UnivariateHW

Sarah Kell

January 30, 2018

```
trees=read.csv("C:/Users/Sarah/Desktop/CofC/Quant_Methods/Univariate/treedata
_subset.csv")
head(trees)

##         plotID  spcode       species cover elev       tci streamdist
## 1 ATBN-01-0403 ABIEFRA Abies fraseri     1 1660  5.701460      490.9
## 2 ATBN-01-0532 ABIEFRA Abies fraseri     8 1712  3.823586      454.0
## 3 ATBN-01-0533 ABIEFRA Abies fraseri     3 1722  3.893762      453.4
## 4 ATBN-01-0536 ABIEFRA Abies fraseri     3 1754  3.145527      492.5
## 5 FRID-01-0003 ABIEFRA Abies fraseri     5 1570 11.850000        0.0
## 6 PITT-01-0045 ABIEFRA Abies fraseri     2 1504  4.373741      237.1
##   disturb     beers
## 1 CORPLOG 0.2244286
## 2  VIRGIN 0.8340878
## 3  LT-SEL 1.3332586
## 4  SETTLE 1.4712484
## 5  LT-SEL 0.4961189
## 6  VIRGIN 1.6558421

Acer<-trees[trees$species=="Acer rubrum",]
Abies<-trees[trees$species=="Abies fraseri",]
library(car)
```

1. Exploratory analysis of species cover

*The exploratory linear model for Abies with all the variables included does not do a great job at explaining species abundance (R2 adjusted = 0.501). Elevation is the only variable with a siginificant effect.

The linear model for Acer is not good (R2 adjusted=0.036). It does tell you however which variables have an significant effect (elevation, distance to stream and heat load index) on abundance to explore more.

```
Abies_ols<-lm(Abies$cover~Abies$elev+Abies$tci+Abies$streamdist+Abies$disturb
+Abies$beers)
Anova(Abies_ols, type=3)

## Anova Table (Type III tests)
##
## Response: Abies$cover
##                  Sum Sq Df F value    Pr(>F)
## (Intercept)      59.401  1 23.1710 2.652e-05 ***
## Abies$elev       61.618  1 24.0358 2.022e-05 ***
```

```
## Abies$tci         5.667  1  2.2105    0.1458
## Abies$streamdist  1.636  1  0.6382    0.4296
## Abies$disturb    10.089  3  1.3118    0.2855
## Abies$beers       0.014  1  0.0056    0.9406
## Residuals        92.289 36
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Abies_ols)
```

```
##
## Call:
## lm(formula = Abies$cover ~ Abies$elev + Abies$tci + Abies$streamdist +
##     Abies$disturb + Abies$beers)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4630 -0.6472  0.0788  1.0872  3.8017
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -20.561173   4.271449  -4.814 2.65e-05 ***
## Abies$elev           0.012370   0.002523   4.903 2.02e-05 ***
## Abies$tci            0.287641   0.193467   1.487   0.1458
## Abies$streamdist    -0.001266   0.001585  -0.799   0.4296
## Abies$disturbLT-SEL  2.188367   2.097905   1.043   0.3038
## Abies$disturbSETTLE  1.527604   2.341471   0.652   0.5183
## Abies$disturbVIRGIN  3.025596   1.735921   1.743   0.0899 .
## Abies$beers          0.037551   0.500269   0.075   0.9406
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.601 on 36 degrees of freedom
## Multiple R-squared:  0.5824, Adjusted R-squared:  0.5011
## F-statistic: 7.171 on 7 and 36 DF,  p-value: 2.215e-05
```

```
Acer_ols<-lm(Acer$cover~Acer$elev+Acer$tci+Acer$streamdist+Acer$disturb+Acer$
beers)
Anova(Acer_ols, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: Acer$cover
##                  Sum Sq  Df  F value    Pr(>F)
## (Intercept)      765.43   1 193.5096 < 2.2e-16 ***
## Acer$elev         40.44   1  10.2233  0.001448 **
## Acer$tci          12.58   1   3.1805  0.074947 .
## Acer$streamdist   29.09   1   7.3531  0.006856 **
## Acer$disturb       9.45   3   0.7962  0.496166
## Acer$beers        35.61   1   9.0034  0.002789 **
## Residuals       2828.21 715
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(Acer_ols)
```

```
##
## Call:
## lm(formula = Acer$cover ~ Acer$elev + Acer$tci + Acer$streamdist +
##     Acer$disturb + Acer$beers)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7073 -1.2446  0.3409  1.3575  5.2732
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          6.3502303  0.4564973  13.911  < 2e-16 ***
## Acer$elev           -0.0010108  0.0003161  -3.197  0.00145 **
## Acer$tci            -0.0627613  0.0351922  -1.783  0.07495 .
## Acer$streamdist      0.0012895  0.0004756   2.712  0.00686 **
## Acer$disturbLT-SEL   0.0829610  0.2166747   0.383  0.70192
## Acer$disturbSETTLE  -0.1044556  0.2804213  -0.372  0.70963
## Acer$disturbVIRGIN   0.3088364  0.2518161   1.226  0.22044
## Acer$beers          -0.3269597  0.1089662  -3.001  0.00279 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.989 on 715 degrees of freedom
## Multiple R-squared:  0.04493,    Adjusted R-squared:  0.03558
## F-statistic: 4.805 on 7 and 715 DF,  p-value: 2.669e-05
```

