homework_003

Mike James

March 1, 2019

Question I

A fixed factor is one in which the levels are set *before* the experiment and we are looking to see what differences occur between the differing levels.

A random factor is one that represents the population as a whole; meaning, the factor can be just about anything as opposed to a set level.

They both share the fact that they are linear models, but obviously, a fixed factor represents the population average while the random factor represents individual effects.

An example of a fixed factor is drug testing. There is the control, the $\frac{1}{2}$ dose, and a full dose. Each level is set *before* the experiment and we are looking to see how each levels affects the outcomes.

An example of a random factor is the bear weight based on location example from class. We do not have any specific quantities to compare, so we don't know how they are correlated.

Question II

a. Determine if adding different sugars to the growth media in Drosophila has an effect in the development of ocular units.

b. Provide the mean treatment effect that each factor level has, ie, individual α_i's. What can you conclude from that information?

Justification

H_0: Sugars have no effect of the number of ocular units
H_a: Sugars do have an effect on the number of ocular units

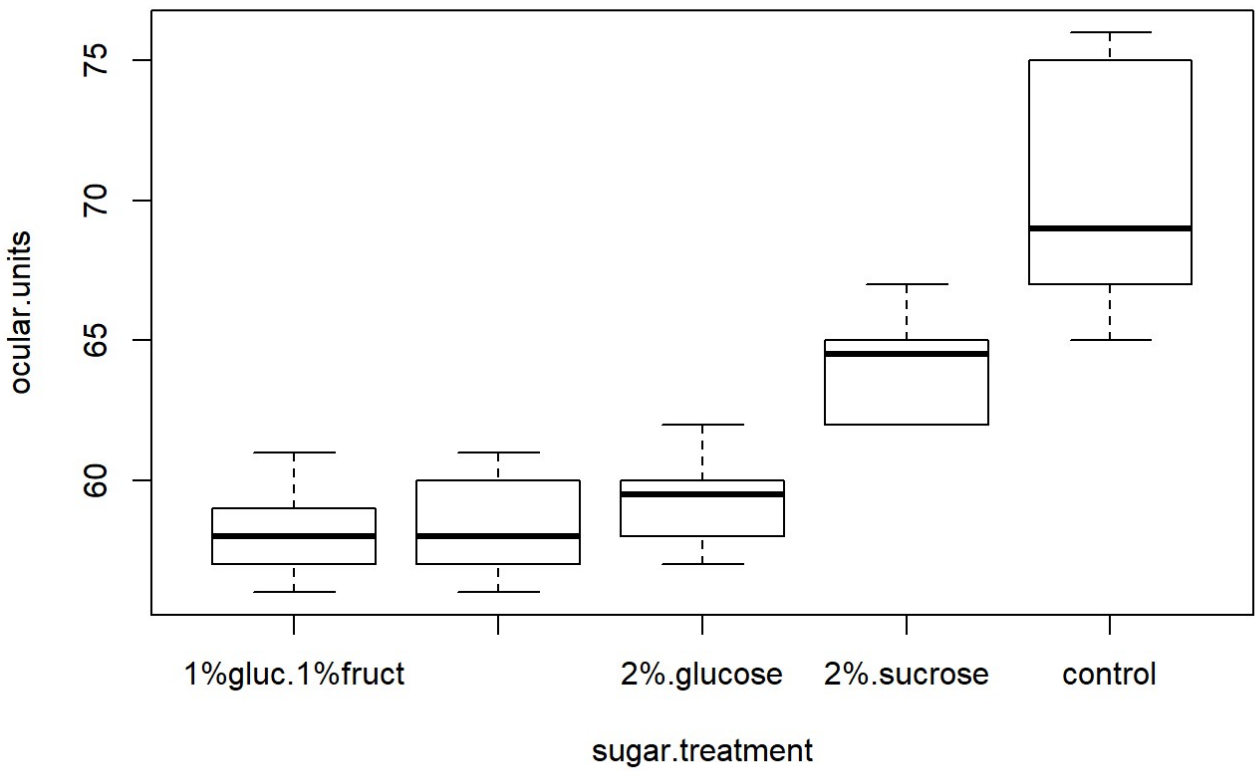
Looking at a simple boxplot of the data, the control has a relatively high number of ocular units compared to **any** of the factors tested for. The 2% sucrose treatment appears to be the closest as far as the number of ocular units, while all other treatments appear to have about the same effect – and much lower ocular units.

Results

```
data2 <- read.delim("h3p2.txt")
attach(data2)
summary(data2)
```

```
##        sugar.treatment  ocular.units
##  1%gluc.1%fruct:10     Min.   :56.00
##  2%.fructose   :10     1st Qu.:58.00
##  2%.glucose    :10     Median :60.50
##  2%.sucrose    :10     Mean   :61.94
##  control       :10     3rd Qu.:65.00
##                        Max.   :76.00
```

```
plot(ocular.units~sugar.treatment)
```




```
anova(lm(ocular.units~sugar.treatment))
```

```
## Analysis of Variance Table
##
## Response: ocular.units
##                 Df Sum Sq Mean Sq F value    Pr(>F)
## sugar.treatment  4 1077.3 269.330  49.368 6.737e-16 ***
## Residuals       45  245.5   5.456
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
data2.lm <- lm(ocular.units~sugar.treatment)
summary(data2.lm)
```

```
##
## Call:
## lm(formula = ocular.units ~ sugar.treatment)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.100 -1.825 -0.150  0.975  5.900
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  58.0000     0.7386  78.525  < 2e-16 ***
## sugar.treatment2%.fructose    0.2000     1.0446   0.191    0.849
## sugar.treatment2%.glucose     1.3000     1.0446   1.245    0.220
## sugar.treatment2%.sucrose     6.1000     1.0446   5.840 5.40e-07 ***
## sugar.treatmentcontrol       12.1000     1.0446  11.584 4.27e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.336 on 45 degrees of freedom
## Multiple R-squared:  0.8144, Adjusted R-squared:  0.7979
## F-statistic: 49.37 on 4 and 45 DF,  p-value: 6.737e-16
```

```
data2.aov <- aov(ocular.units~sugar.treatment)
TukeyHSD(data2.aov)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = ocular.units ~ sugar.treatment)
##
## $sugar.treatment
##                              diff       lwr       upr     p adj
## 2%.fructose-1%gluc.1%fruct    0.2 -2.768072  3.168072 0.9996878
## 2%.glucose-1%gluc.1%fruct     1.3 -1.668072  4.268072 0.7256157
## 2%.sucrose-1%gluc.1%fruct     6.1  3.131928  9.068072 0.0000052
## control-1%gluc.1%fruct       12.1  9.131928 15.068072 0.0000000
## 2%.glucose-2%.fructose        1.1 -1.868072  4.068072 0.8291029
## 2%.sucrose-2%.fructose        5.9  2.931928  8.868072 0.0000100
## control-2%.fructose          11.9  8.931928 14.868072 0.0000000
## 2%.sucrose-2%.glucose         4.8  1.831928  7.768072 0.0003242
## control-2%.glucose           10.8  7.831928 13.768072 0.0000000
## control-2%.sucrose            6.0  3.031928  8.968072 0.0000072
```

```
summary(lm(data2.aov))
```

```
##
## Call:
## lm(formula = data2.aov)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.100 -1.825 -0.150  0.975  5.900
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  58.0000     0.7386  78.525  < 2e-16 ***
## sugar.treatment2%.fructose    0.2000     1.0446   0.191    0.849
## sugar.treatment2%.glucose     1.3000     1.0446   1.245    0.220
## sugar.treatment2%.sucrose     6.1000     1.0446   5.840 5.40e-07 ***
## sugar.treatmentcontrol       12.1000     1.0446  11.584 4.27e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.336 on 45 degrees of freedom
## Multiple R-squared:  0.8144, Adjusted R-squared:  0.7979
## F-statistic: 49.37 on 4 and 45 DF,  p-value: 6.737e-16
```

