Kelly Wang

112037466

AMS 315 Project 2

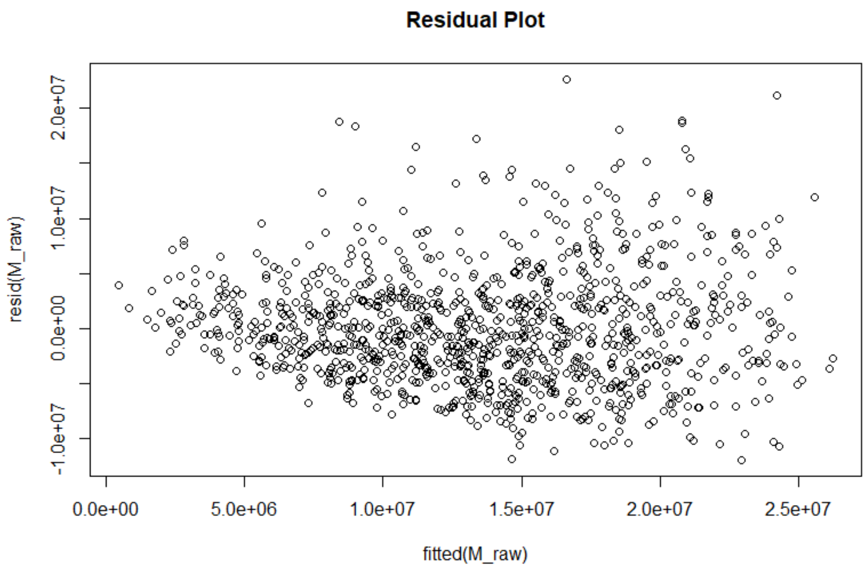
Introduction

The data provided to me includes an interaction between E, environmental variables, and G, genetic variables. The goal of my study is to find the estimated model function that was used to generate my given data. This is provided in the context of finding GxE interactions as studied by Caspi et al. and others in the previous study. In my data, I have 1081 observations with 4 environmental variables and 20 genetic variables.

Methods

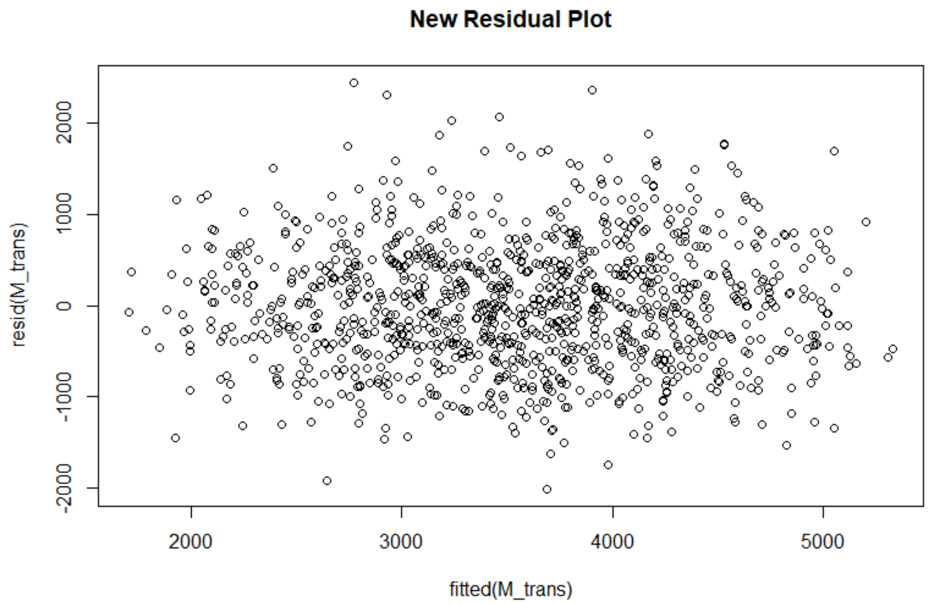
First, I opened my file using read.table() in order to read the .csv file that was assigned to me. I then used a lm function for only the E_i values, attributed to the name, “M_E”. This was to control for the environmental variables. I used the adjusted r-squared function to get me my r-squared value which was 0.4962557. After controlling for the environmental variables, I assessed the contribution of genetic variables by attributing the function to the name “M_raw”. After taking both environmental and genetic variables into consideration, I learned that the adjusted-r square value is now 0.503905. I noted that there is only about a 0.004 increase in the r-squared value, which is not impressive by its scalar degree.

After that, I made sure to graph my residual plot and received the following plot.



From the plot, I noticed that it was not close to be a flat ellipse. To resolve this, I ran the boxcox function to create an optimal transformation that would help me to attain a flat ellipse. From there, I noticed that my estimated lambda value appeared to be around 0.5. Thus, I altered the function to $Y^{0.5}$. After that, I printed the summary of the r-squared value with all the data again. The r-squared value of the transformed data was 0.5385358. Because my r-squared value increased, this shows that the transformation created a more optimal graph.

Afterwards, I graphed the new residual plot and was able to receive the graph below. As noted, the new residual plot was able to come closer to attaining a flat-ellipse shape.



