**Written Report – Biochar Study**

**Report Outline:**

* Research Objective
* Model Proposal
* Model Diagnostics
* Data Analysis
* Conclusion
* Programming (code from R)

**Objective:**

Biochar (charcoal) is commonly used as a soil amendment. In this experiment, we are interested to see if there’s a significant statistical effect of the biochar treatments towards plant growth. In addition to the biochar treatment, we also want to test if certain soil typing and wheat varieties also play an effect on plant growth. This results in the experiment containing three main factors: biochar treatment (yes or no), wheat varieties (76 or 1RS), and soil types (soil or soil-sand). We also consider the possibility of an interaction between any of these factors. Each factor is performed with four replicates each and our test is a completely randomized design. In summary, we have three main factors that we want to statistically test on the yield for five different response variables which includes the plant’s shoot weight, shallow root weight, deep root weight, total root weight, and root to shoot ratio. Each response variable will be met with its own statistical model as well as the proper model diagnostics.

**Proposed Model:**

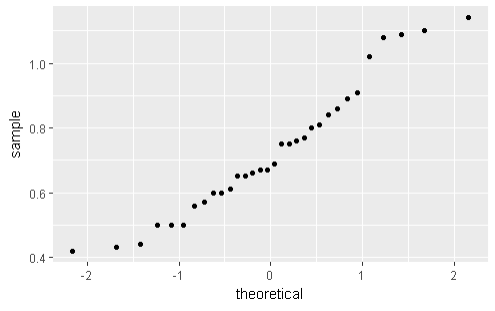
23 Factorial Design  
w/ Assumptions:

* Data follows normal distribution
* Homogeneity of variance

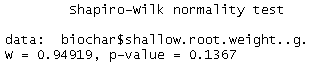
**Model Diagnostics:**

**(Shallow Root Weight)**

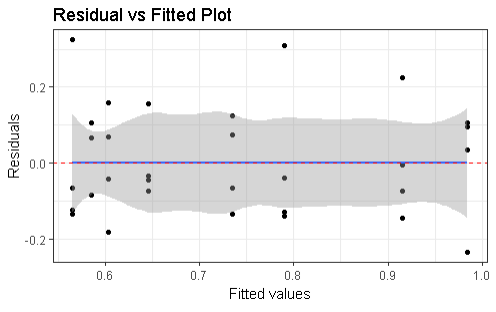
Before we begin the analysis, we check our model assumptions with a normal QQ plot and a Residual vs. Fitted plot.



Based on our output, we can see that the data points regarding Shallow Root Weight falls along the shape of a straight line which indicates a visual fit for normality. We further enhance our findings with a Shapiro-Wilks Test:

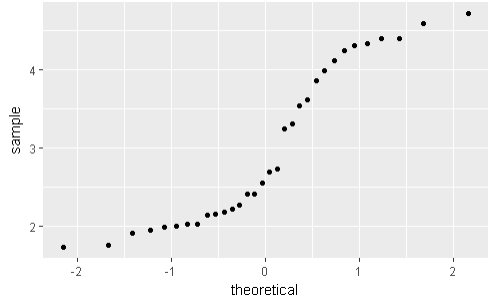


Since our p-value = 0.1367 > α =.05, there is insufficient evidence to indicate our data does not follow normality. In addition, we test for variance homogeneity with Residual vs Fitted plot:

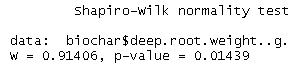


We can see from our output that we do not violate our variance assumption. Hence, a transformation on the data is not necessary.

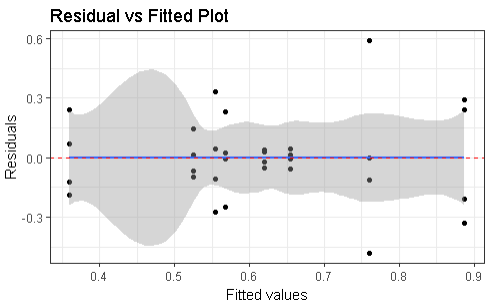
**(Deep Root Weight)**

****

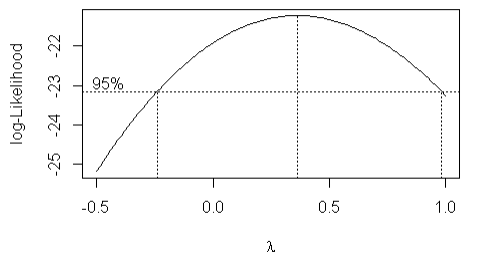
Based on our output, it seems clear that our data does not run along a straight line and there are tail ends of the plot that suggest a pattern. Our visual assumptions of normality appear to have been violated. We follow up our findings with a Shapiro-Wilks test.