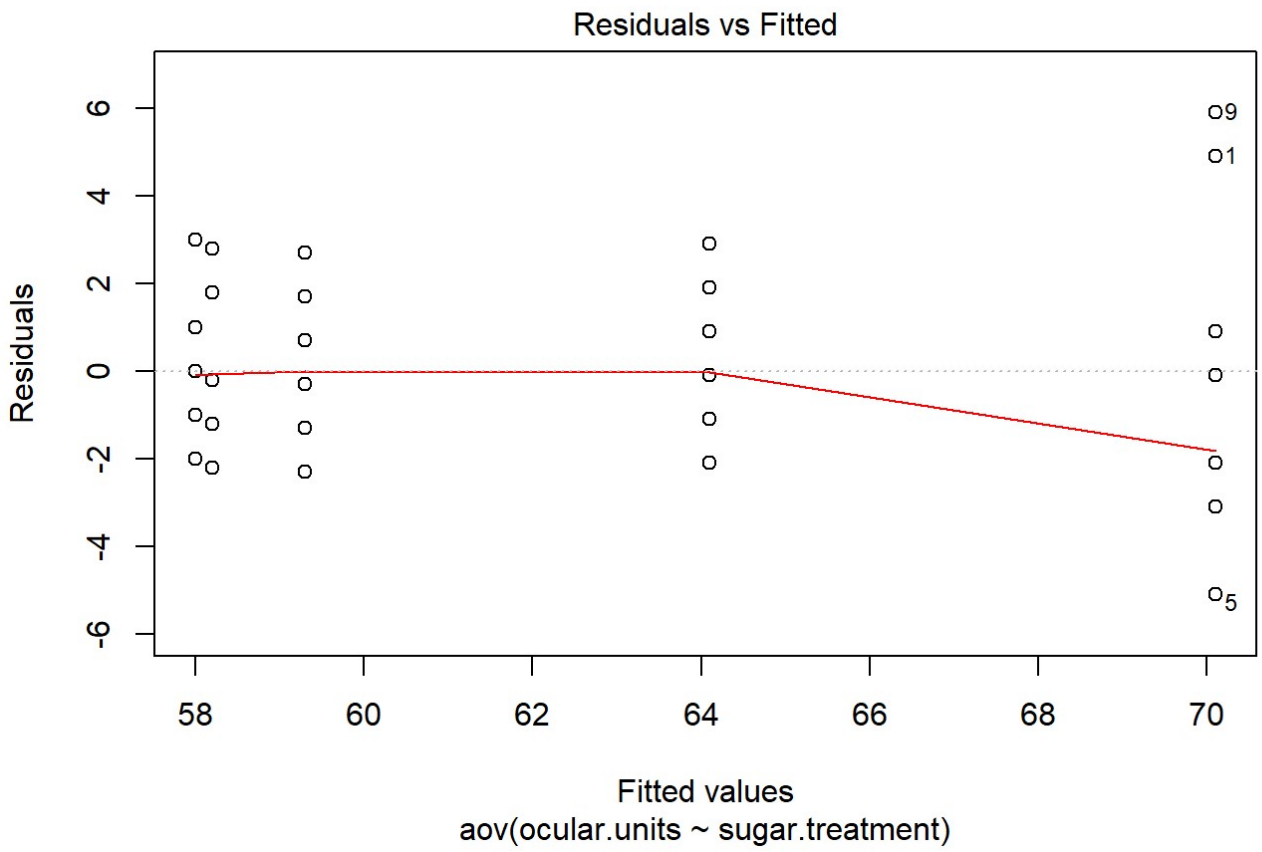
```
bartlett.test(ocular.units~sugar.treatment)
```

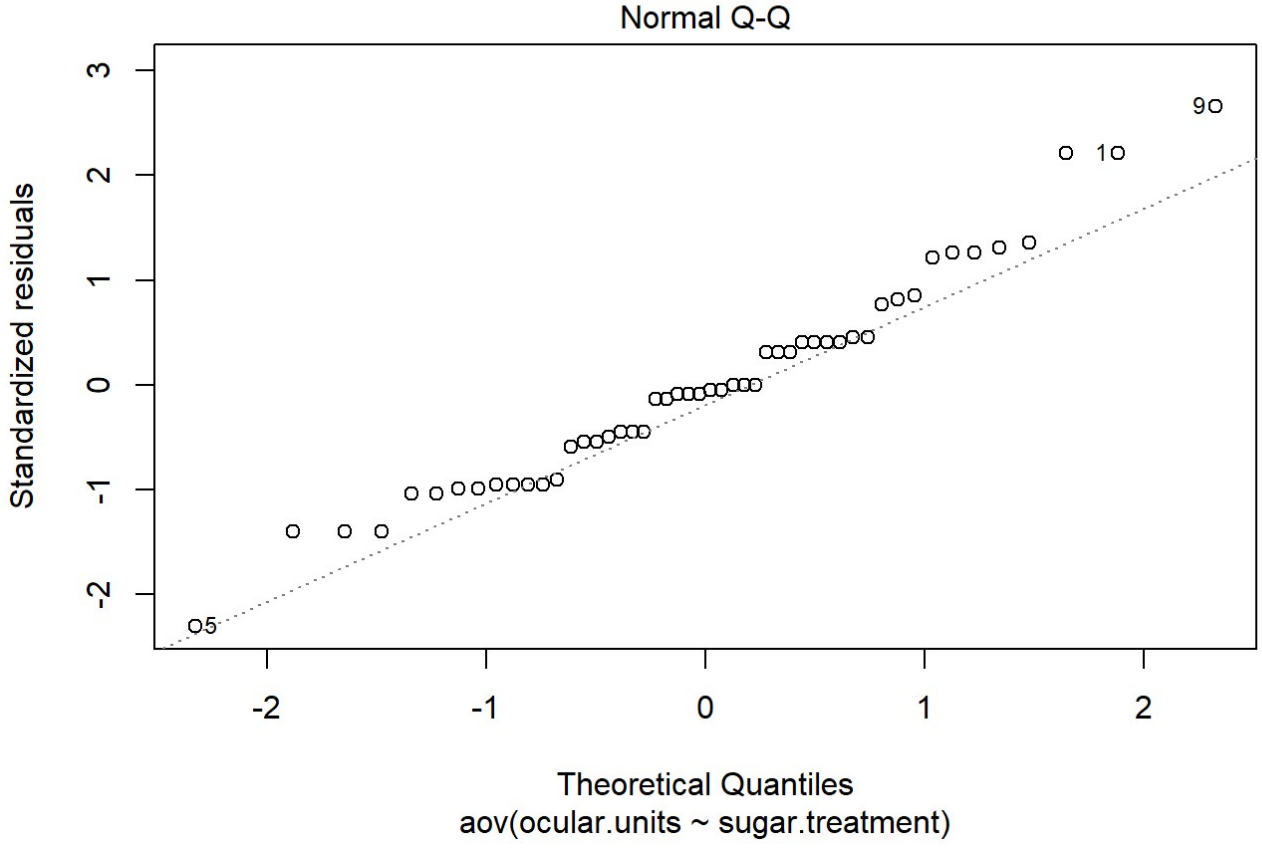
```
##
##  Bartlett test of homogeneity of variances
##
## data:  ocular.units by sugar.treatment
## Bartlett's K-squared = 13.939, df = 4, p-value = 0.007494
```

