06.10.15

Using linear mixed effects model to investigate leaf herbivory with percents asin(sqrt(x)) transformed and soil.ca included:

Linear mixed model fit by REML

t-tests use Satterthwaite approximations to degrees of freedom ['merModLmerTest']

Formula: leaf.pct.herb ~ pct.urban + leaf.pct.n + soil.no3.n + soil.nh4.n + soil.ca + (1 | site)

REML criterion at convergence: -100.5

Scaled residuals:

Min 1Q Median 3Q Max

-3.4918 -0.3746 -0.0740 0.3243 3.6651

Random effects:

Groups Name Variance Std.Dev.

site (Intercept) 0.0002212 0.01487

Residual 0.0034422 0.05867

Number of obs: 54, groups: site, 11

Fixed effects:

Estimate Std. Error df t value Pr(>|t|)

(Intercept) -1.185e-01 5.734e-02 4.640e+01 -2.067 0.044346 \*

pct.urban 2.852e-02 3.550e-02 9.450e+00 0.803 0.441560

leaf.pct.n 3.181e+00 2.701e+00 4.756e+01 1.177 0.244859

soil.no3.n 6.550e-03 2.177e-03 4.795e+01 3.008 0.004175 \*\*

soil.nh4.n -3.409e-03 1.428e-03 4.797e+01 -2.387 0.020963 \*

soil.ca 7.642e-06 1.893e-06 2.942e+01 4.038 0.000354 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:

(Intr) pct.rb lf.pc. sl.n3. sl.n4.

pct.urban -0.204

leaf.pct.n -0.862 0.036

soil.no3.n -0.199 -0.083 -0.182

soil.nh4.n 0.140 0.125 -0.099 -0.526

soil.ca 0.129 -0.426 -0.241 0.096 -0.151

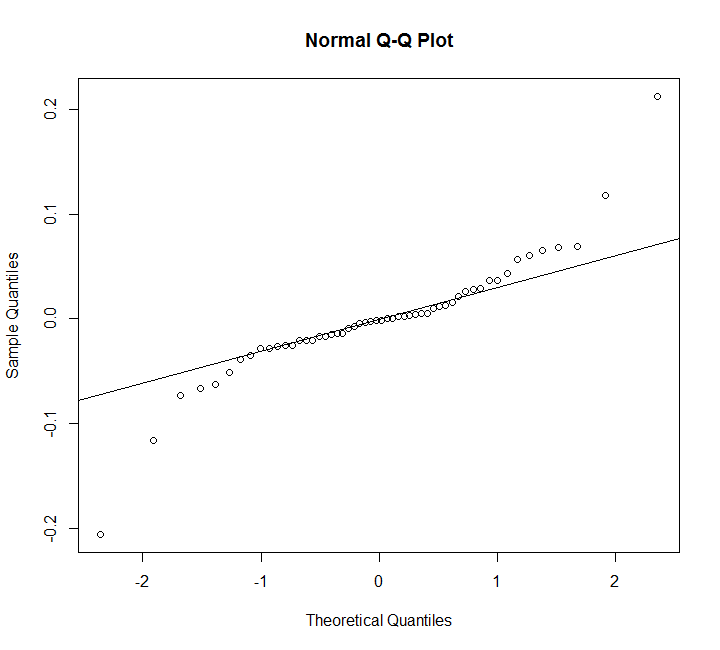
Warning messages:

1: Some predictor variables are on very different scales: consider rescaling

2: Some predictor variables are on very different scales: consider rescaling

3: Some predictor variables are on very different scales: consider rescaling

See comments on next page.

* I found that soil NO3-N, soil NH4-N, and soil Ca are significantly related to leaf herbivory
* Did the same thing with soil.p added, but it wasn’t significant
* So does this mean that my final equation for the influence of variables on leaf herbivory is:
  + leaf.pct.herb ~ soil.nox.n + soil.nh4.n + soil.ca + (1|site) ?
* QQplot for this model: 
* What do I do with this info now?

