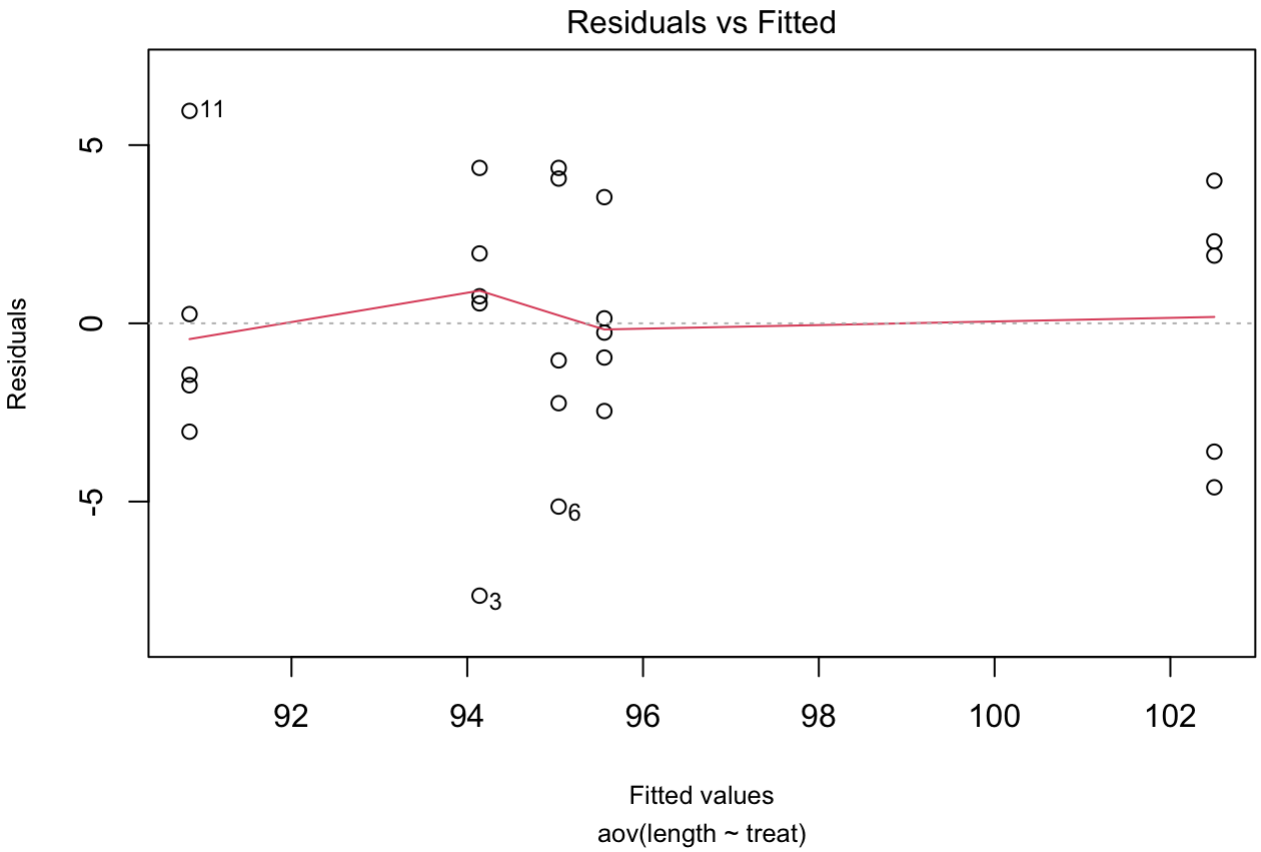
```
boxplot(length~treat, data = results)
points(length~treat, data = results, pch=25,col=4)
```



2. write the equation μ̂ + τ̂1 = 94.14 ; μ̂+τ̂2= 94.14+ 0.90 = 95.04 μ̂+τ̂3 = 94.14-3.30=90.84 μ̂+τ̂4 = 94.14 + 1.42 = 95.56 μ̂+τ̂5 = 94.14 + 8.36 =112.50

Look at diagnostic plots to check the assumptions of the model.

```
plot(model1, which=1, cex.lab=0.8, cex.sub=0.8)
```



Since we have 5 replications for each treatment level and there are only 5 x-values, the points appear vertically aligned at these values. We look in this graphs for constant variance. We see that values in some cases appear to be more spread than in others, and this may be a sign of non-constant variance. However, we only have a few points and this is difficult to determine.