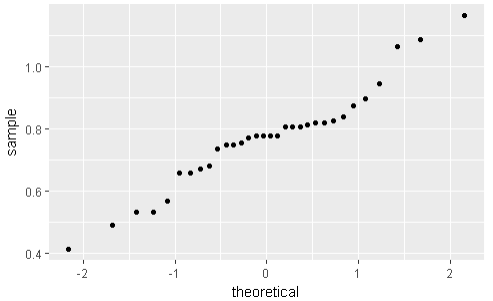
Since our p-value < α = 0.05, we can reject the null hypothesis that our data is normally distributed. We also check our residuals vs fitted plot:

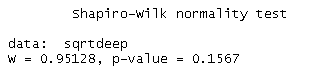


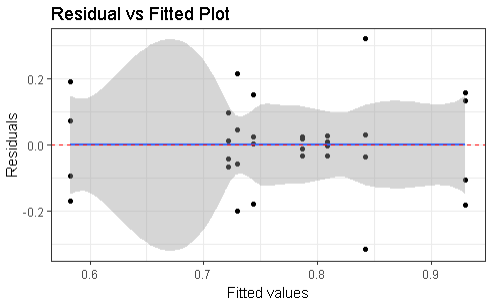
It would appear visually that we have also violated our homogeneity of variance assumption as we can see a fanning out pattern in our plot indicating a variance that is subject to increase. Our plots indicate that we need to transform our data so we visualize a box-cox plot:



According to our output, since our 95% limits for λ captures 0.5, we should proceed with a log transformation on our data for deep root weight. After the necessary transformation, we evaluate our updated plots for model diagnostics:

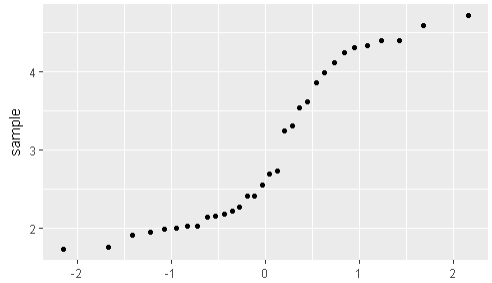




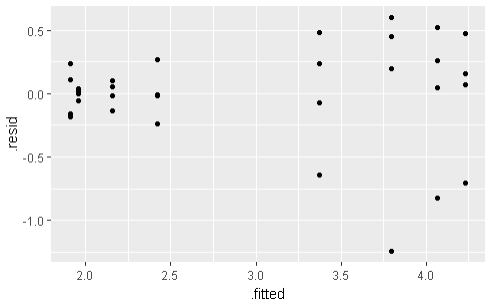


Based on our qq-plot, we can now see that the data points follow a more straight line pattern indicating visually that we have satisfied our normality assumption. In supports of our findings, our updated Shapiro-Wilk p-value also contains a value higher than .05 which complements our claim. Additionally, our updated residual vs fitted plot appears to have improved as we can see less of a fanning pattern and more towards a random scattering of points indicating homogeneity of variance.

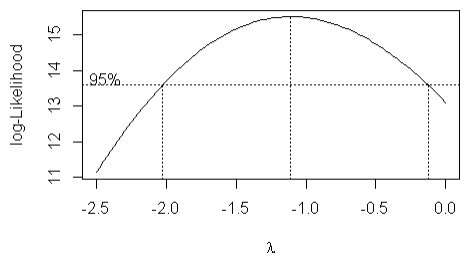
**(Shoot Weight)**



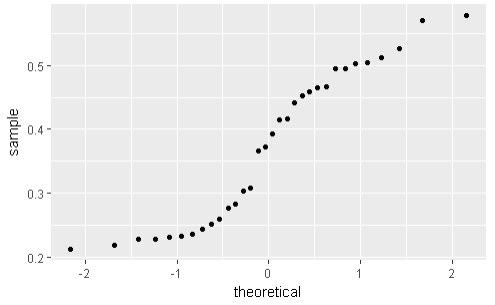
Based on our output, it seems clear that our data does not run along a straight line and there are tail ends of the plot that suggest a pattern. Our visual assumptions of normality appear to have been violated. Our resulting Shapiro-Wilks p-value was also close to 0 supporting our claim.

