

Assignment_6

Kai Ellis

March 6, 2020

1. The data are in: BirdsInForestPatches.csv, and recall from the last assignment that you log10-transformed fragment area, distance to nearest forest patch and distance to largest forest patch. You'll those log10-transformed predictors again here.

```
birdDat <- read.csv("/Users/kelli/Desktop/Biol343/Assignment6/BirdsInForestPatches.csv")
```

```
summary(birdDat)
```

```
##      i..abundance      area_ha      yr_isol      dist_nearest_km
##  Min.       : 1.50    Min.       : 0.10    Min.       :1890    Min.       : 26.0
## 1st Qu.:12.40    1st Qu.: 2.00    1st Qu.:1928    1st Qu.: 93.0
## Median :21.05    Median : 7.50    Median :1962    Median : 234.0
## Mean      :19.51    Mean      : 69.27    Mean      :1950    Mean      : 240.4
## 3rd Qu.:28.30    3rd Qu.: 29.75    3rd Qu.:1966    3rd Qu.: 333.2
## Max.      :39.60    Max.      :1771.00    Max.      :1976    Max.      :1427.0
## dist_large_km      grazing      altitude_masl
##  Min.       : 26.0    Min.       :1.000    Min.       : 60.0
## 1st Qu.: 158.2    1st Qu.:2.000    1st Qu.:120.0
## Median : 338.5    Median :3.000    Median :140.0
## Mean      : 733.3    Mean      :2.982    Mean      :146.2
## 3rd Qu.: 913.8    3rd Qu.:4.000    3rd Qu.:182.5
## Max.      :4426.0    Max.      :5.000    Max.      :260.0
```

```
birdDat <- birdDat %>% mutate(logArea = log10(area_ha),
                             logLarge = log10(dist_large_km),
                             logNear = log10(dist_nearest_km),
                             nyrs_isol = 1985-yr_isol)
```

```
## Warning: The `printer` argument is deprecated as of rlang 0.3.0.
## This warning is displayed once per session.
```

```
#this column had a bugged name, fixed it
names(birdDat)[names(birdDat)=="i..abundance"] <- "abundance"
```

2. Perform backwards selection using the protocol we developed in lecture. Start with the full model and eliminate one predictor at a time chosen as the predictor with the highest P value from an analysis using `car::Anova()`. Larger and smaller models should be compared using likelihood-ratio tests implemented with `anova()`.

```
full <- lm(abundance~ grazing + altitude_masl + logArea + logLarge + logNear + nyrs_isol ,
           data = birdDat)
```

```
Anova(full) #logLarge largest P value at 0.76
```

```
## Anova Table (Type II tests)
```

```
##
```

```
## Response: abundance
```

```
##           Sum Sq Df F value    Pr(>F)
## grazing      131.07  1  3.2163  0.07908 .
## altitude_masl  27.02  1  0.6630  0.41945
## logArea     1059.75  1 26.0049 5.494e-06 ***
## logLarge       3.80  1  0.0933  0.76130
## logNear       4.68  1  0.1149  0.73609
## nyrs_isol     108.83  1  2.6705  0.10864
## Residuals    1996.85 49
```

```
## ---
```