*Which explanatory variables are the most important?

For Abies - While elevation has the greatest effect on abundance, site water potential is also important. Adding disturbance in lowers the AIC score a little but only due to the virgin ‘undisturbed’ areas which makes sense

```r
Abies_olsmod<-lm(Abies$cover~Abies$elev)
Abies_olsmod2<-lm(Abies$cover~Abies$elev+Abies$disturb)
Abies_olsmod3<-lm(Abies$cover~Abies$elev+Abies$tci)
Abies_olsmod4<-lm(Abies$cover~Abies$elev+Abies$tci+Abies$disturb)
AIC(Abies_olsmod)
```

```
## [1] 173.2266
```

```r
AIC(Abies_olsmod2)
```

```
## [1] 173.3299
```

```r
AIC(Abies_olsmod3) #best model
```

```
## [1] 171.5494
```

```
AIC(Abies_olsmod4) #also looked at Adj R2 by summary

## [1] 172.2526
```

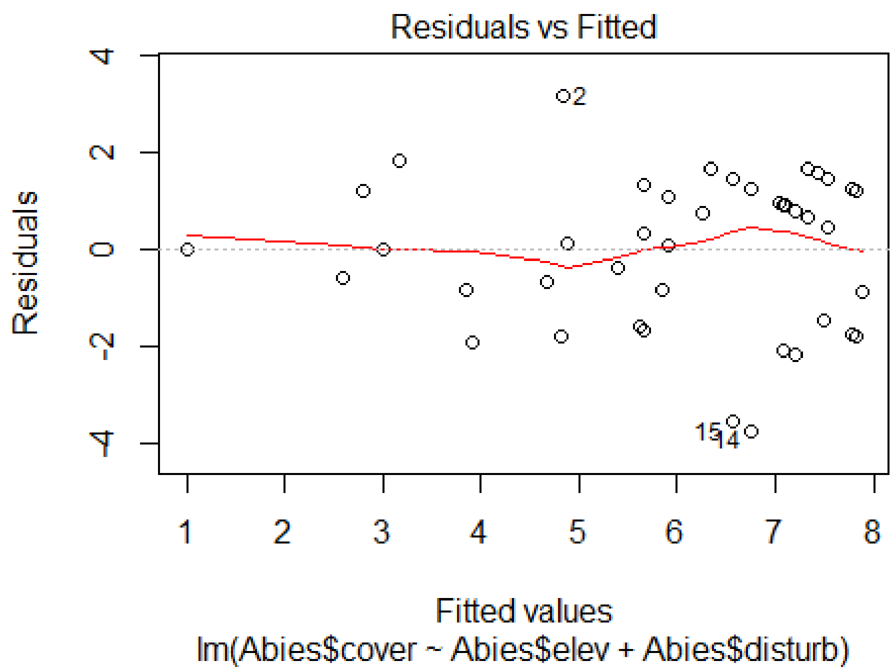
For Acer - Elevation, distance to stream and heat load index are important. The interaction of the three variables has the greatest effect on abundance