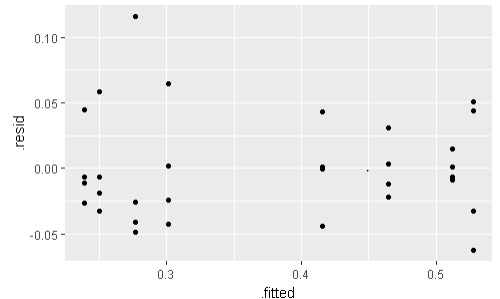


It would appear visually that we have also violated our homogeneity of variance assumption as we can see a fanning out pattern in our plot indicating a variance that is subject to increase. Our plots indicate that we need to transform our data so we visualize a box-cox plot:



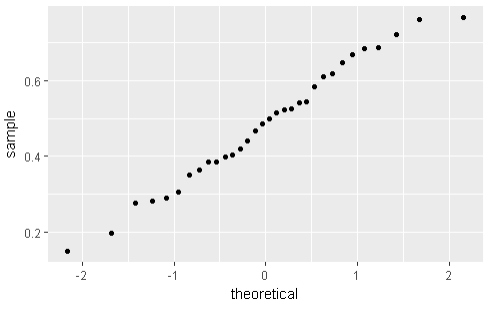
According to our output, since our 95% limits for λ captures -1.0, we should proceed with an inverse transformation on our data for shoot weight. After the necessary transformation, we evaluate our updated plots for model diagnostics:



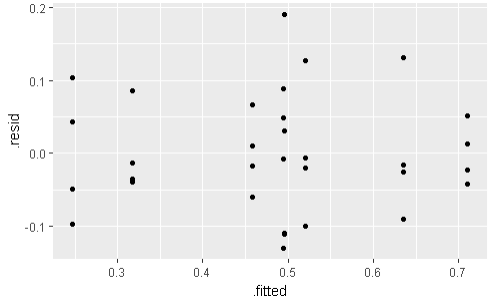


Based on our updated outputs, we appear to have only slightly fixed our qq-plot pattern towards a straight line. This may have been resulted due to the manner of our data set for shoot weight as it appears that we have a large gap in data values which also explains the gap in our plot. Although our normally assumption may not be satisfied with our plot visualization, we may proceed with the analysis because of our sample size for shoot weight being greater than 30 and acknowledging our distance in data point values. It would be safe to assume that if we increased our sample size for shoot weight, our qq-plot would result in our desired behavior. Additionally, it seems quite noticeable disregarding the gap in our data, that we have significantly improved our model assumption for homogeneity of variance.

**(Root to Shoot Ratio)**

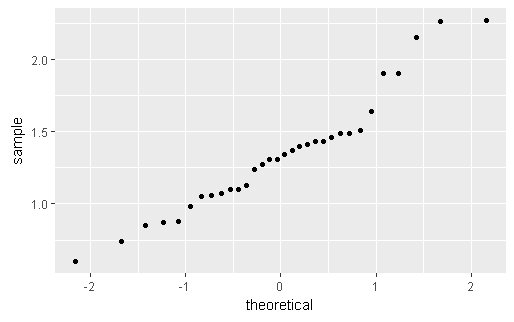


Based on our qq-plot, we can see that our data points follow along the shape of a straight line which is a clear indication that the data points for root-shoot ratio supports the normality assumption. (Shapiro-Wilk p-value = .7781).

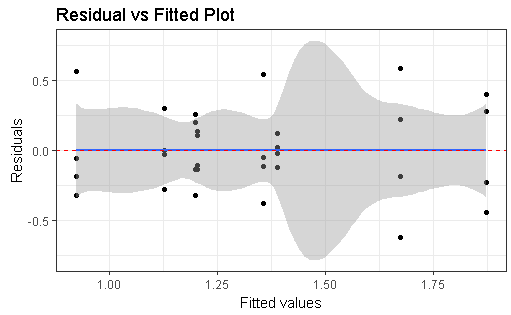


We can also see by the result of our residual vs fitted plot, that we have clear, random scattering of points which supports our model assumption for homogeneity of variances. For root-shoot ratio, there are no necessary requirements for data transformation.

**(Total Root Weight)**



Based on our qq-plot, we can see the data points follow along the pattern of a straight line. Although the last few plotted points seem to tail off (potential outliers), we can still visually assume normality. (Shapiro-Wilk p-value = .1134)



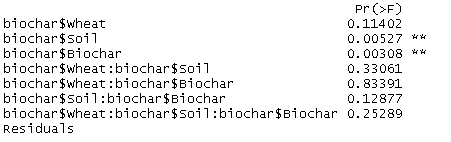
We can see from our residual vs fitted plot that our data points seem to be represented by a random scattering of points supporting our homogeneity of variance assumption. A transformation for total root weight does not appear to be necessary.