Also went back and looked at regular linear regression models with all possibly important variables:

Call:

lm(formula = leaf.pct.herb ~ pct.urban + leaf.pct.n + soil.no3.n +

soil.nh4.n + soil.ca)

Residuals:

Min 1Q Median 3Q Max

-0.201038 -0.023261 -0.001699 0.022768 0.218184

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.180e-01 5.686e-02 -2.075 0.0434 \*

pct.urban 2.706e-02 3.218e-02 0.841 0.4046

leaf.pct.n 3.159e+00 2.686e+00 1.176 0.2454

soil.no3.n 6.614e-03 2.181e-03 3.033 0.0039 \*\*

soil.nh4.n -3.571e-03 1.426e-03 -2.504 0.0157 \*

soil.ca 7.848e-06 1.816e-06 4.322 7.75e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06015 on 48 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.4848, Adjusted R-squared: 0.4311

F-statistic: 9.032 on 5 and 48 DF, p-value: 4.204e-06

* Only soil.no3.n, soil.nh4.n, and soil.ca come out as significant

I also did a backwards stepwise regression to see what would happen:

I also did a backwards stepwise regression to see what would happen:

Start: AIC=-297.94

leaf.pct.herb ~ pct.urban + leaf.pct.n + soil.no3.n + soil.nh4.n +

soil.ca

Df Sum of Sq RSS AIC

- pct.urban 1 0.002558 0.17623 -299.15

- leaf.pct.n 1 0.005004 0.17868 -298.40

<none> 0.17367 -297.94

- soil.nh4.n 1 0.022691 0.19636 -293.31

- soil.no3.n 1 0.033280 0.20695 -290.47

- soil.ca 1 0.067590 0.24126 -282.19

Step: AIC=-299.15

leaf.pct.herb ~ leaf.pct.n + soil.no3.n + soil.nh4.n + soil.ca

Df Sum of Sq RSS AIC

- leaf.pct.n 1 0.004769 0.18100 -299.71

<none> 0.17623 -299.15

- soil.nh4.n 1 0.025621 0.20185 -293.82

- soil.no3.n 1 0.035469 0.21170 -291.25

- soil.ca 1 0.100517 0.27675 -276.78

Step: AIC=-299.71

leaf.pct.herb ~ soil.no3.n + soil.nh4.n + soil.ca

Df Sum of Sq RSS AIC

<none> 0.18100 -299.71

- soil.nh4.n 1 0.023752 0.20475 -295.05

- soil.no3.n 1 0.041275 0.22227 -290.61

- soil.ca 1 0.119166 0.30016 -274.39

Call:

lm(formula = leaf.pct.herb ~ soil.no3.n + soil.nh4.n + soil.ca)

Coefficients:

(Intercept) soil.no3.n soil.nh4.n soil.ca

-5.313e-02 7.221e-03 -3.595e-03 9.001e-06

* Final model includes soil.no3.n, soil.nh4.n, and soil.ca
* Final model info:

Call:

lm(formula = leaf.pct.herb ~ soil.no3.n + soil.nh4.n + soil.ca)

Residuals:

Min 1Q Median 3Q Max

-0.203746 -0.021830 -0.000913 0.021896 0.214609

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -5.313e-02 2.670e-02 -1.990 0.05210 .

soil.no3.n 7.221e-03 2.138e-03 3.377 0.00143 \*\*

soil.nh4.n -3.595e-03 1.404e-03 -2.562 0.01348 \*

soil.ca 9.001e-06 1.569e-06 5.738 5.58e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06017 on 50 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.463, Adjusted R-squared: 0.4308

F-statistic: 14.37 on 3 and 50 DF, p-value: 7.063e-07

BUT, if I do an anova to see if there’s a significant difference between the full model and the final model, there isn’t:

