

ENV 710: Lab 8

Jiahuan Li

Spring 2023

Problem 1

Research Question and Hypothesis

Research questions for problem 1 are 1) what variables (including mean tree diameter, mean height, mean wood density, mean basal area, and presence of tree falls in the plot) influence plot biomass (AGBH.Mg.ha) in the given dataset? And 2) what is the direction and magnitude of their effect on biomass? The null hypothesis is that none of the variables have a significant impact on plot biomass, while the alternative hypothesis is that at least one variable significantly influences biomass.

Method

Normality assessment

The QQ plot indicates that the dependent variable biomass satisfies the normality assumption, rendering log transformation unnecessary. However, a log-transformed sequence of biomass was also tested and found to be less satisfactory than the original sequence.

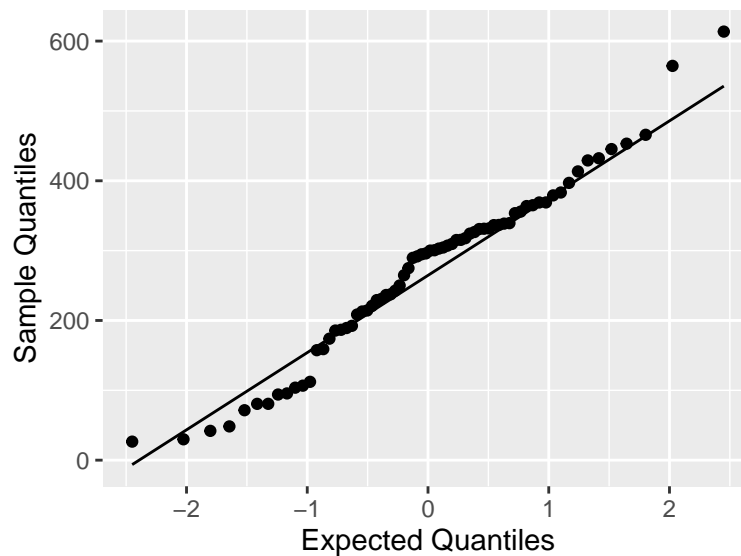


Figure 1: Normal QQ Plot for Biomass

Initial model

Therefore, the initial model encompassing all potential influencing variables is formulated as follows:

```
##
## Call:
## lm(formula = AGBH.Mg.ha ~ mDBH.cm + mH.m + mWD.g.m3 + mBA.cm2 +
##     factor(Tree.Fall), data = na.omit(Tree))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -278.30  -42.78   13.42   40.80  235.97
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -423.8723    173.2292  -2.447 0.017258 *
## mDBH.cm         -72.9733     70.1023  -1.041 0.301939
## mH.m             91.9557     81.5644   1.127 0.263918
## mWD.g.m3        624.5868    169.1739   3.692 0.000472 ***
## mBA.cm2          0.7226      0.4976   1.452 0.151472
## factor(Tree.Fall)major  40.9792     43.1042   0.951 0.345448
## factor(Tree.Fall)minor  -2.9520     38.0787  -0.078 0.938456
## factor(Tree.Fall)none   -5.2892     40.1616  -0.132 0.895649
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 87.29 on 62 degrees of freedom
## Multiple R-squared:  0.5598, Adjusted R-squared:  0.5101
## F-statistic: 11.26 on 7 and 62 DF,  p-value: 4.133e-09
```

The regression summary reveals that only the `intercept` and `mWD.g.m3` terms are statistically significant, suggesting that the model is likely not the minimum adequate model for the problem.

Model reduction

To determine the minimum adequate model, a backward stepwise regression approach is employed. Starting with the full model, the `step()` function is used to perform step-wise model selection.

```
## Start:  AIC=633.19
## AGBH.Mg.ha ~ mDBH.cm + mH.m + mWD.g.m3 + mBA.cm2 + factor(Tree.Fall)
##
##              Df Sum of Sq    RSS    AIC
## - factor(Tree.Fall)  3    19054 491441 629.96
## - mDBH.cm           1     8256 480643 632.41
## - mH.m              1     9684 482071 632.61
## <none>                                472387 633.19
## - mBA.cm2           1     16069 488456 633.54
## - mWD.g.m3          1    103854 576241 645.11
##
## Step:  AIC=629.96
## AGBH.Mg.ha ~ mDBH.cm + mH.m + mWD.g.m3 + mBA.cm2
##
##              Df Sum of Sq    RSS    AIC
## - mDBH.cm       1     9142 500583 629.25
## - mH.m           1     9729 501170 629.33
## <none>                                491441 629.96
## - mBA.cm2       1    21037 512478 630.90
## - mWD.g.m3      1    98626 590067 640.76
##
```

```
## Step: AIC=629.25
## AGBH.Mg.ha ~ mH.m + mWD.g.m3 + mBA.cm2
##
##           Df Sum of Sq    RSS    AIC
## - mH.m      1       729 501312 627.35
## <none>                        500583 629.25
## - mBA.cm2    1     62208 562791 635.45
## - mWD.g.m3   1     94699 595282 639.38
##
## Step: AIC=627.35
## AGBH.Mg.ha ~ mWD.g.m3 + mBA.cm2
##
##           Df Sum of Sq    RSS    AIC
## <none>                        501312 627.35
## - mWD.g.m3   1    110961 612274 639.35
## - mBA.cm2    1    242602 743915 652.98
##
## Call:
## lm(formula = AGBH.Mg.ha ~ mWD.g.m3 + mBA.cm2, data = na.omit(Tree))
##
## Coefficients:
## (Intercept)      mWD.g.m3      mBA.cm2
##   -286.6753      587.5988       0.3127
```

I use the Akaike Information Criterion (AIC) for model comparison. The `step(lm1)` reports the AIC number for each step (i.e., each nested model). For each step, the table displays the AIC numbers associated with the removal of specific predictors indicated by their respective row names.