**Data Analysis**

For the analysis of variance, we introduce two outputs: ANOVA table for full model and ANOVA table for reduced model. The process from the full model towards the reduced model consisted of using backward elimination, eliminating higher significant terms (with interactions taking priority) and aiming towards a model of best fit for each response variable relative to plant growth. We determine higher significant terms by the larger p-value or smaller F-statistic. After this reduction procedure, we introduce our reduced model with a deterministic model to summarize our important effects for each variable.

**(Shallow Root Weight)**

Full ANOVA :



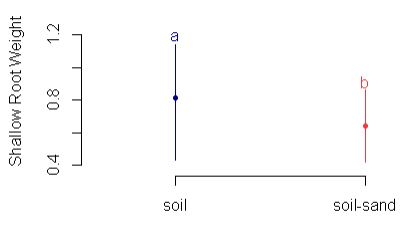
Reduced ANOVA:



Deterministic Model:

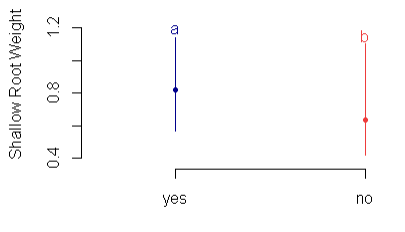


After retrieving our reduced model of best fit for our response variable of interest (in this case shoot weight), we conduct necessary comparisons of means using plots and confidence intervals (adjusted using Tukey’s method). Since we don’t have a significant higher order term for this particular model, we don’t utilize interaction plots.



What we have in this output is our LSD plot to compare means among significant factors which in this case is soil typing. If the plot presents lines of different colors, it indicates that significant difference in means among those compared lines. In our case, the regular soil typing yielded different averages compared to soil-sand type. We further analyze this difference using our Tukey’s adjusted confidence interval: 

This confidence interval compares the difference in averages between soil-sand – soil. Since we have both negative values for our confidence limits, this agrees with our previous conclusion of a difference as well as the knowledge that the regular soil typing yielded higher average shallow root weight than the soil-sand type. We continue similar analysis with our other significant variable: biochar treatment.

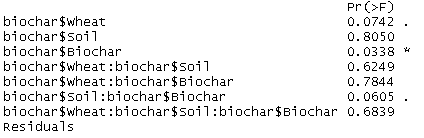




Following similar analysis procedures, we can conclude that the use of biochar treatment yielded significantly higher shallow root weight averages as opposed to our control group.

**(Deep Root Weight)**

Full ANOVA:

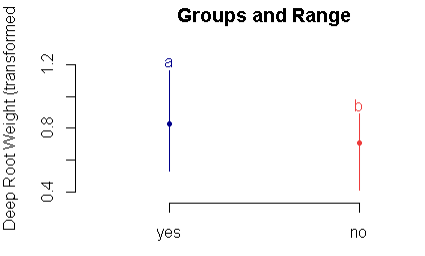


Reduced ANOVA:



Deterministic Model:



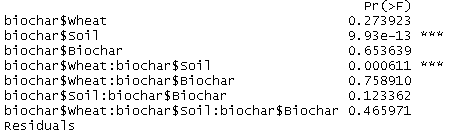




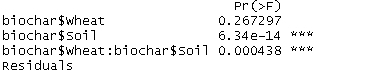
Based on our analysis, we can conclude that the use of biochar treatment yielded significantly higher deep root weight than the control group.

**(Shoot Weight)**

Full ANOVA:



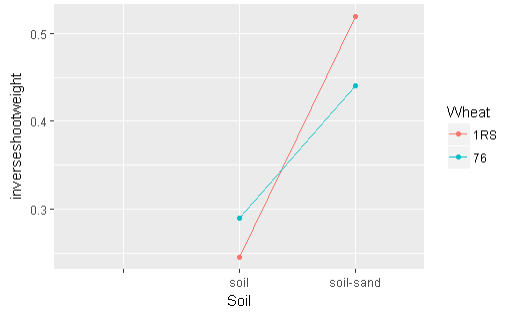
Reduced ANOVA:

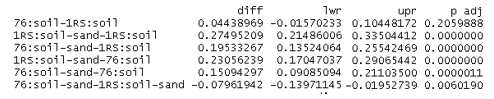


Deterministic Model:



Since for this response variable we have a significant interaction term (wheat and soil typing), we include wheat into our model even though it is not significant as a main effect. We also utilize and interactions plot to visualize our significant differences in means.