Analysis of Variance Table

Model 1: leaf.pct.herb ~ pct.urban + leaf.pct.n + soil.no3.n + soil.nh4.n +

soil.ca

Model 2: leaf.pct.herb ~ soil.no3.n + soil.nh4.n + soil.ca

Res.Df RSS Df Sum of Sq F Pr(>F)

1 48 0.17367

2 50 0.18100 -2 -0.0073268 1.0125 0.3709

I attempted an ANCOVA with soil.ca as the covariate. This is what I did:

# run anova to see whether the groups differ in their levels of leaf herbivory without the covariate included

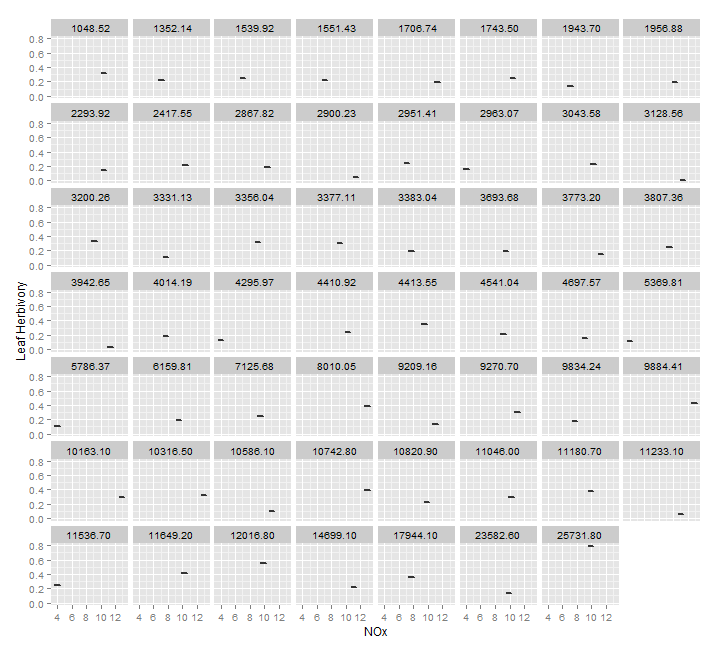
Df Sum Sq Mean Sq F value Pr(>F)

nox.yr.2013 1 0.0425 0.04251 2.514 0.119

Residuals 53 0.8960 0.01691

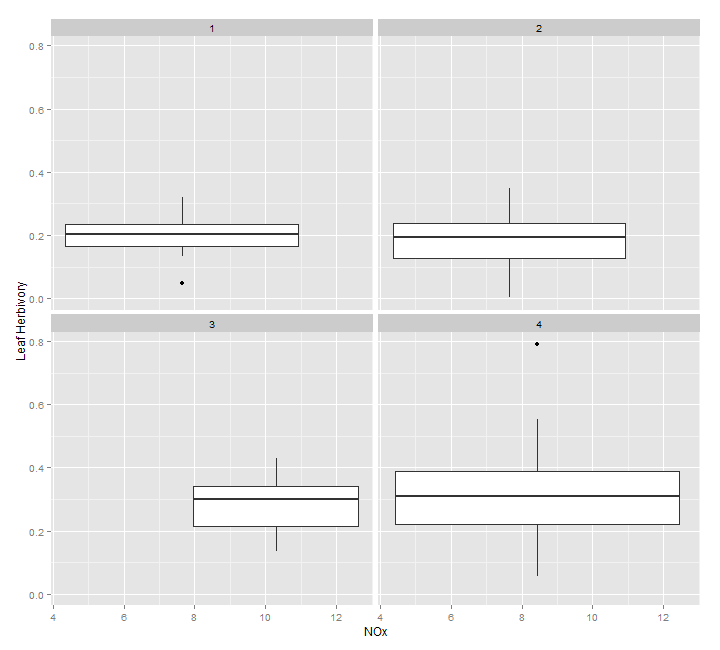
# p value 0.119, no they do not differ

#boxplot - this is insane b/c each value of Ca is getting it's own plot



#convert Ca into factor with 4 groups based on quartiles ?