For instance, the starting AIC number is 633.19. In this case, the AIC numbers for `factor(Tree.Fall)`, `mDBH.cm`, and `mH.m` are 629.96, 632.41, and 632.61, respectively, which are smaller than the model AIC of 633.19. Thus, removing these variables would result in a reduction of the model AIC number, indicating an improvement in the model. Then, examining the final step in which these three predictors were eliminated, we find an AIC number of 627.35. In this case, removing any of the remaining variables would cause an increase in AIC, indicating that the model has achieved the minimum adequate model.

Minimum adequate model

```
lm1.best <- lm(formula = AGBH.Mg.ha ~ mWD.g.m3 + mBA.cm2, data = na.omit(Tree))
summary(lm1.best)
```

```
##
## Call:
## lm(formula = AGBH.Mg.ha ~ mWD.g.m3 + mBA.cm2, data = na.omit(Tree))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -304.74  -30.83   11.75   46.51  233.10
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -286.67527    86.14930  -3.328  0.001425 **
## mWD.g.m3      587.59879   152.58513   3.851  0.000266 ***
## mBA.cm2        0.31271    0.05492   5.694 2.99e-07 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 86.5 on 67 degrees of freedom
## Multiple R-squared:  0.5328, Adjusted R-squared:  0.5189
## F-statistic: 38.21 on 2 and 67 DF,  p-value: 8.451e-12
```

Compare to the initial model, the adjusted R-squared increases from 0.5101 to 0.5189, and the residual standard error falls from 87.29 to 86.5. Thus, we can conclude that this model is better than the initial one with only significant independent variables remained.

Results

The regression model is significant ($R^2 = 0.5189$, $F_{2,67} = 38.21$, $p = 8.451e-12$) with p value less than 0.05. The adjusted R-squared value is 0.5189, which means that the model with two independent variables explains about 51.89% of the variance in mean biomass while adjusting for the degrees of freedom and the number of predictors in the model.

Both independent variables, mean wood density ($p = .000266$) and mean basal area ($p = 2.99e-07$), are significant predictors of mean biomass. The estimated coefficient for mean wood density is 587.59879, which means that for every one-unit increase in the mean wood density (g/m^3), the mean biomass is expected to increase by 587.59879 units, holding all other variables constant. Similarly, the estimated coefficient for mean basal area (cm^2) is 0.31271, which means that for every one-unit increase in the mean basal area, the mean biomass is expected to increase by 0.31271 units, holding all other variables constant.

Assumptions check

The diagnosis plots were utilized to assess the assumptions of the statistical test.

- The residual plot indicates that the residuals are dispersed randomly around the centerline, although the line does not consistently remain horizontal at 0.
- The QQ plot displayed a linear trend for the majority of residuals, indicating generally good normality of the residuals.
- In the figure of the standardized residual plot, the variance seems to increase with the treatment means, which can be attributed to the presence of observation 47 as an outlier.
- The leverage plot identifies outliers 8 and 47, which exceed the Cook's distance limit of 0.5.

Overall, the diagnostic plots are not satisfactory, indicating that the model assumptions may not be fully met. To explore possible solutions, I tried transforming the dependent variable using a logarithmic function, but this did not lead to significant improvements in the diagnosis plots. Additionally, I attempted to remove the outliers from the original dataset. While this did improve the adjusted R-squared value and diagnosis plots, it did not substantially change the estimated parameters of the predictors and the conclusions. As I cannot determine whether these observations are true outliers or simply extreme values, I decided to keep them in the dataset.

```
mean(residuals(lm1.best))
```

```
## [1] 2.043207e-15
```

```
vif(lm1.best)
```

```
## mWD.g.m3  mBA.cm2
```

```
## 1.184032 1.184032
```

Additionally, the calculated mean error is very small and close to 0 as expected, indicating that the model has a good fit to the data. Furthermore, the variance inflation factor (VIF) is around 1.18 in this case, indicating that multicollinearity is not an issue. That is, the independent variables are not highly correlated with each other.