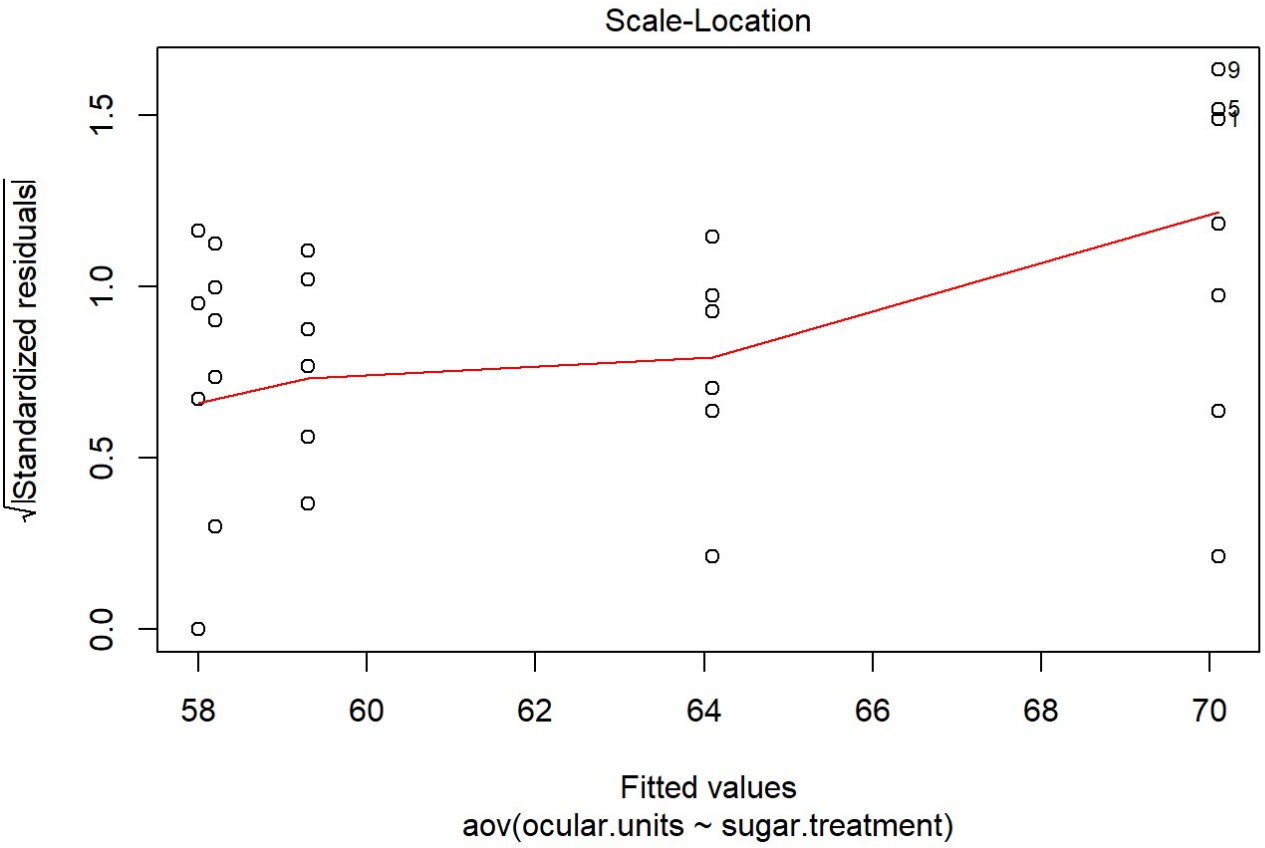
```
shapiro.test(data2.aov$residuals)
```

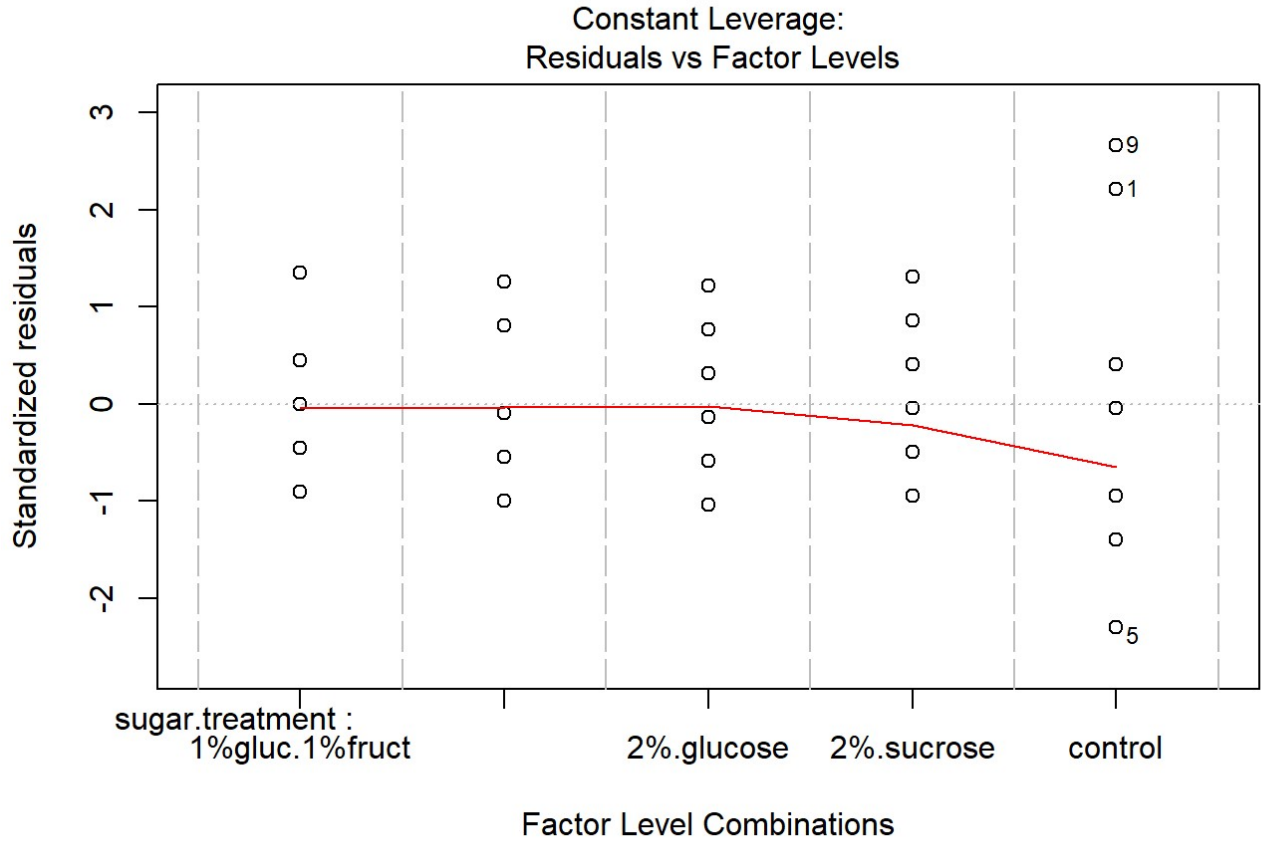
```
##
##  Shapiro-Wilk normality test
##
## data:  data2.aov$residuals
## W = 0.96928, p-value = 0.2164
```

```
plot(data2.aov)
```


Residuals vs Fitted
Residuals
6
4
2
0
-2
-4
-6
9
1
5
58
60
62
64
66
68
70
Fitted values
aov(ocular.units ~ sugar.treatment)


Normal Q-Q
Standardized residuals
3
2
1
0
-1
-2
9
10
5
-2
-1
0
1
2
Theoretical Quantiles
aov(ocular.units ~ sugar.treatment)


Scale-Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
58
60
62
64
66
68
70
Fitted values
aov(ocular.units ~ sugar.treatment)


Constant Leverage:
Residuals vs Factor Levels
Standardized residuals
-2
-1
0
1
2
3
sugar.treatment :
1%gluc.1%fruct
2%.glucose
2%.sucrose
control
Factor Level Combinations

```
detach(data2)
```

Conclusion

Based on the results from ANOVA, we see that F(4,45) = 49.37; p < 0.000, showing significant effects from the use of 2% sucrose compared to the other treatments. The Scale-Location graph shows residuals increase with the fitted values; therefore, we reject the null hypothesis.

As for why there are **fewer** ocular units **all** sugar treatments, I have no idea, it's not my data. Obviously, there is some other factor affecting the number of ocular units, but 2% sucrose does lead to more ocular units than the other treatments.

Looking at Tukey's HSD, we can ignore "2%.fructose-1%gluc.1%fruct", "2%.glucose-1%gluc.1%fruct", and "2%.glucose-2%.fructose" as they are not statistically significant. As for the rest, glucose vs. sucrose have very similar means, the rest range from ~5-6, with those compared to the control having a 10.8 - 12.1 difference of means. The closest to the control would be the 2% sucrose, as noted earlier.

Question III

The Australian brown tree snake was introduced in Guam and has cause the extinction of two native birds. This snake seems to reproduce more often and less seasonally than in its native range in Australia.

The data provided represents the grams of fat in the sexual segment of the kidney, which is a good indication of sperm production in males.

All of the data was collected in the non-breeding.

Determine if there are any differences among the populations.