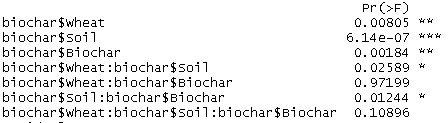




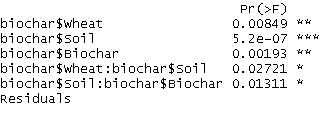
We can see from the intersection within the interaction plot that we have a significance in soil typing with the different wheat types. We can also use our Tukey’s adjusted confidence intervals to check whether there are significant differences in average shoot weight based on all possible combinations of the interaction term for wheat and soil. In this case, we highest yield of shoot weight with 1RS wheat typing with soil-sand.

**(Root-Shoot Ratio)**

Full ANOVA:

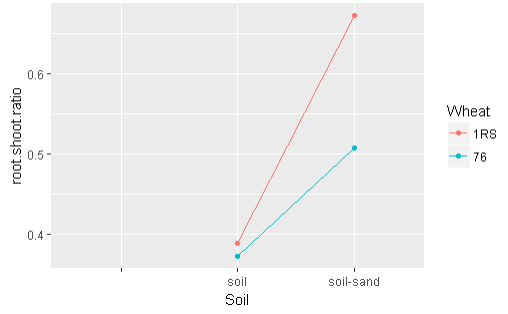


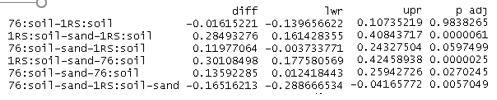
Reduced ANOVA:



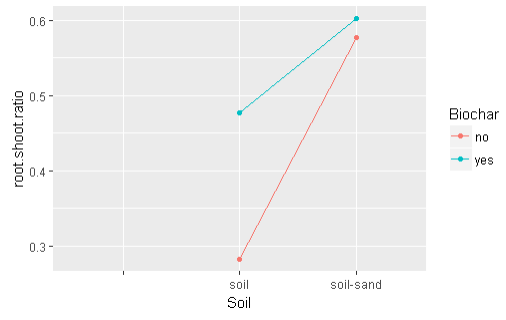
Deterministic Model:

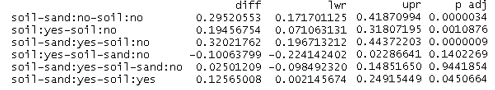






We can see by our output for the interaction between soil typing and wheat, that the combination that yields the highest root-shoot-ratio is soil-sand with 1RS wheat typing. This average was significantly different compared to the alternative wheat typing of 76 with the same soil type.

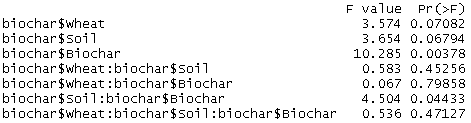




Based on our interaction plot for soil type with biochar treatment, we can see that the highest yield in average root to shoot ratios for this interaction term came from soil-sand. However, we see the points for biochar among soil-sand at a close distance from each other. To see if there’s a difference in those averages we check the Tukey’s adjusted confidence intervals. Since the p-value is .9442 for the difference in average ratios for soil-sand w/ biochar vs. soil-sand without biochar, we can conclude that there isn’t a significant difference from the two. This means that biochar does not play a significant factor in higher average ratios if our soil typing in soil-sand. However, biochar plays a factor when the soil typing is of the regular soil type in which biochar does yield higher average ratios.

**(Total Root Weight)**

Full ANOVA:

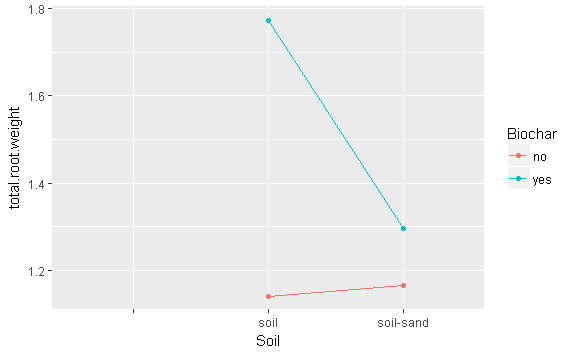


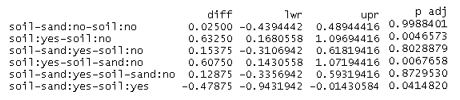
Reduced ANOVA:



Deterministic Model:







Based on our output for the interaction plot on soil typing, we can see visually that we get the highest yield in average root weight through the regular soil type with the treatment of biochar. When focused in on the soil-sand typing, we can tell by our confidence intervals that there is no significant difference in average total root weight yields between the batch with the biochar treatment contrary to our regular soil typing.