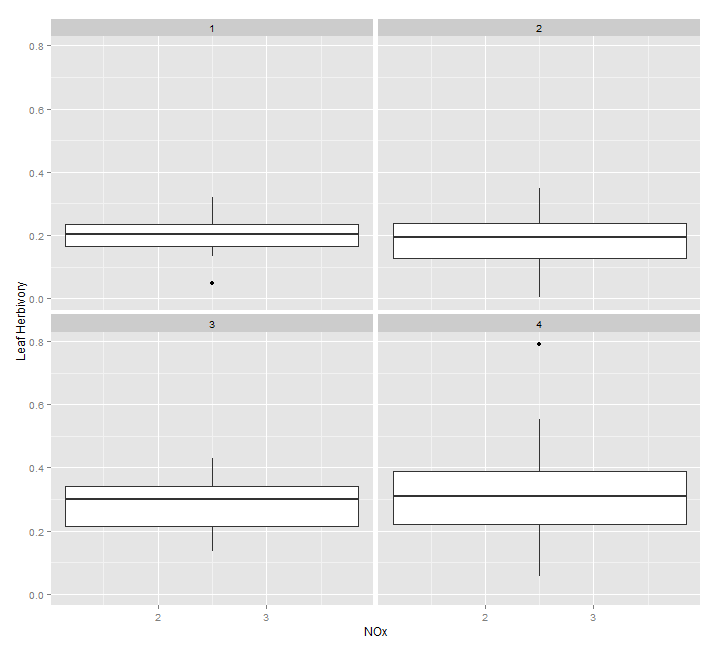
# try boxplot now



#not sure if this is a good plot or not? maybe converting nox into groups is good too?

# convert nox.yr.2013 into factor with 4 groups based on quartiles

# try boxplot again



# doing Levene's test to see whether the variance in leaf herb varies across the interaction of different groups experiencing different Ca levels and the level of NOx

Levene's Test for Homogeneity of Variance (center = median)

Df F value Pr(>F)

group 20 1.5778 0.1178

34

# p value = 0.1178, which means that the assumption of homogeneity of variance is NOT violated

# conduct anova to test whether Ca (covariate in groups) is independent of NOx (indep var)

Df Sum Sq Mean Sq F value Pr(>F)

nox.yr.2013 1 4.95 4.951 4.098 0.048 \*

Residuals 53 64.03 1.208

# p value = 0.048, which means that the soil.ca.gp is significantly different among nox values

# this means it's inappropriate to use soil.ca.gp as a covariate ??

# but if I use the nox.yr.2013.gp in addition to the soil.ca.gp, then it's not significant

Df Sum Sq Mean Sq F value Pr(>F)

nox.yr.2013.gp 1 4.30 4.302 3.525 0.066 .

Residuals 53 64.68 1.220

# p value = 0.066

#trying ancova anyway with soil.ca.gp

Anova Table (Type III tests)

Response: leaf.pct.herb

Sum Sq Df F value Pr(>F)

(Intercept) 0.07562 1 5.0096 0.02951 \*

nox.yr.2013.gp 0.00260 1 0.1722 0.67983

soil.ca.gp 0.13209 1 8.7501 0.00465 \*\*

Residuals 0.78497 52

---

# Sooo.... nox is not important but soil.ca is?

#trying without soil.ca in groups

Anova Table (Type III tests)

Response: leaf.pct.herb

Sum Sq Df F value Pr(>F)

(Intercept) 0.02672 1 1.9737 0.1660064

nox.yr.2013 0.01443 1 1.0659 0.3066561

soil.ca 0.19197 1 14.1796 0.0004245 \*\*\*

Residuals 0.70401 52

# soil.ca even more significant?

I have no idea what to do with this.