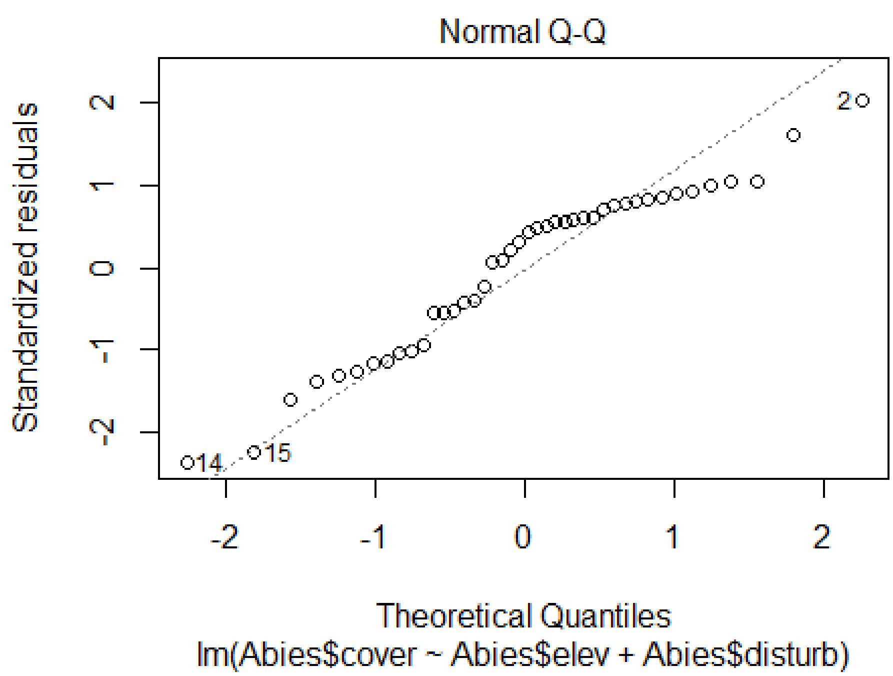
```
Acer_olsmod<-lm(Acer$cover~Acer$elev+Acer$streamdist+Acer$beers)
Acer_olsmod1<-lm(Acer$cover~Acer$elev+Acer$streamdist+Acer$elev+Acer$elev*Ace
r$streamdist+Acer$elev*Acer$beers+Acer$elev*Acer$beers+Acer$elev*Acer$streamd
ist*Acer$beers)
AIC(Acer_olsmod)

## [1] 3054.014

AIC(Acer_olsmod1)#best model

## [1] 3037.797
```

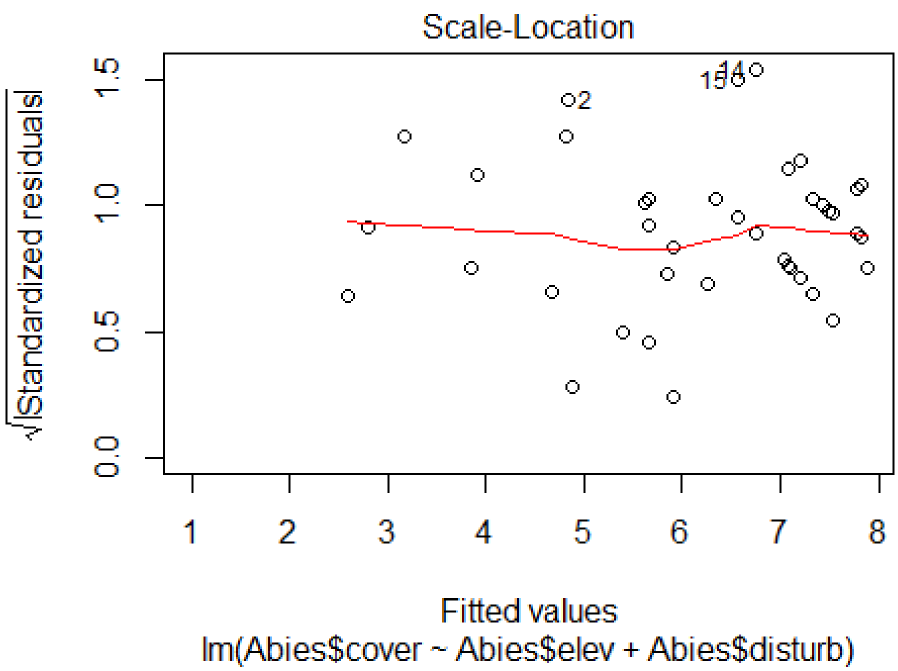
*Do model diagnositcs indicate any problems with violations of OLS assumptions?