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

```
noLarge <- lm(abundance~ grazing + altitude_masl + logArea + logNear + nyrs_isol ,data =
              birdDat)
```

```
anova(full, noLarge) #no sig p, keep going
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: abundance ~ grazing + altitude_masl + logArea + logLarge + logNear +
##      nyrs_isol
```

```
## Model 2: abundance ~ grazing + altitude_masl + logArea + logNear + nyrs_isol
```

```
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      49 1996.8
## 2      50 2000.7 -1    -3.8027 0.0933 0.7613
```

```
Anova(noLarge) #logLarge largest P value at 0.576
```

```
## Anova Table (Type II tests)
##
## Response: abundance
##           Sum Sq Df F value    Pr(>F)
## grazing      132.44  1   3.3098  0.07486 .
## altitude_masl  35.12  1   0.8778  0.35331
## logArea     1193.04  1  29.8161 1.489e-06 ***
## logNear       12.64  1   0.3159  0.57661
## nyrs_isol     121.64  1   3.0399  0.08739 .
## Residuals    2000.66 50
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
noNear <- lm(abundance~ grazing + altitude_masl + logArea + nyrs_isol ,data =
             birdDat)

anova( noLarge, noNear)#no sig p, keep going
```

```
## Analysis of Variance Table
##
## Model 1: abundance ~ grazing + altitude_masl + logArea + logNear + nyrs_isol
## Model 2: abundance ~ grazing + altitude_masl + logArea + nyrs_isol
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      50 2000.7
## 2      51 2013.3 -1    -12.639 0.3159 0.5766
```

```
Anova(noNear) #altitude_masl largest p at 0.231
```

```
## Anova Table (Type II tests)
##
## Response: abundance
##           Sum Sq Df F value    Pr(>F)
## grazing      123.48  1   3.1280  0.08294 .
## altitude_masl  57.84  1   1.4653  0.23167
## logArea     1227.11  1  31.0846 9.412e-07 ***
## nyrs_isol     134.89  1   3.4169  0.07033 .
## Residuals    2013.29 51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
noAlt <- lm(abundance~ grazing + logArea + nyrs_isol ,data =
            birdDat)

anova(noNear, noAlt)#no sig p, keep going
```

```
## Analysis of Variance Table
##
## Model 1: abundance ~ grazing + altitude_masl + logArea + nyrs_isol
## Model 2: abundance ~ grazing + logArea + nyrs_isol
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      51 2013.3
## 2      52 2071.1 -1    -57.844 1.4653 0.2317
```

```
Anova(noAlt) #nyrs_isol largest p at 0.0768
```

```
## Anova Table (Type II tests)
##
## Response: abundance
##           Sum Sq Df F value    Pr(>F)
## grazing      188.45  1  4.7315  0.03418 *
## logArea     1262.97  1 31.7094 7.316e-07 ***
## nyrs_isol    129.81  1  3.2590  0.07682 .
## Residuals  2071.14 52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
noIsol <- lm(abundance~ grazing + logArea ,data = birdDat)
```

```
anova(noAlt, noIsol) #no sig p, keep going
```

```
## Analysis of Variance Table
##
## Model 1: abundance ~ grazing + logArea + nyrs_isol
## Model 2: abundance ~ grazing + logArea
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      52 2071.1
## 2      53 2200.9 -1    -129.81 3.259 0.07682 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(noIsol) #grazing largest p at 0.0001946, probably gonna stop with noIsol but I'
ll test anyway.
```

```
## Anova Table (Type II tests)
##
## Response: abundance
##           Sum Sq Df F value    Pr(>F)
## grazing      666.0  1  16.038 0.0001946 ***
## logArea     1184.6  1  28.527 1.977e-06 ***
## Residuals  2200.9 53
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
noGraze <- lm(abundance~ logArea ,data =
              birdDat)

anova(noIsol,noGraze) #sig P, noIsol is MAM
```

```
## Analysis of Variance Table
##
## Model 1: abundance ~ grazing + logArea
## Model 2: abundance ~ logArea
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1       53 2200.9
## 2       54 2866.9 -1      -666 16.038 0.0001946 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
rm(noAlt, noGraze, noLarge, noNear) #remove all non-full or MAM models to clean up th
e global enviroment
```

3. Once you've arrived at the MAM, see if you get the same model using forward selection implemented by the add1() function.

```
null <- lm(abundance~1, data=birdDat)
add1(null,scope = ~grazing + altitude_masl + logArea + logLarge + logNear + nyrs_iso
1, test = "F") #logArea has lowest p value, add it to the model
```

```
## Single term additions
##
## Model:
## abundance ~ 1
##
```