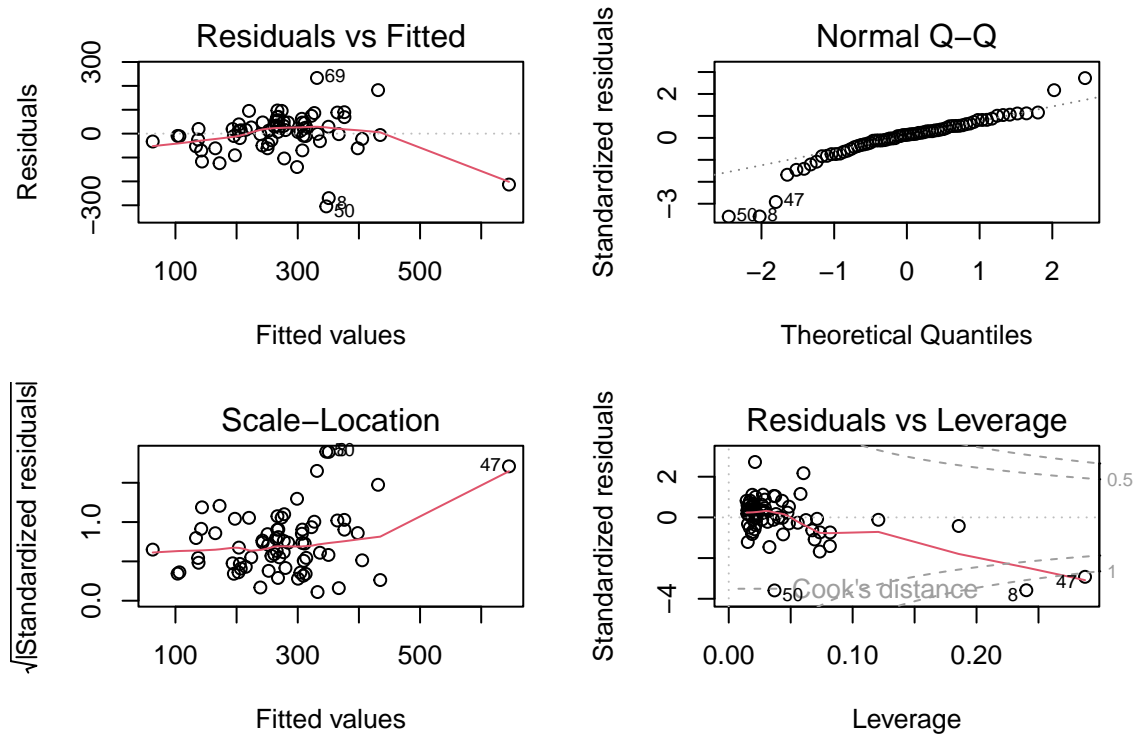


Figure 2: Diagnosis plots

Visualization

Figure 3 and 4 displays the impact of wood density and basal area on biomass, respectively, with each plot holding the other variable constant at its mean value.