Afterwards, I installed the packages for “leaps”, “knitr”, as well as “kableExtra”. I retrieved these packages from my library and then printed the following tables that would provide me insight on creating an optimal model for estimating my given variable values.

In my initial code, I included the data to the second power $(.)^2$ to include for the interaction levels. The initial model chosen was “(Intercept)+E2:E3+E3:G17”. However, later on, I was able to deduce that there was no significant interaction between E3 and G17 as shown by the significant coefficients. Furthermore, later on in my M_2stage table where I found the coefficients for the expected model, there was no mention of G17 as a significant value. This proved to be a contradiction, especially because I could not ignore the fact that G17 produced an observantly large t value. I can conclude through using Occim’s Razor and law of parsimony that I should put my preferential selection over the simplest model that is able to correctly convey and include correct information. From there, I changed the beginning part of my code with the model summary to raise it to the power of 1 and emphasize significance within the scope of the individual variable. Thus, from here on, I decided to change my code so that my data was raised to the first power $(.)^1$.

In my first table, I labelled it Model Summary and I chose the 3rd model, (Intercept)+E2+E3+G17. I chose the 3rd model because there is a small increase in the adjR2 value from the 3rd to the 4th model. Furthermore, another measure that could be used to assess the model is checking the BIC value changes. In the third column of the table, one can deduce that the BIC change from the 3rd to 4th model is significantly much smaller than the decreases from the other models and that the 3rd model has the smallest BIC value.

Model Summary

model	adjR2	BIC
(Intercept)+E3	0.310041050142424	-387.844988462386
(Intercept)+E2+E3	0.529894771830186	-796.232179373285
(Intercept)+E2+E3+G17	0.539809220740668	-813.27136621407
(Intercept)+E2+E3+G3+G17	0.541049665407465	-810.205905611854
(Intercept)+E2+E3+G3+G13+G17	0.541495423885762	-805.275774736494

Thus, I chose the variables associated in the third model, namely E2, E3, and G17. Afterwards, I ran my code under the title “M_main” to look for the significant coefficients. In order to make sure that my model stands valid, I needed to make sure that the main effects are significant.

Sig Coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
E2	236.8734	10.59716	22.352532	0.0e+00
E3	266.6243	10.33195	25.805792	0.0e+00
G17	248.2914	53.31335	4.657208	3.6e-06

As shown by the table, all the variables within the third model were indeed significant coefficients due to their large t values.

Afterwards, I decided to run code under the title, “M_2nd” to take into account for the second interactions. From my results, I was able to learn that there was a significant interaction between E3 and G2 as well as G12 and G20. Because there was no mention of an interaction of the variables that I was interested in for my proposed model, I was able to move past this to the next step.

In terms of generating the final coefficients, I ran my code under the name, “M_2stage”. This allowed me to print the table and receive the following values as shown.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-942.8576	130.10305	-7.247006	0e+00
E2	236.4378	10.48283	22.554760	0e+00
E3	268.9584	10.18812	26.399204	0e+00
G17	259.0149	52.64888	4.919667	1e-06

The intercept value is -942.8576, the coefficient of E2 = 236.4378, the coefficient of E3 = 268.9584, and the coefficient of G17 = 259.0149. From here, I was able to attain the values for my final model.

Result

Initially, based on my summary model table, I chose the third proposed model which was:

$$Y = (Intercept) + E2 + E3 + G17$$

I determined that each of the variables had significant coefficients. Furthermore, from the M_2stage table, I was able to determine that the intercept value is -942.8576, the coefficient of E2 = 236.4378, the coefficient of E3 = 268.9584, and the coefficient of G17 = 259.0149.

The final model based on my printed tables is:

$$Y = -942.8576 + 236.4378E2 + 268.9584E3 + 259.0149G17$$

As shown by the table, because all the p-values of the variables were close to 0, I can reject the null hypothesis.