| | Df | Sum of Sq | RSS | AIC | F value | Pr(>F) | |
|---------------|----|-----------|--------|--------|---------|-----------|-----|
| <none> | | | 6337.9 | 266.82 | | | |
| grazing | 1 | 2952.3 | 3385.6 | 233.71 | 47.0899 | 6.897e-09 | *** |
| altitude_masl | 1 | 943.5 | 5394.4 | 259.80 | 9.4450 | 0.003316 | ** |
| logArea | 1 | 3471.0 | 2866.9 | 224.40 | 65.3774 | 7.178e-11 | *** |
| logLarge | 1 | 88.4 | 6249.5 | 268.04 | 0.7641 | 0.385905 | |
| logNear | 1 | 101.8 | 6236.1 | 267.92 | 0.8813 | 0.352018 | |
| nyrs_isol | 1 | 1605.8 | 4732.1 | 252.46 | 18.3249 | 7.678e-05 | *** |

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

```
add1(update(null,~.+logArea), scope = ~ grazing + altitude_masl + logArea + logLarge
+ logNear + nyrs_isol, test = "F")
```

```
## Single term additions
##
## Model:
## abundance ~ logArea
##
```

| | Df | Sum of Sq | RSS | AIC | F value | Pr(>F) | |
|---------------|----|-----------|--------|--------|---------|-----------|-----|
| <none> | | | 2866.9 | 224.40 | | | |
| grazing | 1 | 666.00 | 2200.9 | 211.59 | 16.0376 | 0.0001946 | *** |
| altitude_masl | 1 | 227.68 | 2639.3 | 221.76 | 4.5722 | 0.0371233 | * |
| logLarge | 1 | 201.93 | 2665.0 | 222.31 | 4.0159 | 0.0501960 | . |
| logNear | 1 | 65.48 | 2801.5 | 225.10 | 1.2388 | 0.2707342 | |
| nyrs_isol | 1 | 607.35 | 2259.6 | 213.06 | 14.2457 | 0.0004067 | *** |