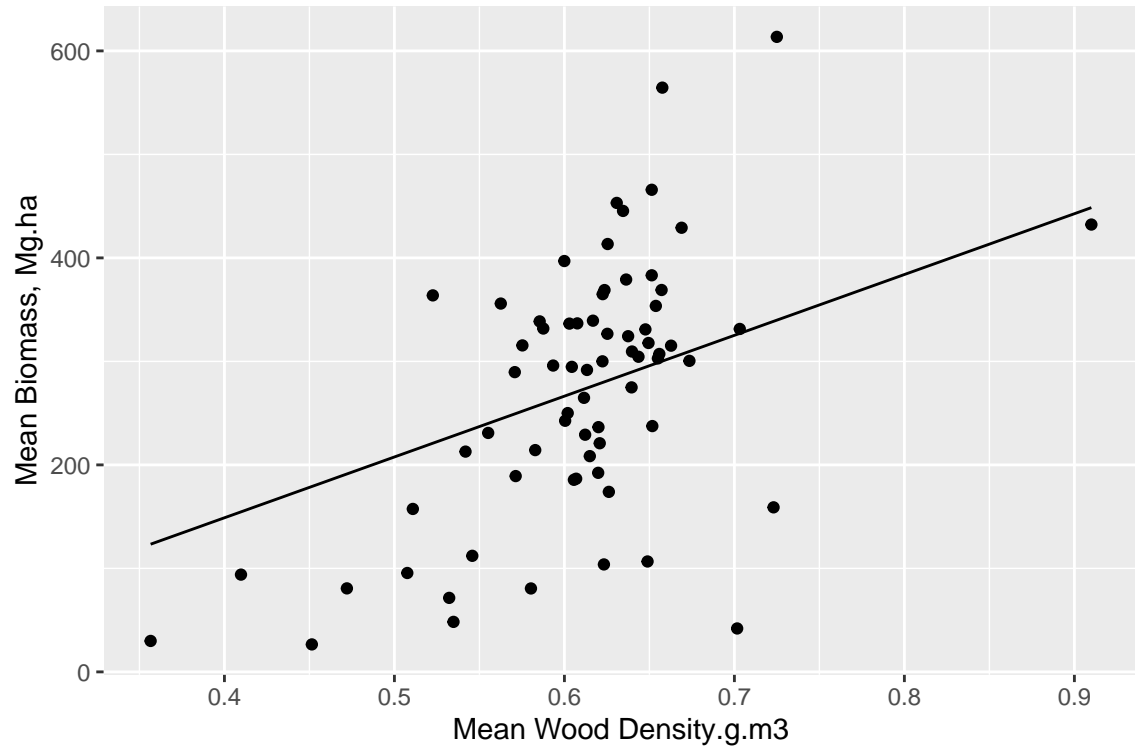


Figure 3: Wood Density Effects on Mean Biomass

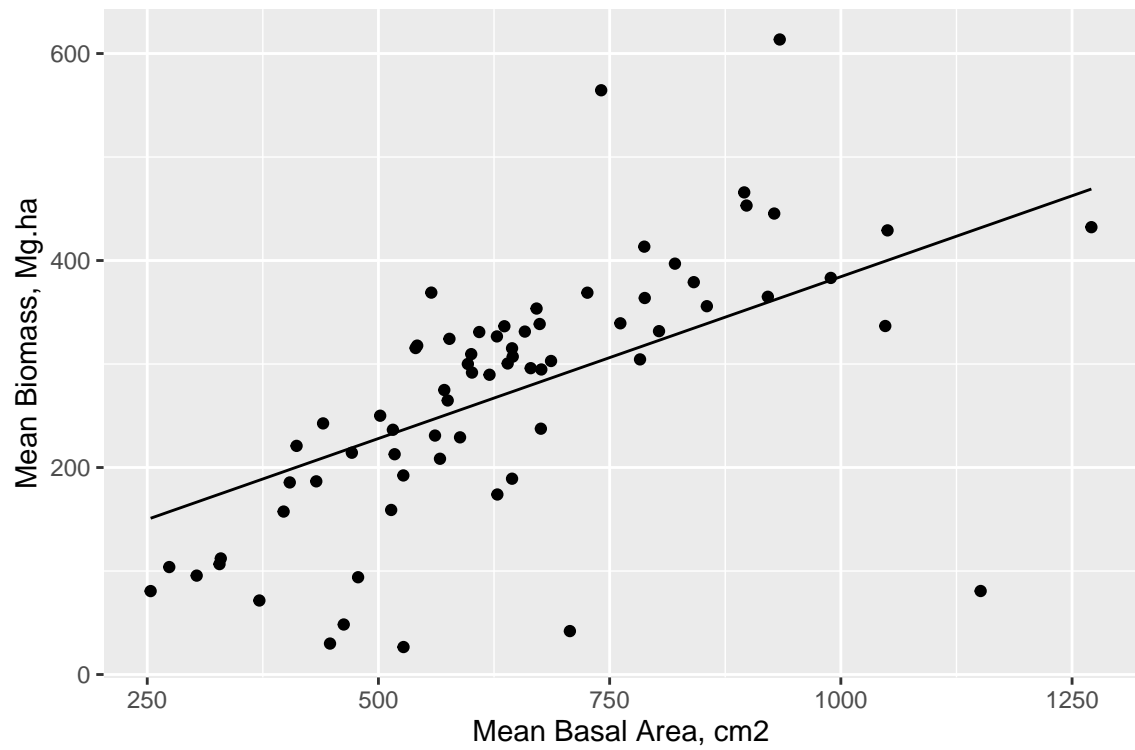


Figure 4: Basal Area Effects on Mean Biomass

Problem 2

Hypothesis

Research questions for problem 2 are 1) what variables (including wind, radiation, temperature and their interactions) influence the ozone concentration? And 2) what is the direction and magnitude of their effects? Null hypothesis for this problem is that there is no significant relationship between any of the weather variables (temperature, solar radiation, and wind speed) or their interaction terms and the ozone concentration. While the alternative hypothesis is that there is a significant relationship between some of the weather variables (and their interactions) and ozone concentration.

Method

Normality Assessment

The QQ plot indicates that the dependent variable does not satisfy the normality assumption. In this case, the log transformation helps the QQ plot look better.

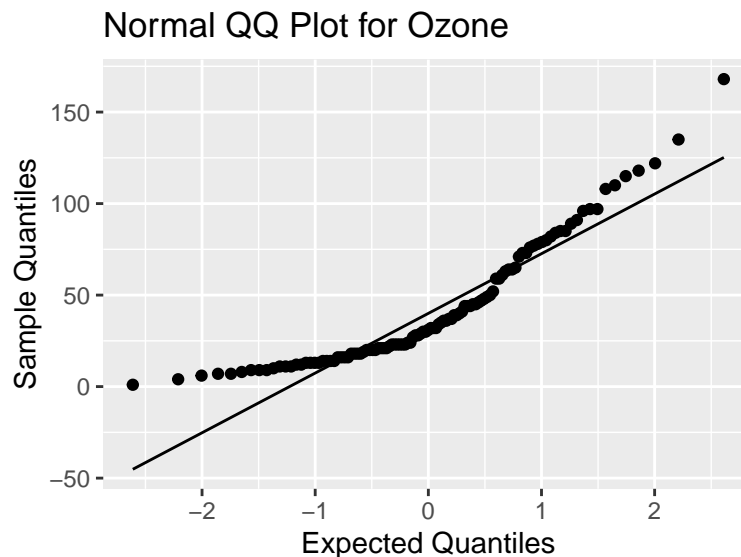


Figure 5: Normal QQ Plots for Ozone and Log(Ozone)

```
##
## Shapiro-Wilk normality test
##
## data:  ozone$ozone
## W = 0.87355, p-value = 2.846e-08

##
## Shapiro-Wilk normality test
##
## data:  log(ozone$ozone)
## W = 0.9712, p-value = 0.01669
```

Furthermore, the results obtained from the `shapiro.test()` suggest that the log-transformed dependent variables exhibit a greater degree of normality than the original distribution, as evidenced by the higher W statistic and p-value.

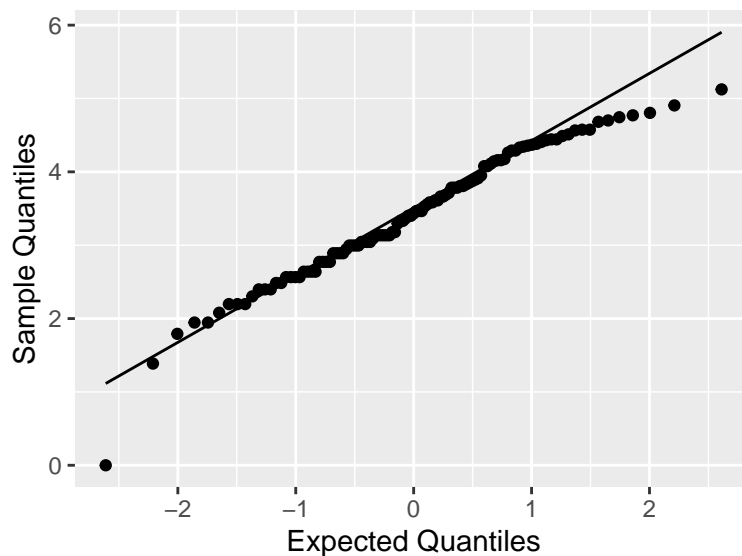


Figure 6: Normal QQ Plots for Ozone and Log(Ozone)