Conclusions and Discussion

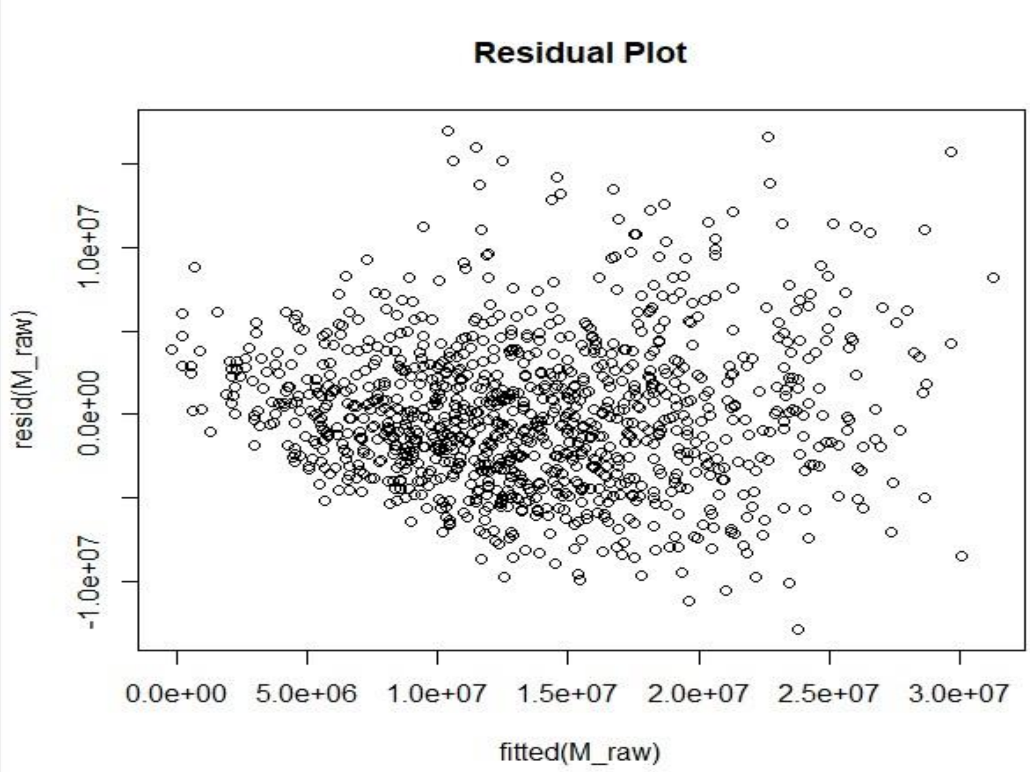
To conclude, I was able to choose the third model from my model summary and produce $Y = -942.8576 + 236.4378E2 + 268.9584E3 + 259.0149G17$ as the estimated model for the data. Because the p value is very close to 0 and 0 is not included in our confidence interval, we can conclude that we reject the null hypothesis. I can also conclude a small change in the r squared value from only the environmental variables, where r^2= 0.4962557 and the data including both environmental and genetic variables, where r^2= 0. 503905. However, we can also conclude that after running the box-cox transformation, my r-squared value was able to achieve an observable increase to r^2= 0. 5385358.

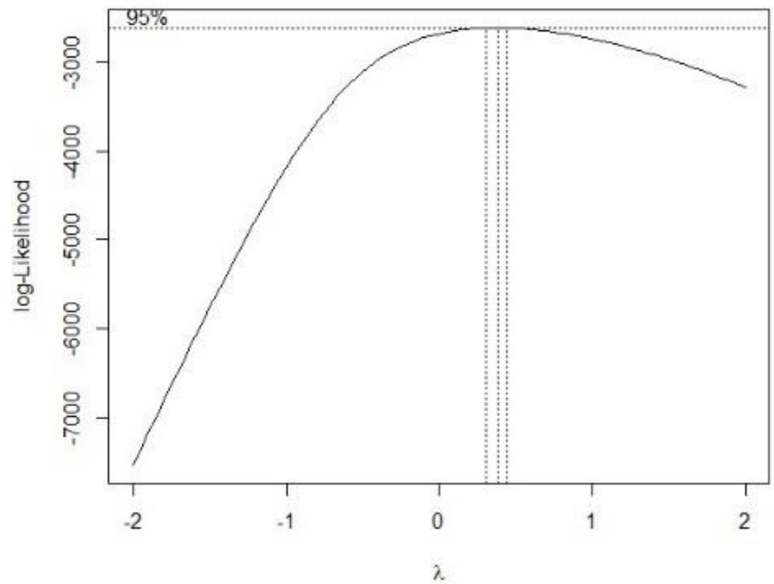
Limitations to this project include the lack of inclusion of other possible models. In such a large pool of models to test, there could have been multiple other models that may have been a better fit for the data. In the data generated, it was under ideal conditions that produced a cleanly fit model. In capricious real-world situations, the model produced may not be as clean given all its confounding and nuisance variables that could have affected the data. Furthermore, I was only able to test up to the 2nd interaction, whereas there may have been more interactions that I could have possibly tested.