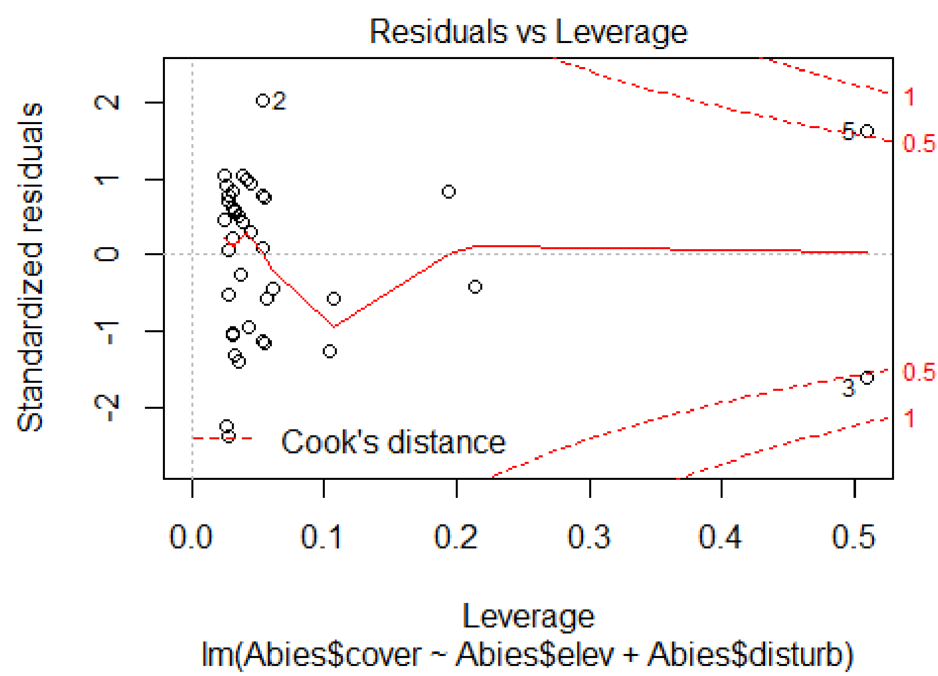
Yes- data is not normal for both species

```
plot(Abies_olsmod2)
```