Initial model

Therefore, the initial model encompassing all potential influencing variables and the log-transformed dependent variable is formulated as follows:

```
lm2 <- lm(log(ozone) ~ rad + temp + wind + rad:temp + rad:wind + temp:wind + rad:temp:wind, data = ozone)
summary(lm2)
```

```
##
## Call:
## lm(formula = log(ozone) ~ rad + temp + wind + rad:temp + rad:wind +
##     temp:wind + rad:temp:wind, data = ozone)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.88015 -0.31393 -0.06175  0.26079  1.21729
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -5.760e+00  2.814e+00  -2.047  0.04320 *
## rad           2.474e-02  1.532e-02   1.615  0.10934
## temp          1.205e-01  3.847e-02   3.133  0.00226 **
## wind          4.557e-01  2.317e-01   1.967  0.05190 .
## rad:temp      -2.844e-04  2.031e-04  -1.401  0.16433
## rad:wind      -2.117e-03  1.351e-03  -1.567  0.12019
## temp:wind     -6.844e-03  3.273e-03  -2.091  0.03897 *
## rad:temp:wind  2.755e-05  1.826e-05   1.509  0.13445
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5006 on 103 degrees of freedom
## Multiple R-squared:  0.687, Adjusted R-squared:  0.6657
## F-statistic: 32.3 on 7 and 103 DF, p-value: < 2.2e-16
```


The interactions among the predictors are considered in this case. The regression summary reveals that some of the predictors are not statistically significant, suggesting that the model is not the minimum adequate model for the problem.

Model reduction

```
##          df          AIC
## lm2      9 171.0912
## lm2.1    8 171.5173
## lm2.2    7 169.5278
## lm2.3    6 167.8976

## Analysis of Variance Table
##
## Model 1: log(ozone) ~ rad + temp + wind + rad:temp + rad:wind + temp:wind +
##          rad:temp:wind
## Model 2: log(ozone) ~ rad + temp + wind + rad:temp + rad:wind + temp:wind
## Model 3: log(ozone) ~ rad + temp + wind + rad:wind + temp:wind
## Model 4: log(ozone) ~ rad + temp + wind + temp:wind
## Model 5: log(ozone) ~ rad + temp + wind
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1     103 25.812
## 2     104 26.383 -1  -0.57038 2.2760 0.13445
## 3     105 26.385 -1  -0.00251 0.0100 0.92044
## 4     106 26.473 -1  -0.08804 0.3513 0.55468
## 5     107 27.667 -1  -1.19416 4.7651 0.03131 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I began by removing the insignificant 3-order interaction terms to create model `lm2.1`. However, further examination of the summary regression revealed that there were still insignificant terms in the model. Therefore, I removed two additional interaction terms with a p-value greater than 0.05. This process led to a decrease in the AIC number, as evidenced in the AIC table, which suggests that the new model is a better fit than the previous one (except that from `lm2` to `lm2.1`).

`lm2.3` is considered as the final model because it consists only of the significant terms along with the `wind` term. Even though itself is not significant, it was instrumental in constructing a significant interaction term. Therefore, it is not advisable to eliminate the `wind` term from the regression model.

The results of the anova test corroborate this finding, as the p-values for all the reduced models, except for `lm2.4`, are not significant. Since the null hypothesis (H_0) in this case is that the reduced model is adequate, a significant p-value indicates that we should reject the reduced model (i.e., the one in which a variable was removed). Therefore, we can conclude that the adjustment made from `lm2.3` to `lm2.4` is not appropriate, and the final model should be `lm2.3`.

Minimum adequate model

```
lm2.best <- lm(log(ozone) ~ rad + temp + wind + temp:wind, data = na.omit(ozone))
summary(lm2.best)
```

```
##
## Call:
## lm(formula = log(ozone) ~ rad + temp + wind + temp:wind, data = na.omit(ozone))
##
## Residuals:
##          Min           1Q       Median           3Q          Max
```

```
## -1.98696 -0.32076 -0.05428 0.30238 1.19016
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.5845520  1.1936395  -2.165  0.0326 *
## rad          0.0025939  0.0005483   4.731 6.93e-06 ***
## temp         0.0782341  0.0145784   5.366 4.76e-07 ***
## wind         0.1661571  0.1052917   1.578  0.1175
## temp:wind    -0.0029299  0.0013399  -2.187  0.0310 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4997 on 106 degrees of freedom
## Multiple R-squared:  0.679, Adjusted R-squared:  0.6669
## F-statistic: 56.05 on 4 and 106 DF, p-value: < 2.2e-16
```

Compared to the initial model, the adjusted R-squared has slightly increased and the residual standard error has slightly decreased in the new model. Therefore, we can conclude that the new model is an improvement over the initial one.

Results

The F-statistic and the very small p-value suggest that the model as a whole is statistically significant ($R^2 = 0.6669$, $F_{4,106} = 56.05$, $p < 2.2e-16$). The adjusted R-squared value of 0.6669 indicates that approximately 66.69% of the variation in the log of ozone can be explained by the predictor variables in the model. The residual standard error of 0.4997 represents the average difference between the observed values of the log of ozone and the predicted values by the model. The coefficient estimates of the linear regression model show that radiation (effect = 0.0026, $t = 4.731$, $p < 0.05$), temperature (effect = 0.078, $t = 5.366$, $p < 0.05$), and the interaction of temperature and wind (effect = -0.0029, $t = -2.187$, $p < 0.05$) have statistically significant impacts on ozone concentration, whereas wind ($p = 0.118 > 0.05$) does not appear to have a significant effect.