Use also Levene's test and Shapiro-Wilk.

```
library(car)
shapiro.test(resid(model1))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(model1)
## W = 0.98028, p-value = 0.8906
```

```
leveneTest(model1)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

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

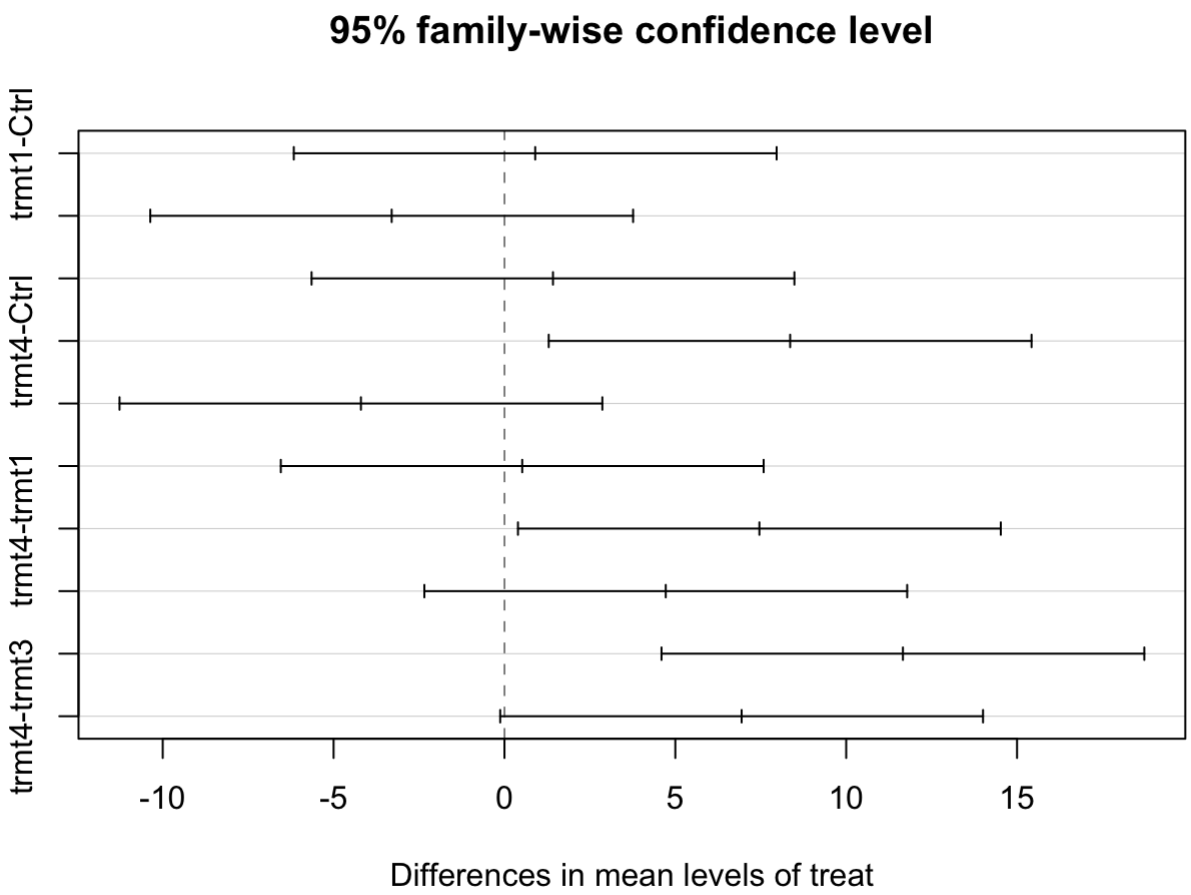
Shapiro-Wilk shows that at the 5% level (or lower levels) we cannot reject the null hypothesis of Gaussianity. leveneTest shows that has a large p-value, saying that hypothesis of homoscedasticy is not rejected.

Use Tukey's HSD procedure to make pairwise comparisons and comment on the results.

```
(model1.tky = TukeyHSD(model1))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = length ~ treat)
##
## $treat
##               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(model1.tky)
```



We get a table with ten comparisons. we can see the value of difference. If we choose alpha = 5%, The trmt4-Ctrl,trmt4-trmt1,trmt4-trmt2 p value smaller than 5%, so that they have significant difference. Others don't have significant difference. If we choose alpha = 2%, The trmt4-Ctrl, trmt4-trmt2 p value smaller than 2%,so that they have significant difference.Others don't have significant difference. If we choose alpha = 1%, The trmt4-trmt2 p value smaller than 1%,so that they have significant difference. Others don't have significant difference.

If the intervals include zero, then the difference is not significant. From figure we can see that trmt4-Ctrl,trmt4-trmt1,trmt4-trmt2 have significant difference.