Justification

H_0: Location has no effect on the fat levels
H_a: Location does have an effect on the fat levels

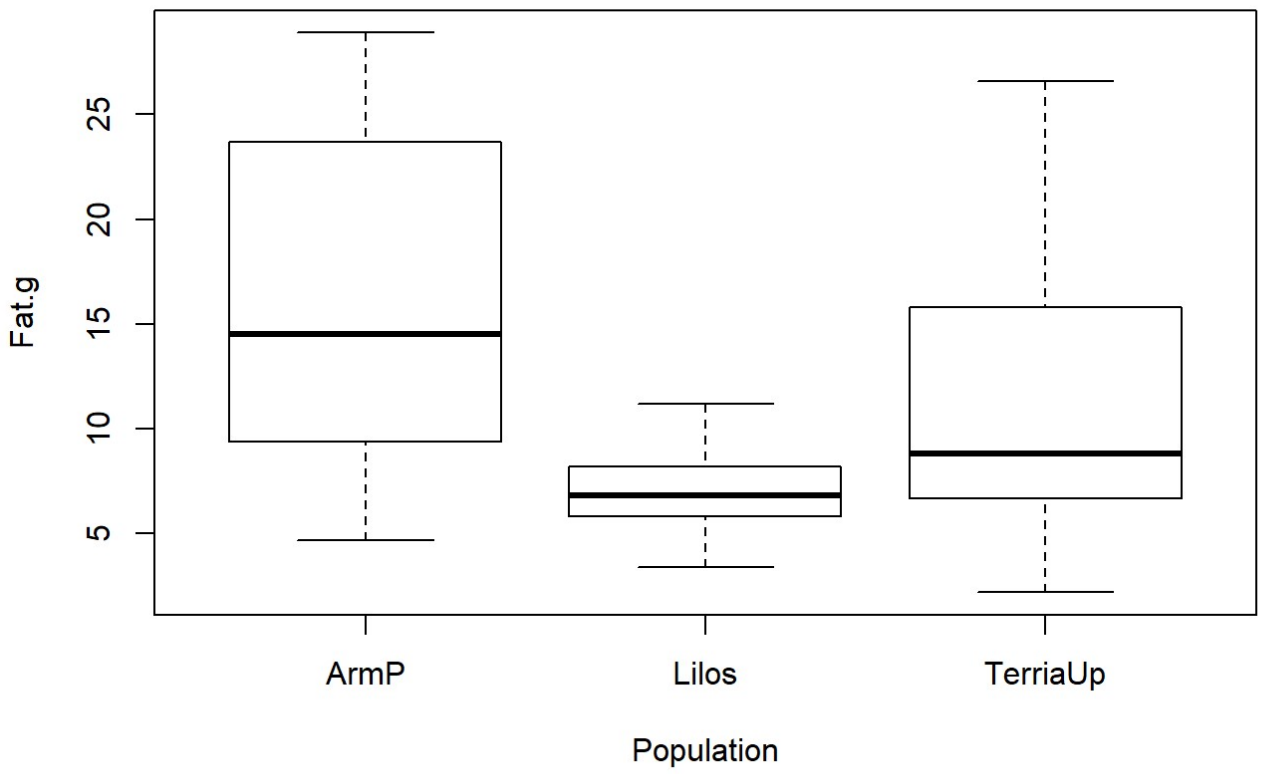
Looking at the boxplot, it would appear that the ArmP location has the highest fat content, and thus, wowuld indicate higher levels of sperm production; at the same time, TerriaUp snakes have a similar range, but a lower mean. Meanwhile, those snakes from the Lilos group have a relatively small range of fat content that is far lower than the previous two locations.

Results

```
data3 <- read.delim("h3p3.txt")
attach(data3)
summary(data3)
```

```
##     Population     Fat.g
##  ArmP    :18   Min.   : 2.20
##  Lilos   :24   1st Qu.: 6.45
##  TerriaUp:21   Median : 8.50
##                Mean   :11.18
##                3rd Qu.:14.30
##                Max.   :28.90
```

```
plot(Fat.g~Population)
```




```
anova(lm(Fat.g~Population))
```

```
## Analysis of Variance Table
##
## Response: Fat.g
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## Population   2  852.04  426.02  11.105 7.887e-05 ***
## Residuals   60 2301.86   38.36
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
data3.aov <- aov(Fat.g~Population)
TukeyHSD(data3.aov)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Fat.g ~ Population)
##
## $Population
##                      diff         lwr        upr     p adj
## Lilos-ArmP     -9.065278 -13.7065799 -4.4239757 0.0000471
## TerriaUp-ArmP  -4.484921  -9.2661928  0.2963516 0.0703636
## TerriaUp-Lilos  4.580357   0.1325284  9.0281859 0.0421695
```

```
summary(lm(data3.aov))
```

```
##
## Call:
## lm(formula = data3.aov)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.4278  -3.5929  -0.8625   2.8973  14.9571
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          16.128      1.460  11.047 4.31e-16 ***
## PopulationLilos      -9.065      1.931  -4.694 1.60e-05 ***
## PopulationTerriaUp   -4.485      1.990  -2.254   0.0278 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.194 on 60 degrees of freedom
## Multiple R-squared:  0.2702, Adjusted R-squared:  0.2458
## F-statistic:  11.1 on 2 and 60 DF,  p-value: 7.887e-05
```