```
##      var      estimate
## [1,] "(Intercept)" "-92.4570137947742"
## [2,] "rad"         "0.259726191882348"
## [3,] "temp"        "8.1375748084872"
## [4,] "wind"        "18.0758529915469"
## [5,] "temp:wind"   "-0.292559725242947"
```

As radiation increases by one unit, the ozone concentration increases by about 0.26%, holding other variables constant. Similarly, a 1-unit increase in temperature leads to a 8.14% increase in ozone concentration. The estimate of the interaction term indicates that one unit change in temperature will cause the impact of changing wind on ozone to decrease by 0.29%, and vice versa.

Assumption checks

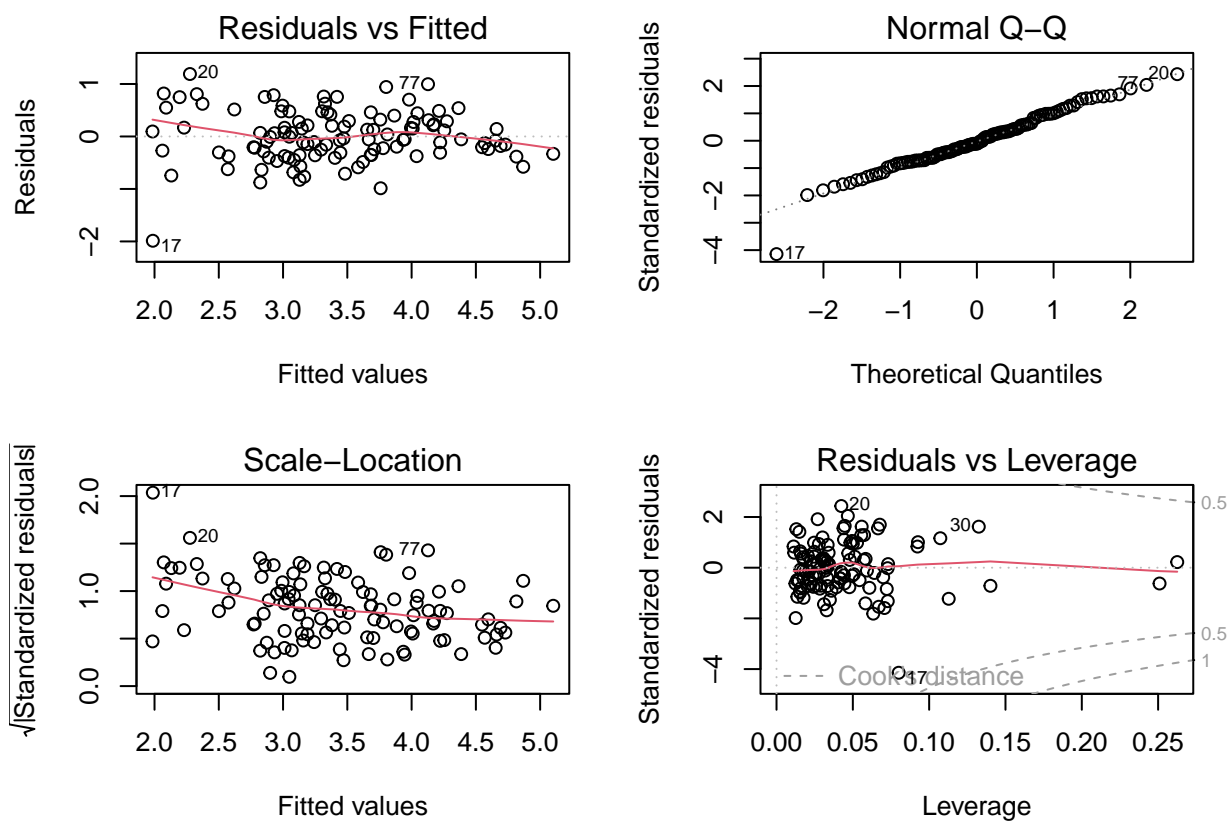


Figure 7: Diagnosis plots

To evaluate how well the model fits the data, I used diagnosis plots, and based on the criteria mentioned above, the plots generally looked good. However, a few outliers were observed, specifically observations 17, 20, and 77. Although it is unclear whether they are true outliers or simply extreme values, removing them did not significantly affect the results. Therefore, I decided to keep these observations in the dataset.

```
mean(residuals(lm2.best))

## [1] 1.04678e-18

shapiro.test(residuals(lm2.best))