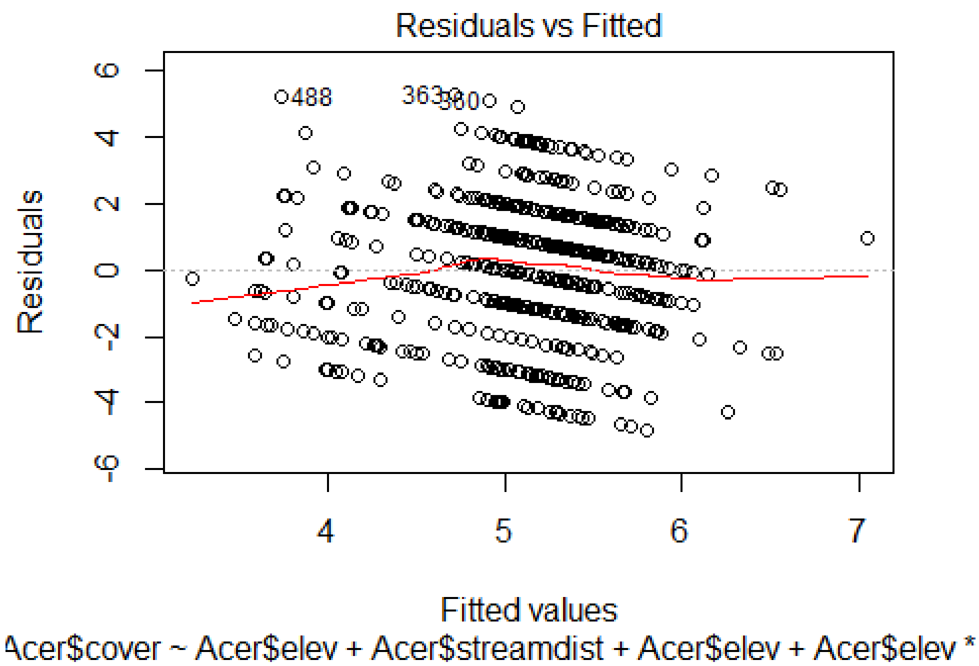

Residuals vs Fitted
Residuals
Fitted values
lm(Abies$cover ~ Abies$elev + Abies$disturb)
2
15
14

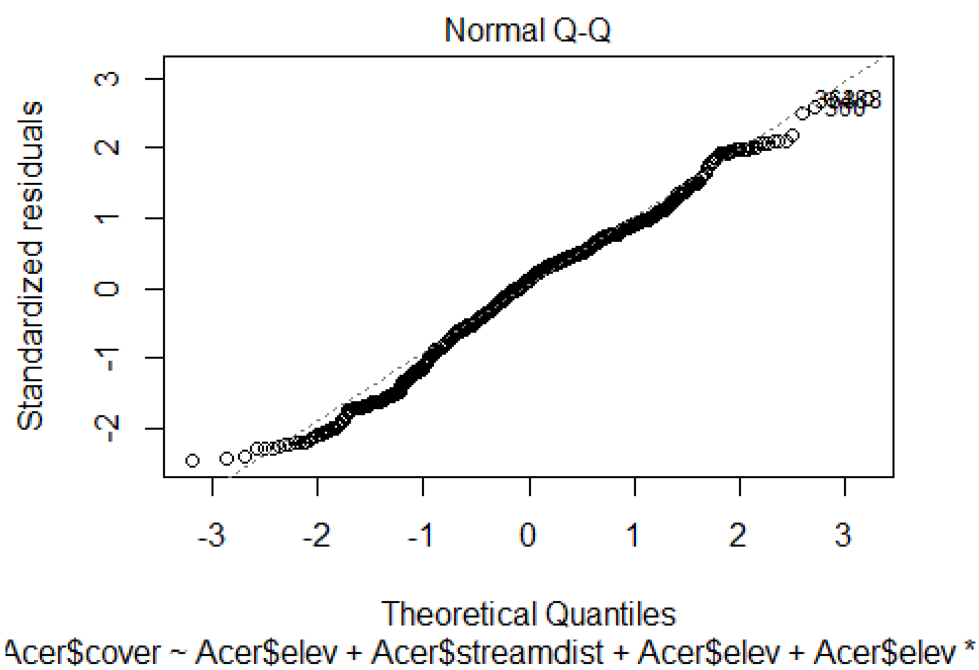

Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(Abies$cover ~ Abies$elev + Abies$disturb)
2
14
15