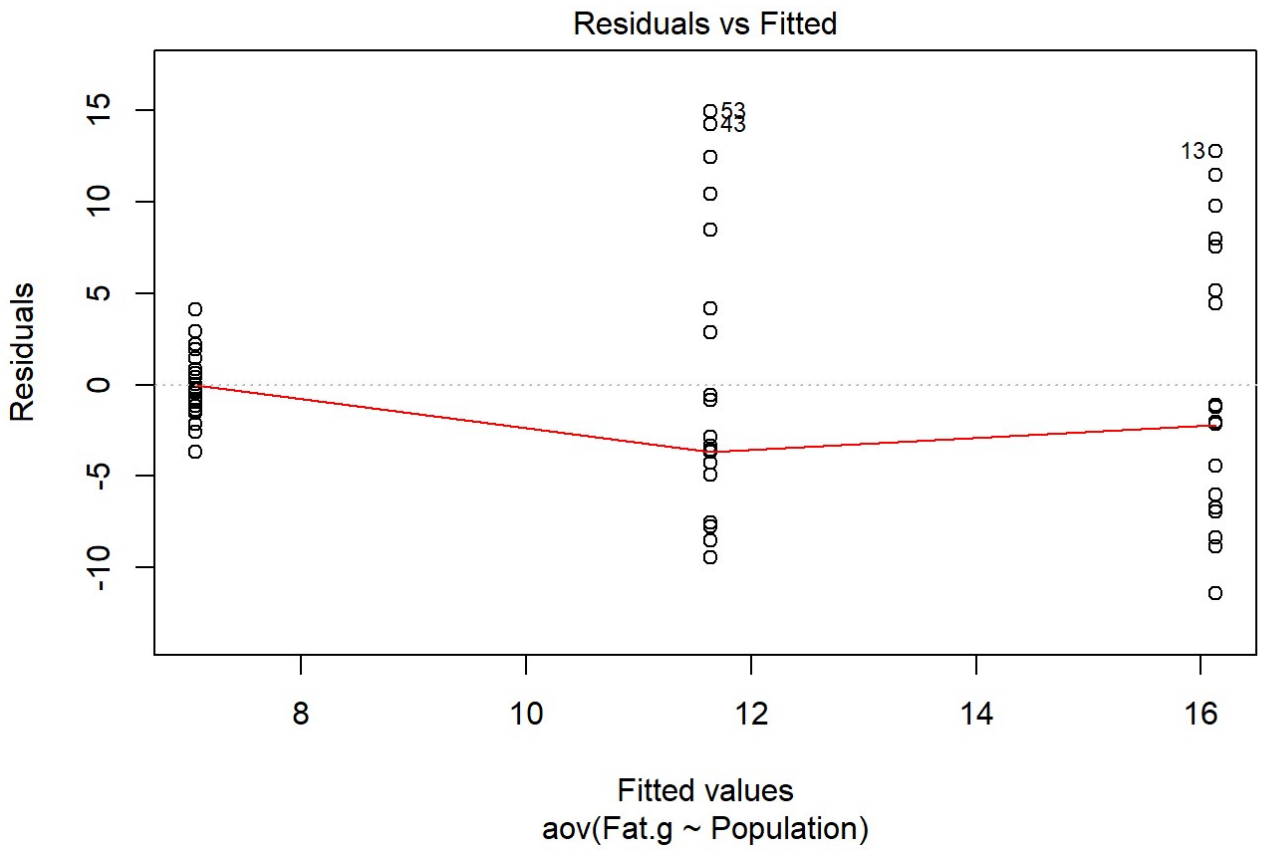
```
bartlett.test(Fat.g~Population)
```

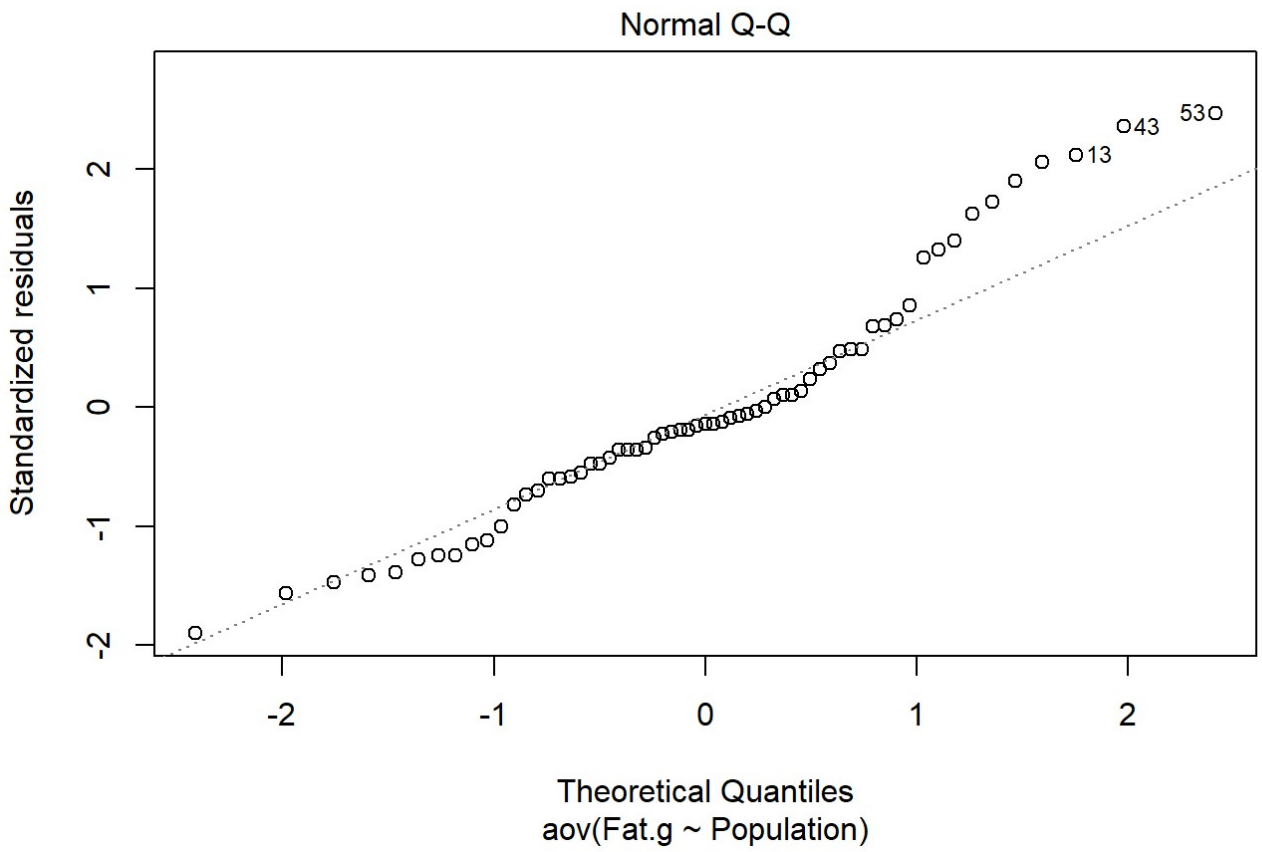
```
##
##  Bartlett test of homogeneity of variances
##
## data:  Fat.g by Population
## Bartlett's K-squared = 36.628, df = 2, p-value = 1.113e-08
```

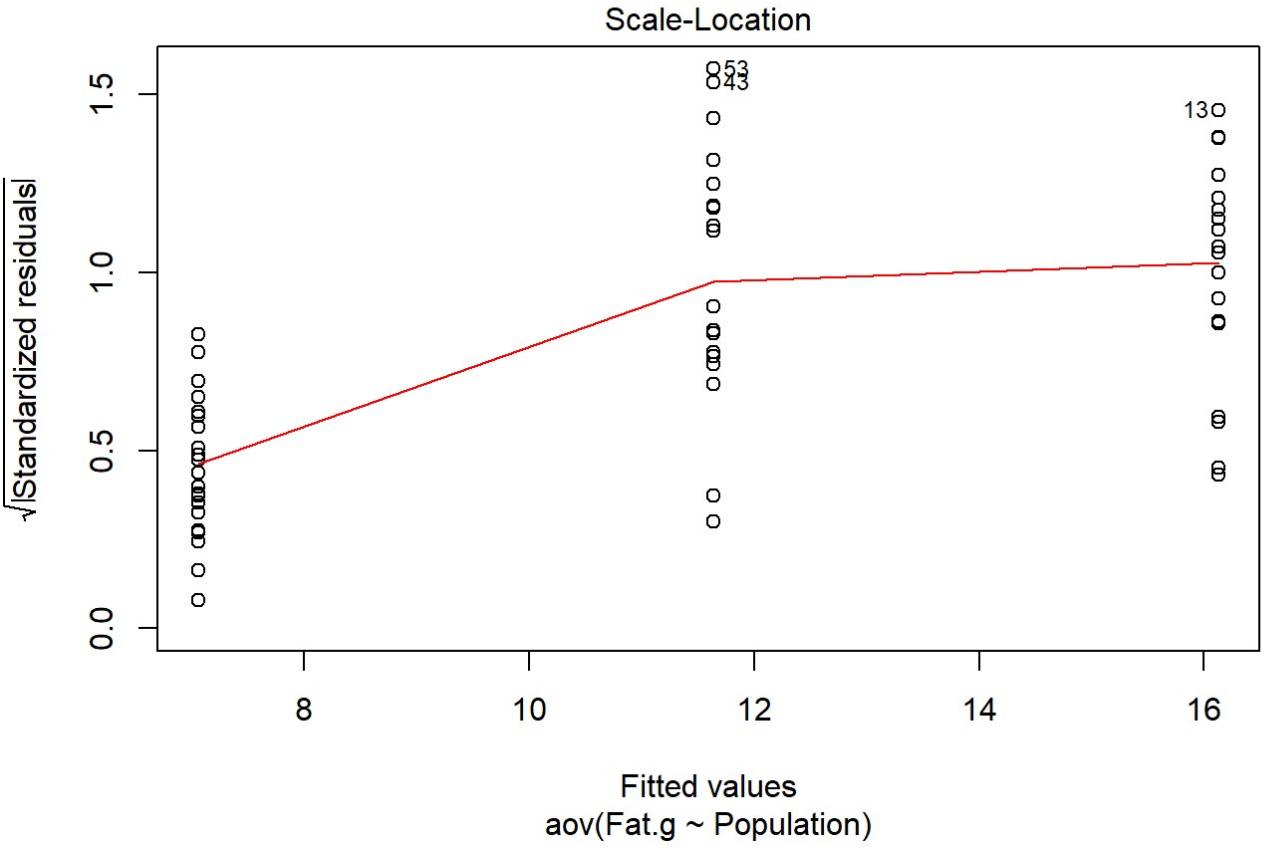
```
shapiro.test(data3.aov$residuals)
```