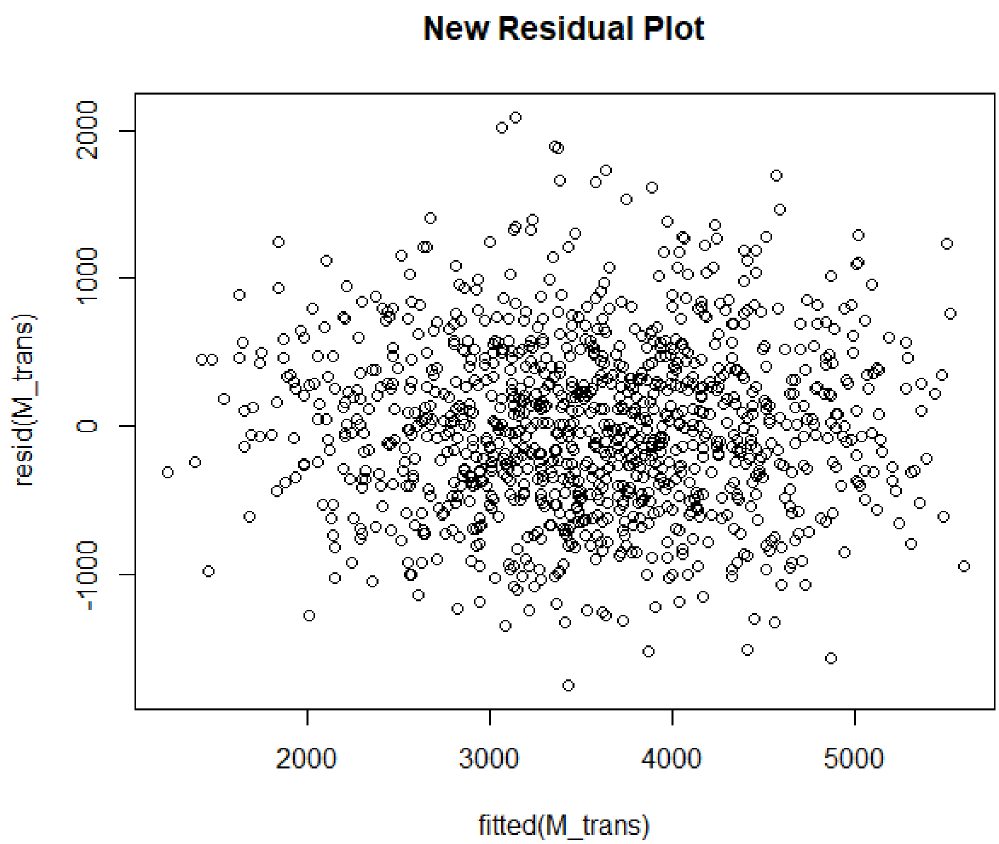
Appendix

First trial

In my first trial, I ran my code with my data raised to the 2nd power $(.)^2$to account for interactions. These were the following results:







In the first attempts, after running with (.)^2 or two interactions:

Model Summary		
model	**adjR2**	**BIC**
(Intercept)+E2:E3	0.511832018864929	-761.495323672839
(Intercept)+E2:E3+E3:G17	0.52497485156692404411	-784.9877351
(Intercept)+E2:E3+E3:G17+G3:G20	0.527319847074736	-784.351332196778
(Intercept)+E2:E3+E3:G17+G3:G20+G12:G13	0.528810750877091	-781.782664958267
(Intercept)+E2:E3+E3:G17+G3:G20+G12:G13+G12:G20	0.530353810131511	-779.345675184741

Sig Coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
E2	236.8734	10.59716	22.352532	0.0e+00
E3	266.6243	10.33195	25.805792	0.0e+00
G17	248.2914	53.31335	4.657208	3.6e-06

2nd Interaction

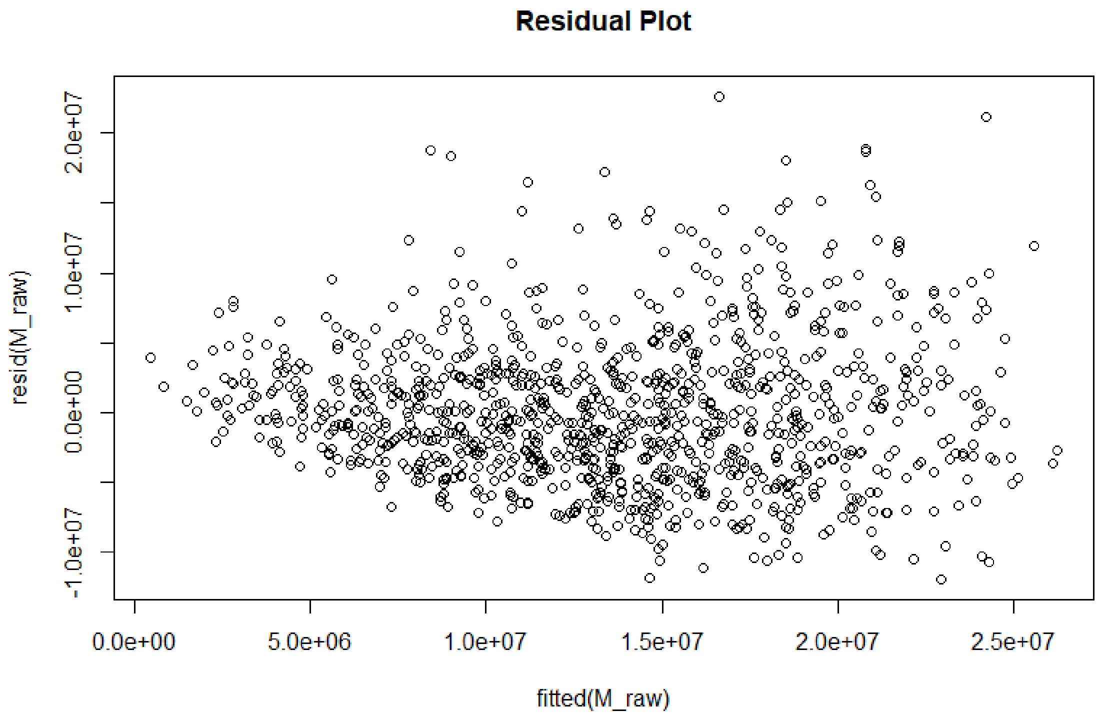
	Estimate	Std. Error	t value	Pr(>|t|)
E2	450.22196	153.1956	2.938871	0.0033914
E3:G2	84.37606	31.7837	2.654696	0.0081000
G12:G20	459.65138	171.8968	2.673997	0.0076524