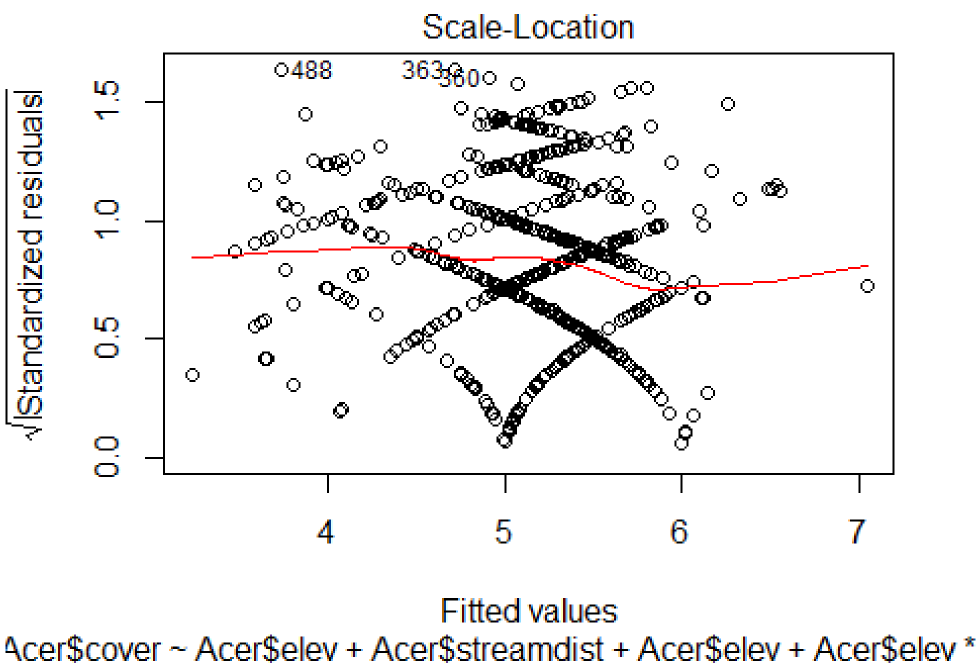

Scale-Location
√|Standardized residuals|
Fitted values
lm(Abies$cover ~ Abies$elev + Abies$disturb)

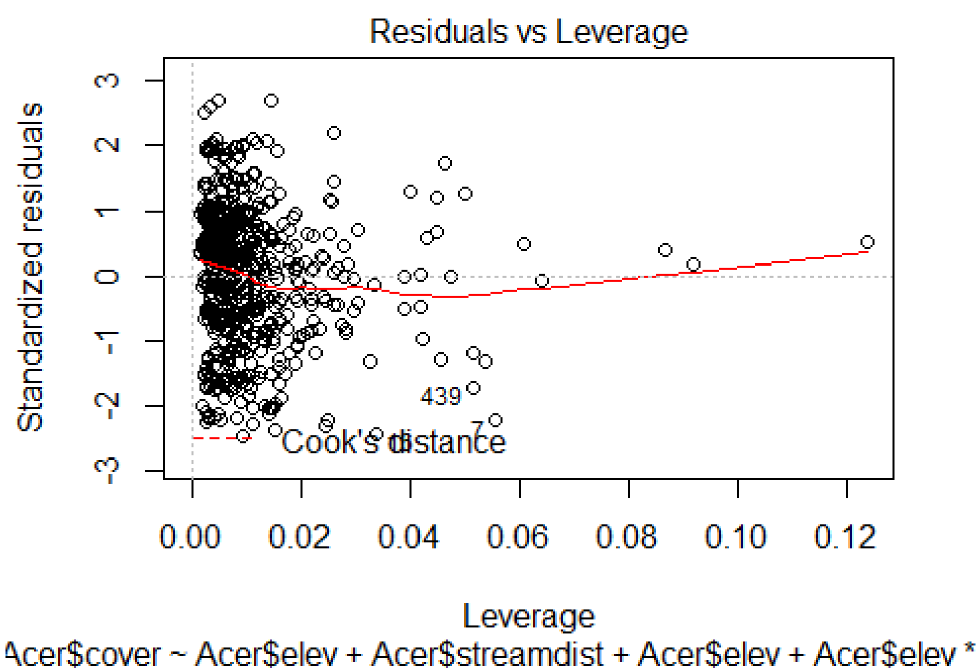

Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage
lm(Abies$cover ~ Abies$elev + Abies$disturb)

```r
plot(Acer_olsmod1)
```