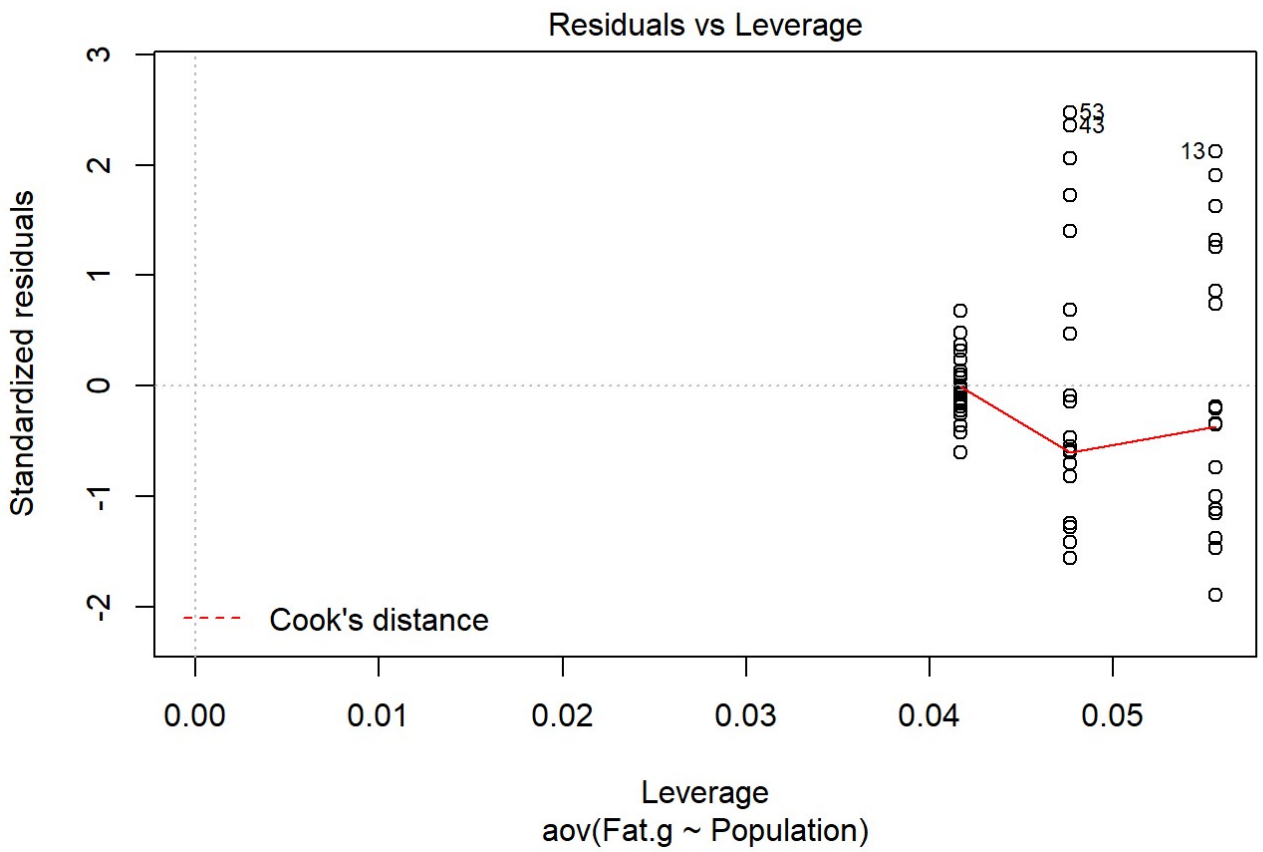
```
##
##  Shapiro-Wilk normality test
##
## data:  data3.aov$residuals
## W = 0.95147, p-value = 0.01458
```

```
plot(data3.aov)
```




Normal Q-Q
Standardized residuals
Theoretical Quantiles
aov(Fat.g ~ Population)
53
43
13


Scale-Location
√|Standardized residuals|
Fitted values
aov(Fat.g ~ Population)
53
43
13




```
detach(data3)
```

Conclusion

Looking at the Normal Q-Q plot, those residuals falling into the 1st and 2nd quantile are higher than expected. The scale-location plot shows the residuals increase with the fitted data.

As for the ANOVA, we have an $F(2,60) = 11.10$; $p < 0.00$, meaning we can reject the null hypothesis and can state that location does play a role in fat content and a likely effect on sexual reproduction.

I am unable to say for sure what each of the locations represents (none of the three specify any country (Guam, Australia, etc.); I could look them up, but again, it's not my data and was not supplied…I'm just a number cruncher).

Question IV

Wood density, and to a great extent hardness, is mostly determined by the number of pores/mm^2 in the wood.

Determine which of the three species has the greatest hardness.

Which one the least?

Justification

H_0: pore density has no effect of the hardness of the wood
H_a: Pore density does have an effect of the hardness of the wood

Note: the data merely states SPP and Cells, not pores/mm^2; I am going to assume – based on Camilo's lecture this morning (where he discussed Teak being extremely hard) – that we are looking for lower values of "cells."

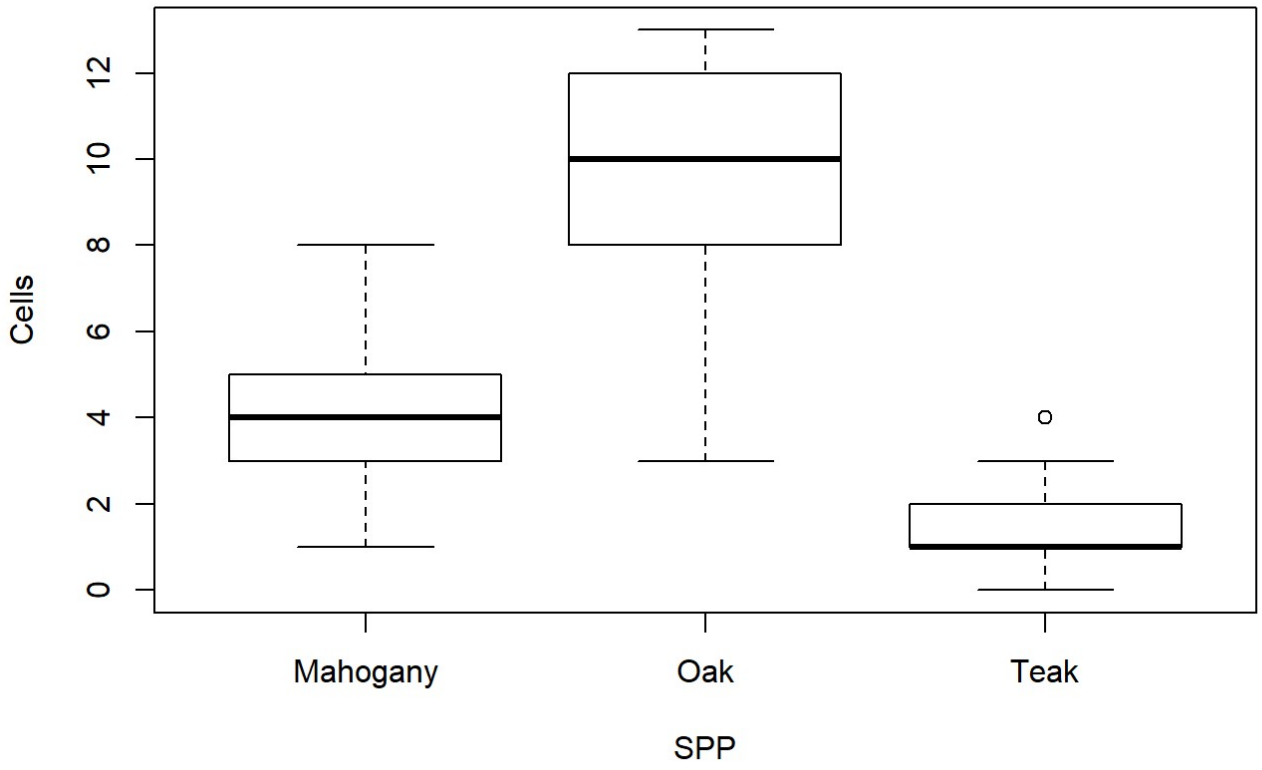
The boxplot of the raw data shows Oak having thy highest number of cells, with mahogany a distant second and teak coming in with approximately half the cells as mahogany.