##
##  Shapiro-Wilk normality test
##
## data:  residuals(lm2.best)
## W = 0.97703, p-value = 0.05235

vif(lm2.best, type = 'predictor')
```

```
##          GVIF Df GVIF^(1/(2*Df)) Interacts With Other Predictors
## rad  1.10004  1      1.048828      --          temp, wind
## temp 1.10004  3      1.016018      wind          rad
## wind 1.10004  3      1.016018      temp          rad
```

In addition to assessing the assumptions graphically, I performed some numerical tests to evaluate the model's quality. Firstly, I checked the mean error of residuals, which was negligible and close to zero, suggesting that the model fits the data well. Secondly, I examined the normality of residuals using the Shapiro-Wilk test, and the p-value was greater than 0.05, indicating that the normality assumption is acceptable and cannot be rejected.

To assess multicollinearity, I used the `vif()` function. However, since the model includes higher-order terms (interactions), there was a warning message that the VIFs may not be calculated correctly by default. Therefore, I added the argument `"type = 'predictor'"` to ensure that the VIFs were correctly calculated. After that, the VIFs were all around 1.1, indicating that multicollinearity was not a problem in this case.

Visualization

The figures below illustrate the effects of independent variables on the log-transformed dependent variable. The figure above shows the relationship between the independent variable `rad` and the dependent variable, which is the predictor that is included only as a main effect. On the other hand, the graph below demonstrates the relationship between the interacting independent variables `temperature` and `wind` with the dependent variable. These three lines visualize the relationship between temperature and ozone while holding wind constant at its mean, 25th percentile, and 75th percentile values. The non-parallel lines indicate the presence of an interaction effect between the two variables. To add more details, we can customize the colors and labels for the three lines created by the `"stat_function"` function.

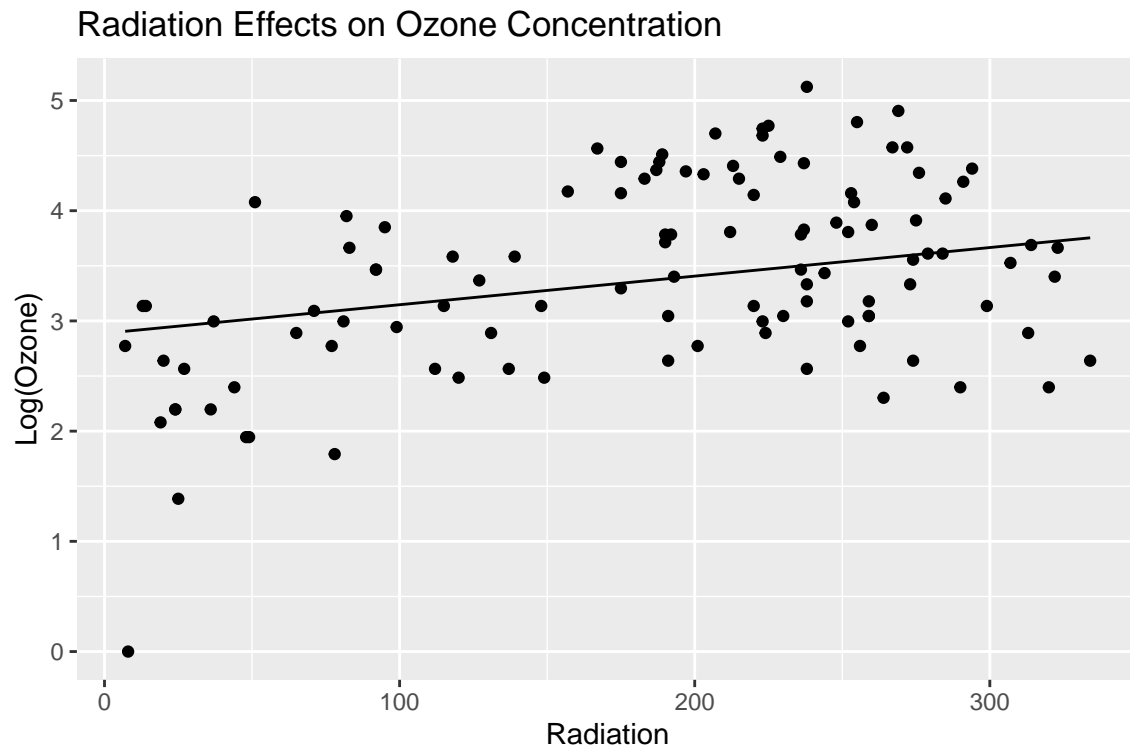


Figure 8: Visualization of independent variables' effects on the dependent variable

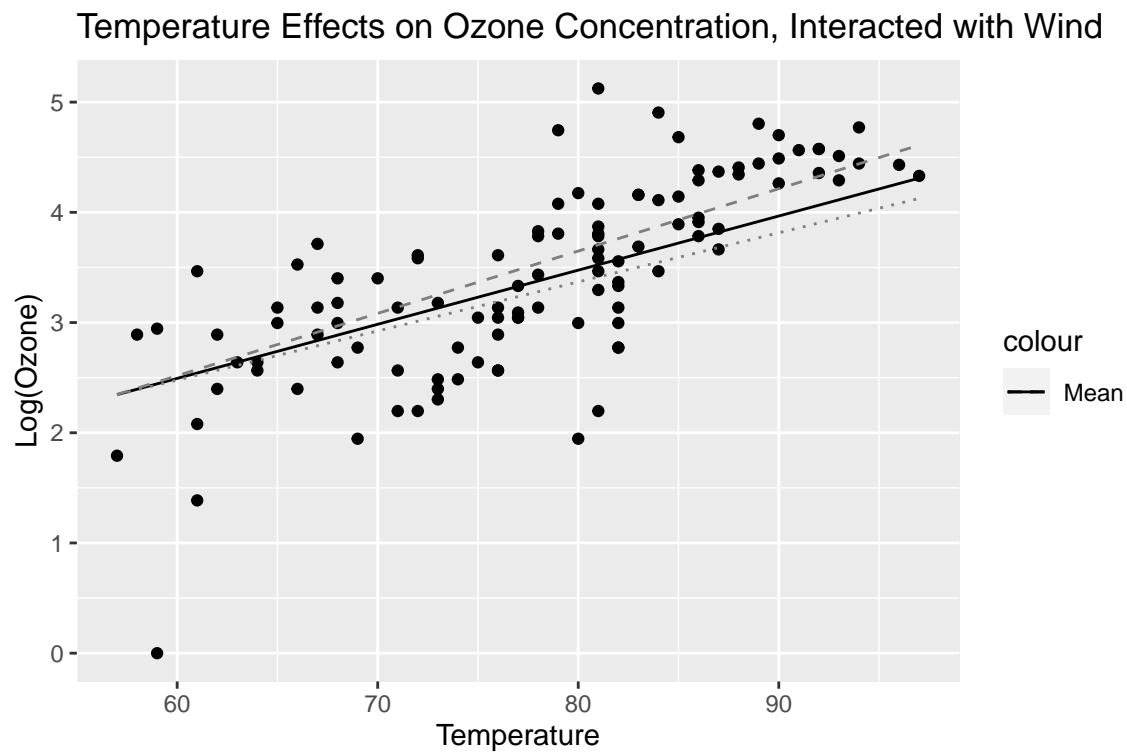


Figure 9: Visualization of independent variables' effects on the dependent variable