Scale-Location
√|Standardized residuals|
Fitted values
Acer$cover ~ Acer$elev + Acer$streamdist + Acer$elev + Acer$elev *
488
363
360


Residuals vs Leverage
Standardized residuals
Leverage
Acer$cover ~ Acer$elev + Acer$streamdist + Acer$elev + Acer$elev *
Cook's distance
439

*Are you able to explain variance in one species better than another- Why? Yes, elevation for Abies is able explain abundance much better than the model with several variables for Acer. The interaction of variables is often harder to explain and predict vs. one variable that exerts a strong effect.

2. Using General Linear Model **used the best fit lm models, also best fit here

```
Abies_glm_mod<-glm(Abies$cover~Abies$elev+Abies$tci, family='poisson') #site
water potential not as important now
Anova(Abies_glm_mod, type=3)

## Analysis of Deviance Table (Type III tests)
##
## Response: Abies$cover
##            LR Chisq Df Pr(>Chisq)
## Abies$elev  21.0649  1   4.44e-06 ***
## Abies$tci    2.1248  1     0.1449
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Acer_glm_mod<-glm(Acer$cover~Acer$elev+Acer$streamdist+Acer$elev+Acer$elev*Ac
er$streamdist+Acer$elev*Acer$beers+Acer$elev*Acer$beers+Acer$elev*Acer$stream
dist*Acer$beers, family = 'poisson')
Anova(Acer_glm_mod, type=3) #streamdistance is not longer significant

## Analysis of Deviance Table (Type III tests)
##
## Response: Acer$cover
##                                     LR Chisq Df Pr(>Chisq)
## Acer$elev                             2.8791  1   0.089736 .
## Acer$streamdist                       0.9151  1   0.338764
## Acer$beers                            5.5316  1   0.018676 *
## Acer$elev:Acer$streamdist             0.5028  1   0.478288
## Acer$elev:Acer$beers                 10.0522  1   0.001522 **
## Acer$streamdist:Acer$beers            0.0423  1   0.837068
## Acer$elev:Acer$streamdist:Acer$beers  0.4725  1   0.491819
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC(Abies_glm_mod)

## [1] 183.2316

AIC(Acer_glm_mod)

## [1] 3086.985
```

Assessing degree of variation by pseudo-R2 **General function, so that any glm model used will be called in when adding your actual model name in after the pseudo R2

```
pseudo_r2 = function(glm_mod) {
     1 -  glm_mod$deviance / glm_mod$null.deviance
```

```
}
pseudo_r2(Abies_glm_mod)

## [1] 0.5140995

pseudo_r2(Acer_glm_mod) ## [1] 0.06274571
```

Comparing OLS and glm models

```
anova(Abies_olsmod2, Abies_glm_mod) #glm model does a better job at lowering
the residual (unexplained) variance

## Analysis of Variance Table
##
## Model 1: Abies$cover ~ Abies$elev + Abies$disturb
## Model 2: Abies$cover ~ Abies$elev + Abies$tci
##   Res.Df     RSS Df Sum of Sq F Pr(>F)
## 1     39 100.776
## 2     41  20.055 -2    80.721

anova(Acer_olsmod1, Acer_glm_mod) #same as above

## Analysis of Variance Table
##
## Model 1: Acer$cover ~ Acer$elev + Acer$streamdist + Acer$elev + Acer$elev
*
##     Acer$streamdist + Acer$elev * Acer$beers + Acer$elev * Acer$beers +
##     Acer$elev * Acer$streamdist * Acer$beers
## Model 2: Acer$cover ~ Acer$elev + Acer$streamdist + Acer$elev + Acer$elev
*
##     Acer$streamdist + Acer$elev * Acer$beers + Acer$elev * Acer$beers +
##     Acer$elev * Acer$streamdist * Acer$beers
##   Res.Df    RSS Df Sum of Sq F Pr(>F)
## 1    715 2758.1
## 2    715  608.6  0    2149.5
```

*Did changing the error distribution change the results much? The glm models were able to lower the residual (unexplained) variance but the adjusted R2 values were pretty much the same. In the Acer glm model, stream distance is no longer an important variable, in the Abies glm model site water potential is not as important.