**Conclusion**

Shallow Root Weight: With our research goal being to see how we can utilize biochar treatment to maximize plant growth, if we want to focus on shallow root weight, we should consider the biochar treatment in addition to soil typing. Since by our analysis, we only consider the single order terms to be significant, we conclude on individual means and we see that using the regular soil typing or using the biochar treatment will yield increased average shallow root weight.

Deep Root Weight: For deep root weight, we deduced from our analysis that biochar treatment was the only significant effect. With our post-ANOVA analysis we were able to compare the means for deep root weight among the use of biochar treatment and it turns out that the use of biochar treatment did improve yields.

Shoot Weight: For our analysis of the plant’s shoot weight, we deduced a model signifying a significant interaction term which included soil and wheat typing. However, we did not see a significant effect for biochar treatment in shoot weight for this response variable. If we are focusing in on soil-sand, wheat typing 1RS will yield significantly larger shoot weight over type 76. If we focus in on the regular soil typing, there was no difference in average shoot weights for the different wheat types.

Root-Shoot Ratio: When analyzing root-shoot ratio, we deduced two significant interaction terms: wheat with soil typing and soil typing with biochar treatment. For the first interaction, if we want to see higher average ratios, we consider soil-sand typing with 1RS wheat as it is significantly higher than with 76 typing. If we were concentrated on the regular soil type, the use of different wheat types was not statistically significant in higher ratios. For our second interaction, we only saw a significance for biochar treatment within the regular soil typing in which the use of the treatment yielded larger average ratios. However, among soil-sand typing, the use of biochar treatment was not significantly different than the control group.

Total Root Weight: If we evaluate our analysis for the total root weight, we saw a significant interaction term in soil typing with biochar treatment. If we want to see the highest yields in total root weight, we pair the regular soil typing with the use of biochar treatment. Using soil-sand will yield significantly smaller average total root weight yields with the addition of biochar treatment not being significant in larger yields.

**Programming (R)**

library(tidyverse)

library(dplyr)

library(gridExtra)

library(MASS)

library(agricolae)

setwd("C:/Users/Brandonkim939/Desktop/School/STAT 183/Individual Project")

biochar <- read.csv("biochar.csv",header=TRUE)

# Model for shoot weight

shootweight <- aov(biochar$shoot.weight..g.~biochar$Wheat\*biochar$Soil\*biochar$Biochar)

summary(shootweight)

# Shoot Weight = wheat + soil + wheat\*soil

Soil <- biochar$Soil

WheatSoil <- biochar$Wheat:biochar$Soil

ShootWeight <- biochar$shoot.weight..g.

biochar %>% filter(!is.na(root.shoot.ratio))

# Model for shallow root weight

shallowrootweight <- aov(biochar$shallow.root.weight..g.~biochar$Wheat\*biochar$Soil\*biochar$Biochar)

(qq1 <- ggplot(biochar, aes(sample=shallow.root.weight..g.)) + stat\_qq())

p1<-ggplot(shallowrootweight, aes(.fitted, .resid))+geom\_point()

p1<-p1+stat\_smooth(method="loess")+geom\_hline(yintercept=0, col="red", linetype="dashed")

p1<-p1+xlab("Fitted values")+ylab("Residuals")

p1<-p1+ggtitle("Residual vs Fitted Plot")+theme\_bw()

p1

shapiro.test(biochar$shallow.root.weight..g.)