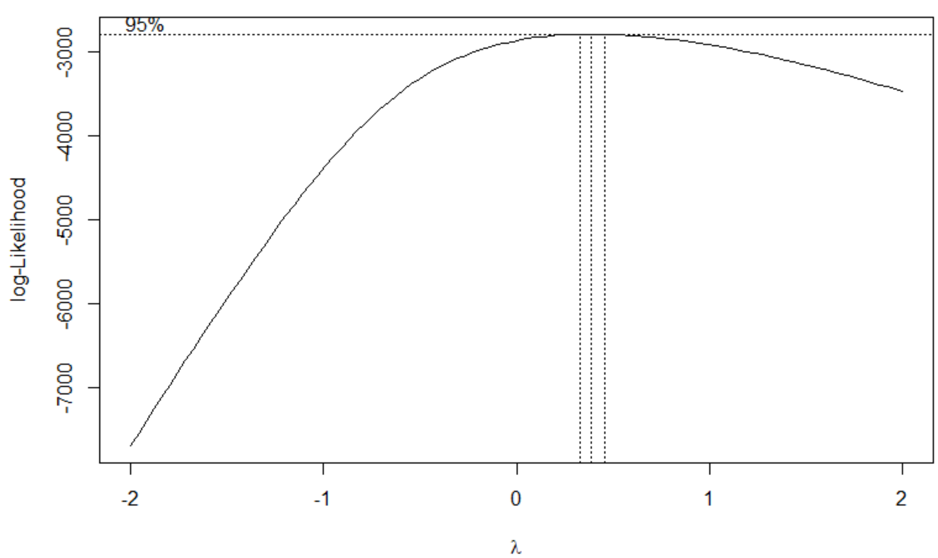
	Estimate	Std. Error	t value	Pr(>|t|)
E2	220.3944	100.82157	2.185985	0.0290325
E3	227.6567	99.62204	2.285204	0.0224954

I realized that this would not work since my significant coefficients show a considerably large t-value for G17 but G17 was not included in the 2^{nd} interaction graph

At this point I changed it to a linear form that accounted for the variables without the interaction effects, by changing (.)^2 to (.)

Second Trial → CURRENT CODE





From the above graph, I was able to determine that lambda = 0.5, which allowed me to use $Y^{0.5}$ for my boxcox transformation.

The residual plot produced a form that was closer to being a “flat ellipse”.

Afterwards, I was able to create the following tables below

Model Summary

model	adjR2	BIC
(Intercept)+E3	0.310041050142424	-387.844988462386
(Intercept)+E2+E3	0.529894771830186	-796.232179373285
(Intercept)+E2+E3+G17	0.539809220740668	-813.27136621407
(Intercept)+E2+E3+G3+G17	0.541049665407465	-810.205905611854
(Intercept)+E2+E3+G3+G13+G17	0.541495423885762	-805.275774736494

Sig Coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
E2	236.8734	10.59716	22.352532	0.0e+00
E3	266.6243	10.33195	25.805792	0.0e+00
G17	248.2914	53.31335	4.657208	3.6e-06

2nd Interaction

	Estimate	Std. Error	t value	Pr(>|t|)
E2	450.22196	153.1956	2.938871	0.0033914
E3:G2	84.37606	31.7837	2.654696	0.0081000
G12:G20	459.65138	171.8968	2.673997	0.0076524

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-942.8576	130.10305	-7.247006	0e+00
E2	236.4378	10.48283	22.554760	0e+00
E3	268.9584	10.18812	26.399204	0e+00
G17	259.0149	52.64888	4.919667	1e-06

From this, I was able to generate my final mode:

$$Y = -942.8576 + 236.4378E2 + 268.9584E3 + 259.0149G17$$

Code

```
#Project 2, AMS 315
#First title the file given to me
Dat <- read.csv('P2_37466.csv', header=TRUE)
#Fit a model with only environmental variables
M_E <- lm(Y ~ E1+E2+E3+E4, data=Dat)
print(summary(M_E))
##Cal:
##lm(formula = Y ~ E1 + E2 + E3 + E4, data = Dat)
##
##Residuals:
##   Min        1Q    Median       3Q       Max
##-11993182  -3528162   -508869   2802659  23941989
##
##Coefficients:
##   Estimate Std. Error t value Pr(>|t|)
##(Intercept) -16529068    1354085 -12.207   <2e-16 ***
##   E1            -28851      80244  -0.360    0.719
##   E2           1677173      80351  20.873   <2e-16 ***
##   E3           1922821      78213  24.584   <2e-16 ***
##   E4            -13708      82188  -0.167    0.868
##---
##   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##Residual standard error: 5268000 on 1075 degrees of freedom
##Multiple R-squared:  0.4981,  Adjusted R-squared:  0.4963
##F-statistic: 266.7 on 4 and 1075 DF,  p-value: < 2.2e-16

#Find the adjusted R squared value
print(summary(M_E)$adj.r.squared)
##[1] 0.4962557

#Now we control environmental variables and assess for the contribution of genetic variables
paste(paste0('E', 1:5), collapse = '+')
M_raw <- lm( Y ~ (E1+E2+E3+E4+G1+G2+G3+G4+G5+G6+G7+G8+G9+G10+G11+G12+G13+G14+G15+G16+G17+G18+G19+G20), data = Dat )
#We will create and examine the residual plot
plot(resid(M_raw) ~ fitted(M_raw), main='Residual Plot')
library(MASS)
boxcox(M_raw)
#Because my estimated lambda value appears to be around 0.5, I use Y^0.5
M_trans <- lm( I(Y^0.5) ~ (.), data=Dat )
print(summary(M_raw)$adj.r.square)
##[1] 0.5039505
#Because my r-squared value increased, this shows that the transformation created a more optimal graph
print(summary(M_trans)$adj.r.square)
##[1] 0.5385358
#Plot the transformed graph now
#plot the residual graph now
plot(resid(M_trans) ~fitted(M_trans), main='New Residual Plot')

#I made sure to install the package for leaps
#install.packages("leaps")
library(leaps)
M <- regsubsets( model.matrix(M_trans)[,-1], I((Dat$Y)^(1/2)),
                 nbest = 1 , nvmax=5,
                 method = 'forward', intercept = TRUE )
temp <- summary(M)
```

```
#print(temp)
##Subset selection object
#install.packages("knitr")
#install.packages("kableExtra")
library(kableExtra)
library(knitr)
Var <- colnames(model.matrix(M_trans))
M_select <- apply(temp$which, 1,
                  function(x) paste0(Var[x], collapse='+'))
print(kable(data.frame(cbind( model = M_select, adjR2 = temp$adjr2, BIC = temp$bic)),
            caption='Model Summary'))
##<table>
##  <caption>Model Summary</caption>
##  <thead>
##  <tr>
##  <th style="text-align:left;"> model </th>
##  <th style="text-align:left;"> adjR2 </th>
##  <th style="text-align:left;"> BIC </th>
##  </tr>
##  </thead>
##  <tbody>
##  <tr>
##  <td style="text-align:left;"> (Intercept)+E3 </td>
##  <td style="text-align:left;"> 0.310041050142424 </td>
##  <td style="text-align:left;"> -387.844988462386 </td>
##  </tr>
##  <tr>
##  <td style="text-align:left;"> (Intercept)+E2+E3 </td>
##  <td style="text-align:left;"> 0.529894771830186 </td>
##  <td style="text-align:left;"> -796.232179373285 </td>
##  </tr>
##  <tr>
##  <td style="text-align:left;"> (Intercept)+E2+E3+G17 </td>
##  <td style="text-align:left;"> 0.539809220740668 </td>
##  <td style="text-align:left;"> -813.27136621407 </td>
##  </tr>
##  <tr>
##  <td style="text-align:left;"> (Intercept)+E2+E3+G3+G17 </td>
##  <td style="text-align:left;"> 0.541049665407465 </td>
##  <td style="text-align:left;"> -810.205905611854 </td>
##  </tr>
##  <tr>
##  <td style="text-align:left;"> (Intercept)+E2+E3+G3+G13+G17 </td>
##  <td style="text-align:left;"> 0.541495423885762 </td>
##  <td style="text-align:left;"> -805.275774736494 </td>
##  </tr>
##  </tbody>
##  </table>
M_main <- lm( I((Y)^(1/2)) ~ ., data=Dat)
temp <- summary(M_main)
print(kable(temp$coefficients[ abs(temp$coefficients[,4]) <= 0.001, ], caption='Sig Coefficients'))
##<table>
##  <caption>Sig Coefficients</caption>
##  <thead>
##  <tr>
##  <th style="text-align:left;">   </th>
```

```
## <th style="text-align:right;"> Estimate </th>
## <th style="text-align:right;"> Std. Error </th>
## <th style="text-align:right;"> t value </th>
## <th style="text-align:right;"> Pr(&gt;|t|) </th>
## </tr>
## </thead>
## <tbody>
## <tr>
## <td style="text-align:left;"> E2 </td>
## <td style="text-align:right;"> 236.8734 </td>
## <td style="text-align:right;"> 10.59716 </td>
## <td style="text-align:right;"> 22.352532 </td>
## <td style="text-align:right;"> 0.0e+00 </td>
## </tr>
## <tr>
## <td style="text-align:left;"> E3 </td>
## <td style="text-align:right;"> 266.6243 </td>
## <td style="text-align:right;"> 10.33195 </td>
## <td style="text-align:right;"> 25.805792 </td>
## <td style="text-align:right;"> 0.0e+00 </td>
## </tr>
## <tr>
## <td style="text-align:left;"> G17 </td>
## <td style="text-align:right;"> 248.2914 </td>
## <td style="text-align:right;"> 53.31335 </td>
## <td style="text-align:right;"> 4.657208 </td>
## <td style="text-align:right;"> 3.6e-06 </td>
## </tr>
## </tbody>
## </table>
M_2nd <- lm( I((Y)^(1/2)) ~ (.)^2, data=Dat)
temp  <- summary(M_2nd)
print(kable(temp$coefficients[ abs(temp$coefficients[,4]) <= 0.01, ], caption='2nd Interaction'))
## <table>
## <caption>2nd Interaction</caption>
## <thead>
## <tr>
## <th style="text-align:left;">   </th>
## <th style="text-align:right;"> Estimate </th>
## <th style="text-align:right;"> Std. Error </th>
## <th style="text-align:right;"> t value </th>
## <th style="text-align:right;"> Pr(&gt;|t|) </th>
## </tr>
## </thead>
## <tbody>
## <tr>
## <td style="text-align:left;"> E2 </td>
## <td style="text-align:right;"> 450.22196 </td>
## <td style="text-align:right;"> 153.1956 </td>
## <td style="text-align:right;"> 2.938871 </td>
## <td style="text-align:right;"> 0.0033914 </td>
## </tr>
## <tr>
## <td style="text-align:left;"> E3:G2 </td>
## <td style="text-align:right;"> 84.37606 </td>
## <td style="text-align:right;"> 31.7837 </td>
```