```

knitr::opts_chunk$set(echo = FALSE, eval = TRUE, warning = F, message = FALSE)
pacman::p_load(ggplot2, car)
Tree <- read.csv("labs/lab8 linear models with multiple predictors/Tree_Plots_Lab8.csv")

ggplot(data.frame(x = Tree$AGBH.Mg.ha), aes(sample = x)) +
  stat_qq() +
  stat_qq_line() +
  labs(x = "Expected Quantiles", y = "Sample Quantiles")
lm1 <- lm(AGBH.Mg.ha ~ mDBH.cm + mH.m + mWD.g.m3 + mBA.cm2 + factor(Tree.Fall),
  data = na.omit(Tree))
summary(lm1)
step(lm1)
lm1.best <- lm(formula = AGBH.Mg.ha ~ mWD.g.m3 + mBA.cm2, data = na.omit(Tree))
summary(lm1.best)
par(mfrow = c(2,2), mar = c(5,4,2,2))
plot(lm1.best)

#tree1 <- Tree[-c(8,47),]
#lm1.best1 <- lm(formula = AGBH.Mg.ha ~ mWD.g.m3 + mBA.cm2, data = na.omit(tree1))
#summary(lm1.best1)
#plot(lm1.best1)
mean(residuals(lm1.best))
vif(lm1.best)
eq1.wd <- function(x)
{ coef(lm1.best)[1] + coef(lm1.best)[2] * x + coef(lm1.best)[3] * mean(Tree$mBA.cm2)}

eq1.ba <- function(x)
{ coef(lm1.best)[1] + coef(lm1.best)[2] * mean(Tree$mWD.g.m3) + coef(lm1.best)[3] * x}

par(mfrow = c(2,1))
ggplot(Tree) +
  geom_point(aes(x = mWD.g.m3, y = AGBH.Mg.ha)) +
  stat_function(aes(x = mWD.g.m3, fun = eq1.wd, geom = "line") +
  labs(x = "Mean Wood Density.g.m3", y = "Mean Biomass, Mg.ha")
ggplot(Tree) +
  geom_point(aes(x = mBA.cm2, y = AGBH.Mg.ha)) +
  stat_function(aes(x = mBA.cm2, fun = eq1.ba, geom = "line") +
  labs(x = "Mean Basal Area, cm2", y = "Mean Biomass, Mg.ha")
ozone <- read.csv("labs/lab8 linear models with multiple predictors/ozone.data.csv")

par(mfrow=c(1,2))
ggplot(data.frame(x = ozone$ozone), aes(sample = x)) +
  stat_qq() +
  stat_qq_line() +
  labs(title = "Normal QQ Plot for Ozone", x = "Expected Quantiles", y = "Sample Quantiles")

ggplot(data.frame(x = log(ozone$ozone)), aes(sample = x)) +
  stat_qq() +
  stat_qq_line() +
  labs(x = "Expected Quantiles", y = "Sample Quantiles")
shapiro.test(ozone$ozone)
shapiro.test(log(ozone$ozone))
lm2 <- lm(log(ozone) ~ rad + temp + wind + rad:temp + rad:wind + temp:wind + rad:temp:wind, data = ozone)

```

```

summary(lm2)
lm2.1 <- update(lm2, ~. - rad:temp:wind)

lm2.2 <- update(lm2.1, ~.- rad:temp)

lm2.3 <- update(lm2.2, ~.- rad:wind)

AIC(lm2,lm2.1,lm2.2,lm2.3)

lm2.4 <- update(lm2.3, ~.- temp:wind)

anova(lm2,lm2.1,lm2.2,lm2.3,lm2.4)
lm2.best <- lm(log(ozone) ~ rad + temp + wind + temp:wind, data = na.omit(ozone))
summary(lm2.best)
estimate <- c()
var <- c()
for (i in c(1:5)){
  estimate[i] <- (exp(coef(lm2.best)[i]) - 1) * 100
  var[i] <- names(coef(lm2.best)[i])
}

cbind(var,estimate)
par(mfrow = c(2,2), mar = c(5,4,2,2))
plot(lm2.best)
mean(residuals(lm2.best))
shapiro.test(residuals(lm2.best))
vif(lm2.best, type = 'predictor')

eq2.rad <- function(x)
{ coef(lm2.best)[1] + coef(lm2.best)[2] * x + coef(lm2.best)[3] * mean(ozone$temp) + coef(lm2.best)[4] *

eq2.temp_mean <- function(x)
{ coef(lm2.best)[1] + coef(lm2.best)[2] * mean(ozone$rad) + coef(lm2.best)[3] * x + coef(lm2.best)[4] *

eq2.temp_25 <- function(x)
{ coef(lm2.best)[1] + coef(lm2.best)[2] * mean(ozone$rad) + coef(lm2.best)[3] * x + coef(lm2.best)[4] *

eq2.temp_75 <- function(x)
{ coef(lm2.best)[1] + coef(lm2.best)[2] * mean(ozone$rad) + coef(lm2.best)[3] * x + coef(lm2.best)[4] *

par(mfrow = c(1,2))

ggplot(ozone) +
  geom_point(aes(x = rad, y = log(ozone))) +
  stat_function(aes(x = rad, fun = eq2.rad, geom = "line") +
  labs(title = "Radiation Effects on Ozone Concentration",
        x = "Radiation", y = "Log(Ozone)")

ggplot(ozone) +
  geom_point(aes(x = temp, y = log(ozone))) +
  stat_function(aes(x = temp, color = "Mean"), fun = eq2.temp_mean, geom = "line", linetype = "solid")

```

```

stat_function(aes(x = temp, color = "25th Quantile"), fun = eq2.temp_25, geom = "line", linetype = "dashed",
stat_function(aes(x = temp, color = "75th Quantile"), fun = eq2.temp_75, geom = "line", linetype = "dotted",
labs(title = "Temperature Effects on Ozone Concentration, Interacted with Wind",
      x = "Temperature", y = "Log(Ozone)") +
scale_color_manual(values = c("Mean" = "black", "25% Quantile" = "blue", "75% Quantile" = "red")) +
scale_linetype_manual(values = c("Mean" = "solid", "25% Quantile" = "dashed", "75% Quantile" = "dotted"))

```