Results

```
data4 <- read.delim("h3p4.txt")
attach(data4)
summary(data4)
```

```
##        SPP         Cells      
##  Mahogany:25   Min.   : 0.000  
##  Oak     :25   1st Qu.: 2.000  
##  Teak    :25   Median : 4.000  
##                Mean   : 4.947  
##                3rd Qu.: 8.000  
##                Max.   :13.000
```

```
plot(Cells~SPP)
```




```
anova(lm(Cells~SPP))
```

```
## Analysis of Variance Table
##
## Response: Cells
##           Df Sum Sq Mean Sq F value    Pr(>F)
## SPP        2 752.83  376.41   80.91 < 2.2e-16 ***
## Residuals 72 334.96    4.65
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
data4.aov <- aov(Cells~SPP)
TukeyHSD(data4.aov)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Cells ~ SPP)
##
## $SPP
##                 diff       lwr        upr     p adj
## Oak-Mahogany   5.16  3.700042  6.6199581 0.0000000
## Teak-Mahogany -2.44 -3.899958 -0.9800419 0.0004421
## Teak-Oak      -7.60 -9.059958 -6.1400419 0.0000000
```

```
summary(lm(data4.aov))
```

```
##
## Call:
## lm(formula = data4.aov)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -6.20  -0.82  -0.04   0.96   3.96
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.0400     0.4314   9.365 4.38e-14 ***
## SPPOak        5.1600     0.6101   8.458 2.14e-12 ***
## SPPTeak      -2.4400     0.6101  -4.000 0.000152 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.157 on 72 degrees of freedom
## Multiple R-squared:  0.6921, Adjusted R-squared:  0.6835
## F-statistic: 80.91 on 2 and 72 DF,  p-value: < 2.2e-16
```

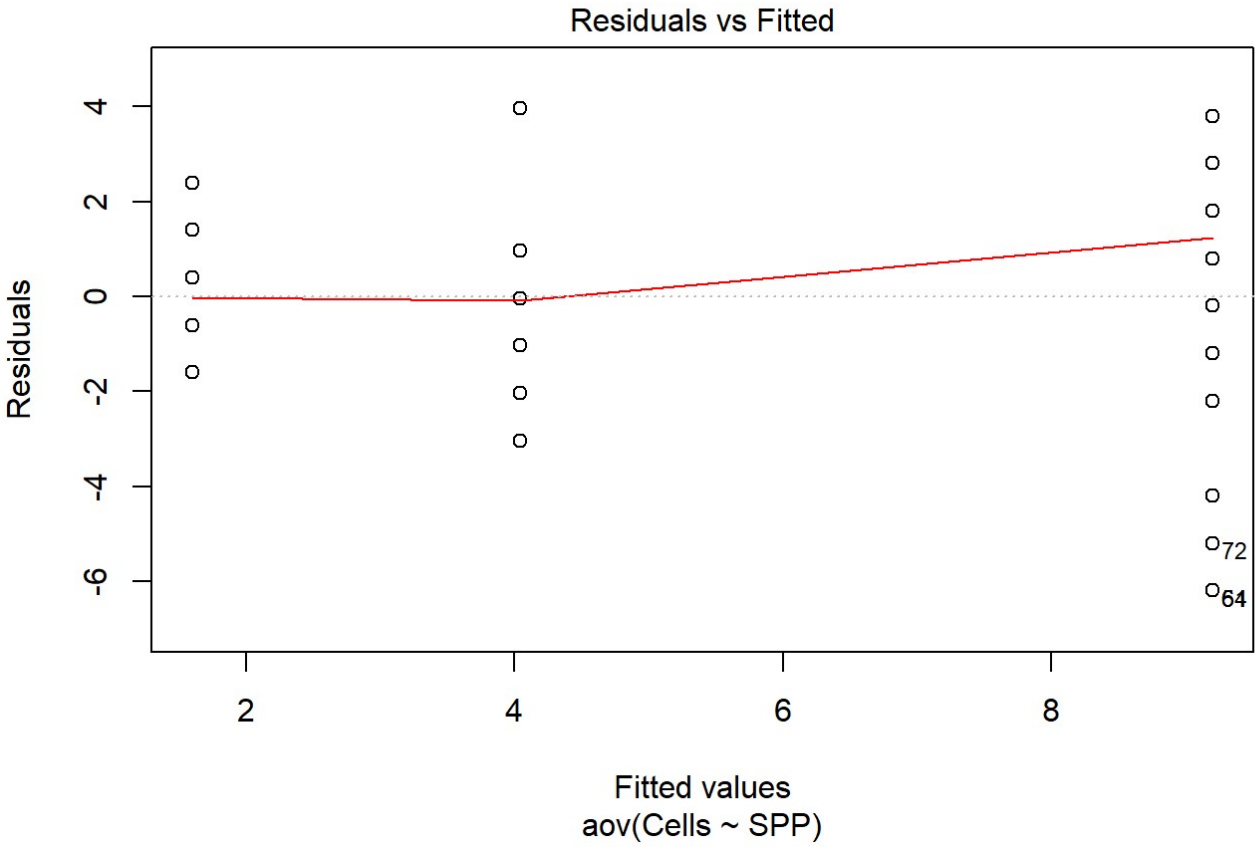
```
bartlett.test(Cells~SPP)
```

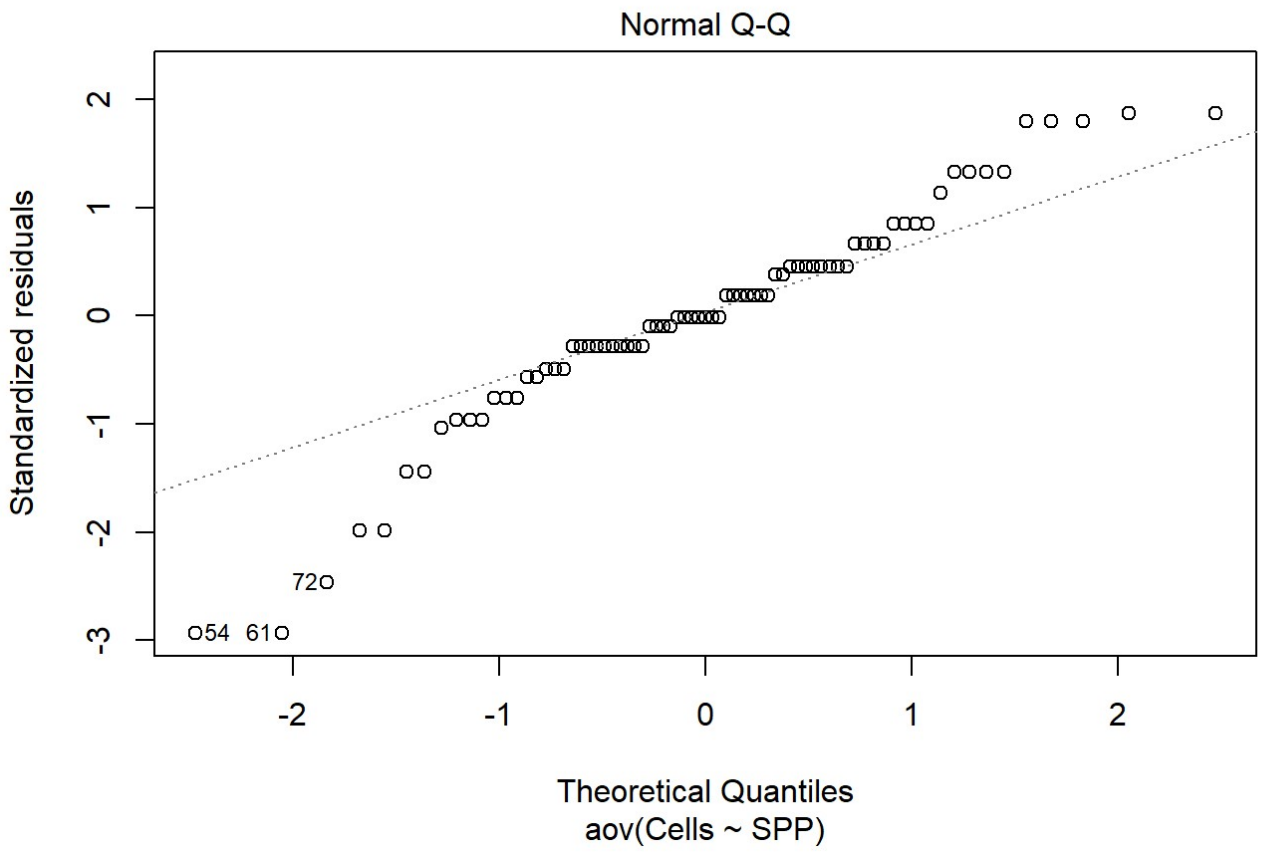
```
##
##  Bartlett test of homogeneity of variances
##
## data:  Cells by SPP
## Bartlett's K-squared = 26.73, df = 2, p-value = 1.569e-06
```