(bx1 <- boxcox(shallowrootweight, data = biochar,

lambda = seq(-2, 2, length = 10))) # no transformation needed ~ 1 is in the interval

(model1 <- summary(shallowrootweight)) # drop 3-way interaction

shallowrootweight <- aov(biochar$shallow.root.weight..g.~biochar$Wheat+biochar$Soil+biochar$Biochar

+biochar$Wheat\*biochar$Soil+biochar$Soil\*biochar$Biochar+biochar$Wheat\*biochar$Biochar)

(model1 <- summary(shallowrootweight)) # drop 2-way interactions

shallowrootweight <- aov(biochar$shallow.root.weight..g.~biochar$Wheat+biochar$Soil+biochar$Biochar)

(model1 <- summary(shallowrootweight))

# Final model = soil + biochar

shallowrootweight <- aov(biochar$shallow.root.weight..g.~biochar$Soil+biochar$Biochar)

(model1 <- summary(shallowrootweight))

# LSD for soil

df<-df.residual(shallowrootweight)

MSerror<-deviance(shallowrootweight)/df

out1 <- with(biochar,LSD.test(shallow.root.weight..g.,Soil,df,MSerror))

plot(out1,ylab="Shallow Root Weight") # LSD plot

(posthocsoil <- TukeyHSD(x=shallowrootweight,'biochar$Soil', conf.level=0.95)) # Use to check if 0 is in interval

# LSD for biochar

out2 <- with(biochar,LSD.test(shallow.root.weight..g.,Biochar,df,MSerror))

plot(out2,ylab="Shallow Root Weight")

(posthocbiochar <- TukeyHSD(x=shallowrootweight,'biochar$Biochar', conf.level=0.95))

# Model for deep root weight

deeprootweight <- aov(biochar$deep.root.weight..g.~biochar$Wheat\*biochar$Soil\*biochar$Biochar)

(qq2 <- ggplot(biochar, aes(sample=shoot.weight..g.)) + stat\_qq())

p2<-ggplot(deeprootweight, aes(.fitted, .resid))+geom\_point()

p2<-p2+stat\_smooth(method="loess")+geom\_hline(yintercept=0, col="red", linetype="dashed")

p2<-p2+xlab("Fitted values")+ylab("Residuals")

p2<-p2+ggtitle("Residual vs Fitted Plot")+theme\_bw()

p2

shapiro.test(biochar$deep.root.weight..g.)

boxcox(deeprootweight, data = biochar,

lambda = seq(-.5, 1, length = 10))

# transformation

sqrtdeep <- sqrt(biochar$deep.root.weight..g.)

model2 <- aov(sqrtdeep ~ biochar$Wheat\*biochar$Soil\*biochar$Biochar)

(qq2 <- ggplot(biochar, aes(sample=sqrtdeep)) + stat\_qq())

p2<-ggplot(model2, aes(.fitted, .resid))+geom\_point()

p2<-p2+stat\_smooth(method="loess")+geom\_hline(yintercept=0, col="red", linetype="dashed")

p2<-p2+xlab("Fitted values")+ylab("Residuals")

p2<-p2+ggtitle("Residual vs Fitted Plot")+theme\_bw()

p2

shapiro.test(sqrtdeep)

summary(model2)

model2 <- aov(sqrtdeep ~ biochar$Biochar)

summary(model2)

df<-df.residual(model2)

MSerror<-deviance(model2)/df

out1 <- with(biochar,LSD.test(sqrtdeep,Biochar,df,MSerror))

plot(out1,ylab="Deep Root Weight (transformed") # LSD plot

(posthocsoil <- TukeyHSD(x=model2,'biochar$Biochar', conf.level=0.95))

# Model for total root weight

totalrootweight <- aov(biochar$total.root.weight~biochar$Wheat\*biochar$Soil\*biochar$Biochar)

summary(totalrootweight)

(qq3 <- ggplot(biochar, aes(sample=total.root.weight)) + stat\_qq())

p3<-ggplot(totalrootweight, aes(.fitted, .resid))+geom\_point()

p3<-p3+stat\_smooth(method="loess")+geom\_hline(yintercept=0, col="red", linetype="dashed")

p3<-p3+xlab("Fitted values")+ylab("Residuals")

p3<-p3+ggtitle("Residual vs Fitted Plot")+theme\_bw()

p3

shapiro.test(biochar$total.root.weight)

boxcox(totalrootweight, data = biochar,

lambda = seq(-2, 2, length = 10))

totalrootweight <- aov(biochar$total.root.weight~biochar$Soil\*biochar$Biochar)

summary(totalrootweight)

# model = soil + biochar + soil\*biochar

df<-df.residual(totalrootweight)

MSerror<-deviance(totalrootweight)/df

out1 <- with(biochar,LSD.test(total.root.weight,Soil,df,MSerror))

(posthocsoil <- TukeyHSD(x=totalrootweight,'biochar$Soil:biochar$Biochar', conf.level=0.95)) # 0 in interval