3. Summary of findings Red maples are habitat generalists which can tolerate a wide range of habitat conditions. Species abundance appears to be affected the most by the combination of elevation and heat load index.

   Abies fraseri are habitat specialists which limits their abundance due to needing specific conditions to thrive. Elevation is the most important factor in limiting their distribution and abundance.

5. Step() function Abies fraseri

```r
library(MASS)
stepAIC(Abies_ols)
```

```
## Start:  AIC=48.59
## Abies$cover ~ Abies$elev + Abies$tci + Abies$streamdist + Abies$disturb +
##     Abies$beers
##
##                    Df Sum of Sq     RSS    AIC
## - Abies$beers       1     0.014  92.304 46.599
## - Abies$disturb     3    10.089 102.379 47.157
## - Abies$streamdist  1     1.636  93.926 47.366
## <none>                           92.289 48.593
## - Abies$tci         1     5.667  97.956 49.215
## - Abies$elev        1    61.618 153.908 69.095
##
## Step:  AIC=46.6
## Abies$cover ~ Abies$elev + Abies$tci + Abies$streamdist + Abies$disturb
##
##                    Df Sum of Sq     RSS    AIC
## - Abies$streamdist  1     1.665  93.969 45.386
## - Abies$disturb     3    10.679 102.983 45.417
## <none>                           92.304 46.599
## - Abies$tci         1     6.745  99.049 47.703
## - Abies$elev        1    64.662 156.966 67.961
##
## Step:  AIC=45.39
## Abies$cover ~ Abies$elev + Abies$tci + Abies$disturb
##
##                 Df Sum of Sq     RSS    AIC
## - Abies$disturb  3    12.021 105.990 44.683
## <none>                        93.969 45.386
## - Abies$tci      1     6.807 100.776 46.463
## - Abies$elev     1    78.687 172.656 70.153
##
## Step:  AIC=44.68
## Abies$cover ~ Abies$elev + Abies$tci
##
##               Df Sum of Sq    RSS    AIC
## <none>                     105.99 44.683
## - Abies$tci    1     9.239 115.23 46.360
## - Abies$elev   1   114.046 220.04 74.822

##
## Call:
## lm(formula = Abies$cover ~ Abies$elev + Abies$tci)
##
## Coefficients:
## (Intercept)   Abies$elev    Abies$tci
##   -18.78984      0.01262      0.30454
```

Acer rubrum

```
stepAIC(Acer_ols)

## Start:  AIC=1002.17
## Acer$cover ~ Acer$elev + Acer$tci + Acer$streamdist + Acer$disturb +
##     Acer$beers
##
##                   Df Sum of Sq    RSS     AIC
## - Acer$disturb     3     9.449 2837.7  998.58
## <none>                         2828.2 1002.17
## - Acer$tci         1    12.581 2840.8 1003.37
## - Acer$streamdist  1    29.085 2857.3 1007.56
## - Acer$beers       1    35.613 2863.8 1009.21
## - Acer$elev        1    40.439 2868.7 1010.43
##
## Step:  AIC=998.58
## Acer$cover ~ Acer$elev + Acer$tci + Acer$streamdist + Acer$beers
##
##                   Df Sum of Sq    RSS     AIC
## <none>                         2837.7  998.58
## - Acer$tci         1    14.370 2852.0 1000.23
## - Acer$streamdist  1    31.491 2869.2 1004.56
## - Acer$beers       1    35.515 2873.2 1005.57
## - Acer$elev        1    45.778 2883.4 1008.15

##
## Call:
## lm(formula = Acer$cover ~ Acer$elev + Acer$tci + Acer$streamdist +
##     Acer$beers)
##
## Coefficients:
##     (Intercept)        Acer$elev         Acer$tci  Acer$streamdist
##       6.3218898       -0.0008868       -0.0668631        0.0013256
##      Acer$beers
##      -0.3204370
```