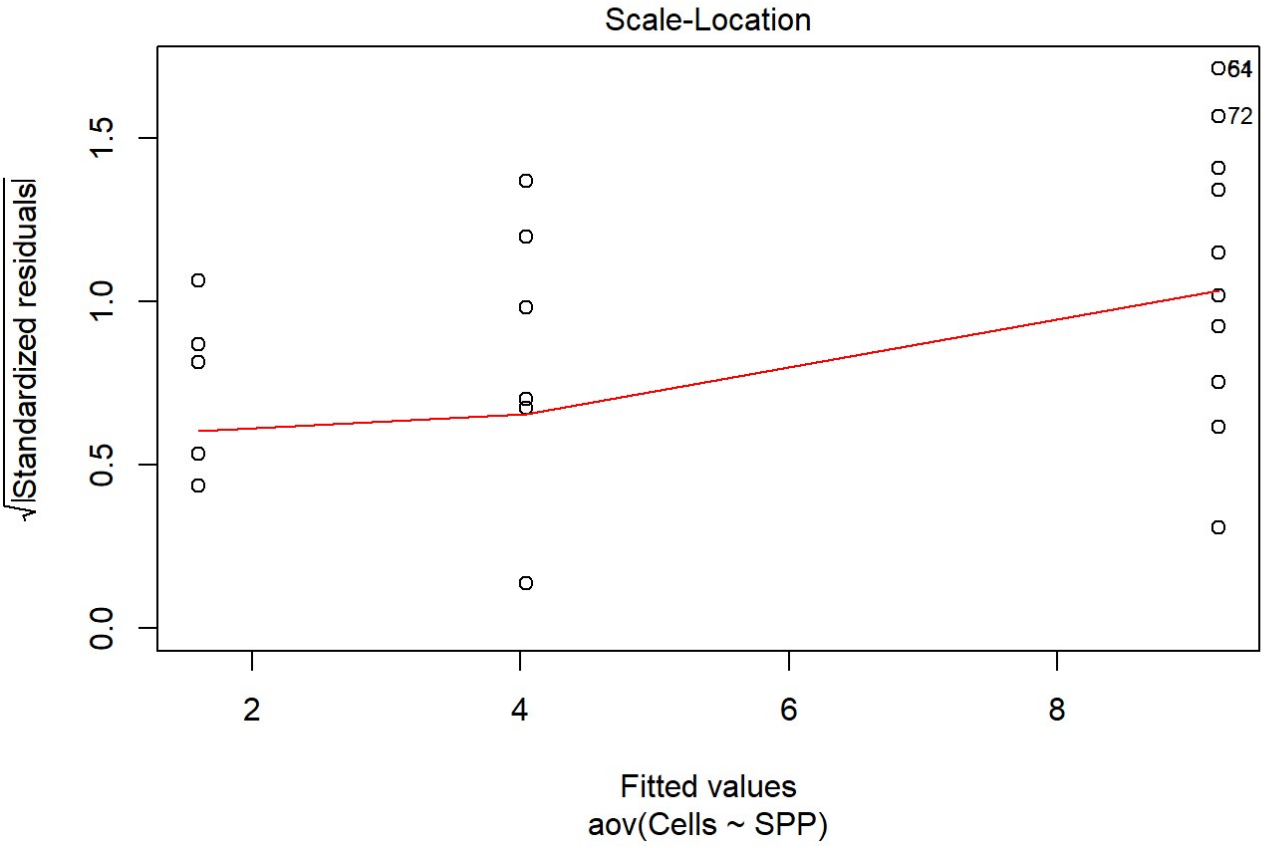
```
shapiro.test(data4.aov$residuals)
```

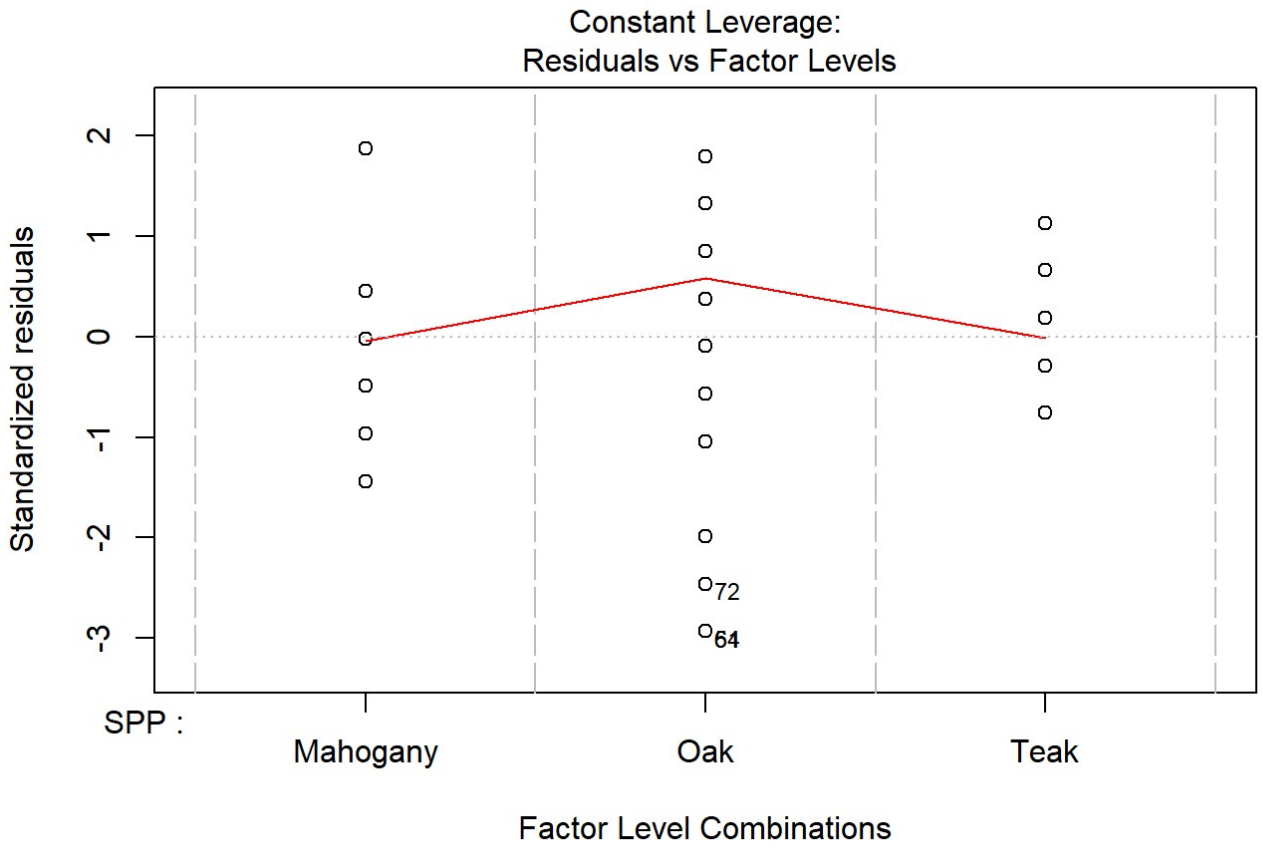
```
##
##  Shapiro-Wilk normality test
##
## data:  data4.aov$residuals
## W = 0.9498, p-value = 0.004864
```

```
plot(data4.aov)
```




Normal Q-Q
Standardized residuals
Theoretical Quantiles
aov(Cells ~ SPP)
72
54
61


Scale-Location
√|Standardized residuals|
Fitted values
aov(Cells ~ SPP)
64
72




```
detach(data4)
```

Conclusion

The Normal Q-Q plot shows a negative skew for the -1 and -2 quantiles and a positive skew for the +1 and +2 quantiles, denoting lower than and higher than expected results respectively. We have an $F(2,72) = 80.91$; $p < 0.00$, showing that we reject the null hypothesis that pore density does not affect hardness and accept the alternate hypothesis.

Scale-location shows that the residuals increase with the fitted data.

Based on these results, Teak is the hardest of the woods, followed by mahogany, and then oak being the softest of the three.