#Project 2, AMS 315

#First title the file given to me

Dat <- read.csv('P2_37466.csv', header=TRUE)

#Fit a model with only environmental variables

M_E <- lm(Y ~ E1+E2+E3+E4, data=Dat)

print(summary(M_E))

##Cal:

##lm(formula = Y ~ E1 + E2 + E3 + E4, data = Dat)

##

```
##Residuals:
##  Min       1Q    Median       3Q      Max
##-11993182  -3528162   -508869   2802659  23941989
##
##Coefficients:
##  Estimate Std. Error t value Pr(>|t|)
##(Intercept) -16529068    1354085 -12.207   <2e-16 ***
##  E1          -28851      80244  -0.360    0.719
##  E2         1677173      80351  20.873   <2e-16 ***
##  E3         1922821      78213  24.584   <2e-16 ***
##  E4          -13708      82188  -0.167    0.868
##---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##Residual standard error: 5268000 on 1075 degrees of freedom
##Multiple R-squared:  0.4981,   Adjusted R-squared:  0.4963
##F-statistic: 266.7 on 4 and 1075 DF,  p-value: < 2.2e-16


#Find the adjusted R squared value
print(summary(M_E)$adj.r.squared)
##[1] 0.4962557


#Now we control environmental variables and assess for the contribution of genetic variables
paste(paste0('E', 1:5), collapse = '+')
M_raw <- lm( Y ~
(E1+E2+E3+E4+G1+G2+G3+G4+G5+G6+G7+G8+G9+G10+G11+G12+G13+G14+G15+G16+G17+G18+G19+
G20), data = Dat )
#We will create and examine the residual plot
plot(resid(M_raw) ~ fitted(M_raw), main='Residual Plot')
```

```
library(MASS)

boxcox(M_raw)

#Because my estimated lambda value appears to be around 0.5, I use Y^0.5

M_trans <- lm( I(Y^0.5) ~ (.), data=Dat )

print(summary(M_raw)$adj.r.square)

##[1] 0.5039505

#Because my r-squared value increased, this shows that the transformation created a more optimal graph

print(summary(M_trans)$adj.r.square)

##[1] 0.5385358

#Plot the transformed graph now

#plot the residual graph now

plot(resid(M_trans) ~fitted(M_trans), main='New Residual Plot')


#I made sure to install the package for leaps

#install.packages("leaps")

library(leaps)

M <- regsubsets( model.matrix(M_trans)[,-1], I((Dat$Y)^(1/2)),

        nbest = 1 , nvmax=5,

        method = 'forward', intercept = TRUE )

temp <- summary(M)

#print(temp)

##Subset selection object

#install.packages("knitr")

#install.packages("kableExtra")

library(kableExtra)

library(knitr)

Var <- colnames(model.matrix(M_trans))

M_select <- apply(temp$which, 1,
```

```
            function(x) paste0(Var[x], collapse='+'))
print(kable(data.frame(cbind( model = M_select, adjR2 = temp$adjr2, BIC = temp$bic)),
        caption='Model Summary'))
##<table>
## <caption>Model Summary</caption>
## <thead>
## <tr>
## <th style="text-align:left;"> model </th>
## <th style="text-align:left;"> adjR2 </th>
## <th style="text-align:left;"> BIC </th>
## </tr>
## </thead>
## <tbody>
## <tr>
## <td style="text-align:left;"> (Intercept)+E3 </td>
## <td style="text-align:left;"> 0.310041050142424 </td>
## <td style="text-align:left;"> -387.844988462386 </td>
## </tr>
## <tr>
## <td style="text-align:left;"> (Intercept)+E2+E3 </td>
## <td style="text-align:left;"> 0.529894771830186 </td>
## <td style="text-align:left;"> -796.232179373285 </td>
## </tr>
## <tr>
## <td style="text-align:left;"> (Intercept)+E2+E3+G17 </td>
## <td style="text-align:left;"> 0.539809220740668 </td>
## <td style="text-align:left;"> -813.27136621407 </td>
## </tr>
## <tr>
```

```
##  <td style="text-align:left;"> (Intercept)+E2+E3+G3+G17 </td>
##  <td style="text-align:left;"> 0.541049665407465 </td>
##  <td style="text-align:left;"> -810.205905611854 </td>
##  </tr>
##  <tr>
##  <td style="text-align:left;"> (Intercept)+E2+E3+G3+G13+G17 </td>
##  <td style="text-align:left;"> 0.541495423885762 </td>
##  <td style="text-align:left;"> -805.275774736494 </td>
##  </tr>
##  </tbody>
##  </table>
M_main <- lm( I((Y)^(1/2)) ~ ., data=Dat)
temp <- summary(M_main)
print(kable(temp$coefficients[ abs(temp$coefficients[,4]) <= 0.001, ], caption='Sig Coefficients'))
##<table>