(biocharsoil <- ggplot(aes(x=Soil,color=Biochar,group=Biochar,y=total.root.weight),data=biochar) +

stat\_summary(fun.y=mean,geom="point") +

stat\_summary(fun.y=mean,geom="line"))

summary(rootshootratio <- aov(biochar$root.shoot.ratio~biochar$Wheat\*biochar$Soil\*biochar$Biochar-biochar$Wheat:biochar$Soil:biochar$Biochar-biochar$Wheat:biochar$Biochar))

(qq4 <- ggplot(biochar, aes(sample=biochar$root.shoot.ratio)) + stat\_qq())

p4<-ggplot(rootshootratio, aes(.fitted, .resid))+geom\_point()

p3<-p3+stat\_smooth(method="loess")+geom\_hline(yintercept=0, col="red", linetype="dashed")

p3<-p3+xlab("Fitted values")+ylab("Residuals")

p3<-p3+ggtitle("Residual vs Fitted Plot")+theme\_bw()

p4

shapiro.test(biochar$root.shoot.ratio)

boxcox(rootshootratio, data = biochar,

lambda = seq(-2, 2, length = 10))

summary(rootshootratio)

rootshootratio <- aov(biochar$root.shoot.ratio ~ biochar$Wheat+biochar$Soil+biochar$Biochar

+biochar$Wheat\*biochar$Soil+biochar$Soil\*biochar$Biochar)

summary(rootshootratio)

out1 <- with(biochar,LSD.test(root.shoot.ratio,Wheat:Soil,df,MSerror))

(posthocint <- TukeyHSD(x=rootshootratio,'biochar$Wheat:biochar$Soil', conf.level=0.95))

plot(out1,ylab="Root-Shoot Ratio")

out2 <- with(biochar,LSD.test(root.shoot.ratio,Biochar,df,MSerror))

(posthocint <- TukeyHSD(x=rootshootratio,'biochar$Soil:biochar$Biochar', conf.level=0.95))

plot(out2,ylab="Root-Shoot Ratio")

wheatsoil <- ggplot(data=biochar) +

aes(x = Soil, color = Wheat, group = Wheat, y = root.shoot.ratio) +

stat\_summary(fun.y = mean, geom = "point") +

stat\_summary(fun.y = mean, geom = "line")

wheatsoil

soilbiochar <- ggplot(data=biochar) +

aes(x = Soil, color = Biochar, group = Biochar, y = root.shoot.ratio) +

stat\_summary(fun.y = mean, geom = "point") +

stat\_summary(fun.y = mean, geom = "line")

soilbiochar

# Model for shoot weight

shootweight <- aov(biochar$shoot.weight..g.~biochar$Wheat\*biochar$Soil\*biochar$Biochar)

(qq5 <- ggplot(biochar, aes(sample=biochar$shoot.weight..g.)) + stat\_qq())

p5<-ggplot(shootweight, aes(.fitted, .resid))+geom\_point()

p3<-p3+stat\_smooth(method="loess")+geom\_hline(yintercept=0, col="red", linetype="dashed")

p3<-p3+xlab("Fitted values")+ylab("Residuals")

p3<-p3+ggtitle("Residual vs Fitted Plot")+theme\_bw()

p5

shapiro.test(biochar$shoot.weight..g.)

boxcox(shootweight, data = biochar,

lambda = seq(-2.5, 0, length = 10))

# Transformed model

inverseshootweight <- 1 / (biochar$shoot.weight..g.)

model5 <- aov(inverseshootweight~biochar$Wheat\*biochar$Soil\*biochar$Biochar)

(qq5 <- ggplot(biochar, aes(sample=inverseshootweight)) + stat\_qq() )

p5<-ggplot(model5, aes(.fitted, .resid))+geom\_point()

p3<-p3+stat\_smooth(method="loess")+geom\_hline(yintercept=0, col="red", linetype="dashed")

p3<-p3+xlab("Fitted values")+ylab("Residuals")

p3<-p3+ggtitle("Residual vs Fitted Plot")+theme\_bw()

p5

shapiro.test(inverseshootweight)

summary(model5)

newmodel5 <- aov(inverseshootweight~biochar$Wheat+biochar$Soil + biochar$Wheat\*biochar$Soil)

summary(newmodel5)

wheatsoil <- ggplot(aes(x=Soil,color=Wheat,group=Wheat,y=inverseshootweight),data=biochar) +

stat\_summary(fun.y=mean,geom="point") +

stat\_summary(fun.y=mean,geom="line")

wheatsoil

(posthocint <- TukeyHSD(x=newmodel5,'biochar$Wheat:biochar$Soil', conf.level=0.95))