##  <caption>Sig Coefficients</caption>
##  <thead>
##  <tr>
##  <th style="text-align:left;">   </th>
##  <th style="text-align:right;"> Estimate </th>
##  <th style="text-align:right;"> Std. Error </th>
##  <th style="text-align:right;"> t value </th>
##  <th style="text-align:right;"> Pr(&gt;|t|) </th>
##  </tr>
##  </thead>
##  <tbody>
##  <tr>
##  <td style="text-align:left;"> E2 </td>
##  <td style="text-align:right;"> 236.8734 </td>
```

```
## <td style="text-align:right;"> 10.59716 </td>
## <td style="text-align:right;"> 22.352532 </td>
## <td style="text-align:right;"> 0.0e+00 </td>
## </tr>
## <tr>
## <td style="text-align:left;"> E3 </td>
## <td style="text-align:right;"> 266.6243 </td>
## <td style="text-align:right;"> 10.33195 </td>
## <td style="text-align:right;"> 25.805792 </td>
## <td style="text-align:right;"> 0.0e+00 </td>
## </tr>
## <tr>
## <td style="text-align:left;"> G17 </td>
## <td style="text-align:right;"> 248.2914 </td>
## <td style="text-align:right;"> 53.31335 </td>
## <td style="text-align:right;"> 4.657208 </td>
## <td style="text-align:right;"> 3.6e-06 </td>
## </tr>
## </tbody>
## </table>
M_2nd <- lm( I((Y)^(1/2)) ~ (.)^2, data=Dat)
temp  <- summary(M_2nd)
print(kable(temp$coefficients[ abs(temp$coefficients[,4]) <= 0.01, ], caption='2nd Interaction'))
## <table>
## <caption>2nd Interaction</caption>
## <thead>
## <tr>
## <th style="text-align:left;">   </th>
## <th style="text-align:right;"> Estimate </th>
```

```
## <th style="text-align:right;"> Std. Error </th>
## <th style="text-align:right;"> t value </th>
## <th style="text-align:right;"> Pr(&gt;|t|) </th>
## </tr>
## </thead>
## <tbody>
## <tr>
## <td style="text-align:left;"> E2 </td>
## <td style="text-align:right;"> 450.22196 </td>
## <td style="text-align:right;"> 153.1956 </td>
## <td style="text-align:right;"> 2.938871 </td>
## <td style="text-align:right;"> 0.0033914 </td>
## </tr>
## <tr>
## <td style="text-align:left;"> E3:G2 </td>
## <td style="text-align:right;"> 84.37606 </td>
## <td style="text-align:right;"> 31.7837 </td>
## <td style="text-align:right;"> 2.654696 </td>
## <td style="text-align:right;"> 0.0081000 </td>
## </tr>
## <tr>
## <td style="text-align:left;"> G12:G20 </td>
## <td style="text-align:right;"> 459.65138 </td>
## <td style="text-align:right;"> 171.8968 </td>
## <td style="text-align:right;"> 2.673997 </td>
## <td style="text-align:right;"> 0.0076524 </td>
## </tr>
## </tbody>
## </table>
```

```
M_2stage <- lm( I((Y)^(1/2)) ~ (E2+E3+G17), data=Dat)

temp <- summary(M_2stage)

print(kable(temp$coefficients[ abs(temp$coefficients[,3]) >= 4, ]))

##<table>
##  <thead>
##  <tr>
##  <th style="text-align:left;">   </th>
##  <th style="text-align:right;"> Estimate </th>
##  <th style="text-align:right;"> Std. Error </th>
##  <th style="text-align:right;"> t value </th>
##  <th style="text-align:right;"> Pr(&gt;|t|) </th>
##  </tr>
##  </thead>
##  <tbody>
##  <tr>
##  <td style="text-align:left;"> (Intercept) </td>
##  <td style="text-align:right;"> -942.8576 </td>
##  <td style="text-align:right;"> 130.10305 </td>
##  <td style="text-align:right;"> -7.247006 </td>
##  <td style="text-align:right;"> 0e+00 </td>
##  </tr>
##  <tr>
##  <td style="text-align:left;"> E2 </td>
##  <td style="text-align:right;"> 236.4378 </td>
##  <td style="text-align:right;"> 10.48283 </td>
##  <td style="text-align:right;"> 22.554760 </td>
##  <td style="text-align:right;"> 0e+00 </td>
##  </tr>
##  <tr>
```

<td style="text-align:left;"> E3 </td>

<td style="text-align:right;"> 268.9584 </td>

<td style="text-align:right;"> 10.18812 </td>

<td style="text-align:right;"> 26.399204 </td>

<td style="text-align:right;"> 0e+00 </td>

</tr>

<tr>

<td style="text-align:left;"> G17 </td>

<td style="text-align:right;"> 259.0149 </td>

<td style="text-align:right;"> 52.64888 </td>

<td style="text-align:right;"> 4.919667 </td>

<td style="text-align:right;"> 1e-06 </td>

</tr>

</tbody>

</table>

#We would use the above values to generate our final function for our picked model

End of Project Report #2