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

#grazing has lowest p value, add it to the model

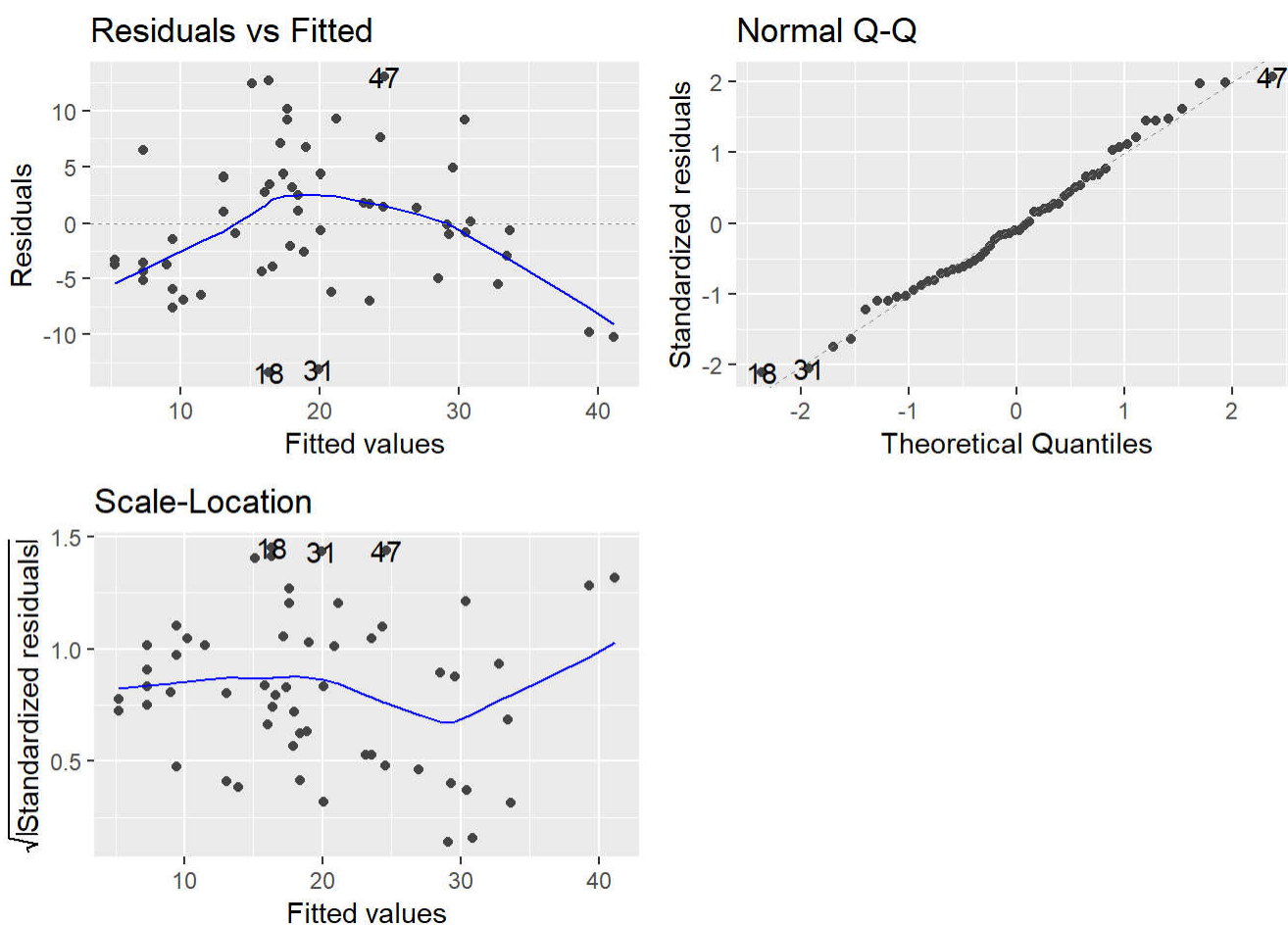
```
add1(update(null,~.+logArea + grazing), scope = ~ grazing + altitude_masl + logArea +
logLarge + logNear + nyrs_isol, test = "F") #no significant p values, stop at logAre
a+grazing
```

```
## Single term additions
##
## Model:
## abundance ~ logArea + grazing
##
##           Df Sum of Sq  RSS   AIC F value  Pr(>F)
## <none>                2200.9 211.59
## altitude_masl    1      52.763 2148.2 212.23  1.2772 0.26361
## logLarge         1      70.324 2130.6 211.77  1.7163 0.19592
## logNear          1      52.376 2148.6 212.24  1.2676 0.26538
## nyrs_isol        1     129.806 2071.1 210.19  3.2590 0.07682 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We arrived at the same MAM with both methods. MAM is abundance~area+grazing.

- Once you have the MAM, check assumptions using a quick 1-function graphical analysis. No figure caption required here.

```
autoplot(noIsol,c(1,2,3))
```



None of the assumptions appear to be violated as there are no consistent trends away from the expected values in the autoplots.

- Determine the relative strength of the predictors in the model by rerunning it to calculate standardized

partial regression coefficients.

```
summary(smam <- lm(scale(abundance) ~ scale(grazing)+scale(logArea), data = birdDat))
```

```
##
## Call:
## lm(formula = scale(abundance) ~ scale(grazing) + scale(logArea),
##     data = birdDat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2510 -0.4023 -0.0589  0.3845  1.2179
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -5.348e-17  8.022e-02   0.000 1.000000
## scale(grazing) -3.910e-01  9.763e-02  -4.005 0.000195 ***
## scale(logArea)  5.214e-01  9.763e-02   5.341 1.98e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6003 on 53 degrees of freedom
## Multiple R-squared:  0.6527, Adjusted R-squared:  0.6396
## F-statistic: 49.81 on 2 and 53 DF,  p-value: 6.723e-13
```

6. Make a formal graph that shows how well the data fit the MAM.

```
birdDat <- birdDat %>% mutate(pred = fitted(noIsol))

ggplot(birdDat)+
  geom_point(aes(x=pred, y = abundance)) +
  geom_abline(intercept = 0, slope = 1, colour = "blue") +
  theme_minimal() +
  scale_y_continuous (name ="Abundance") +
  scale_x_continuous (name ="Predicted")
```

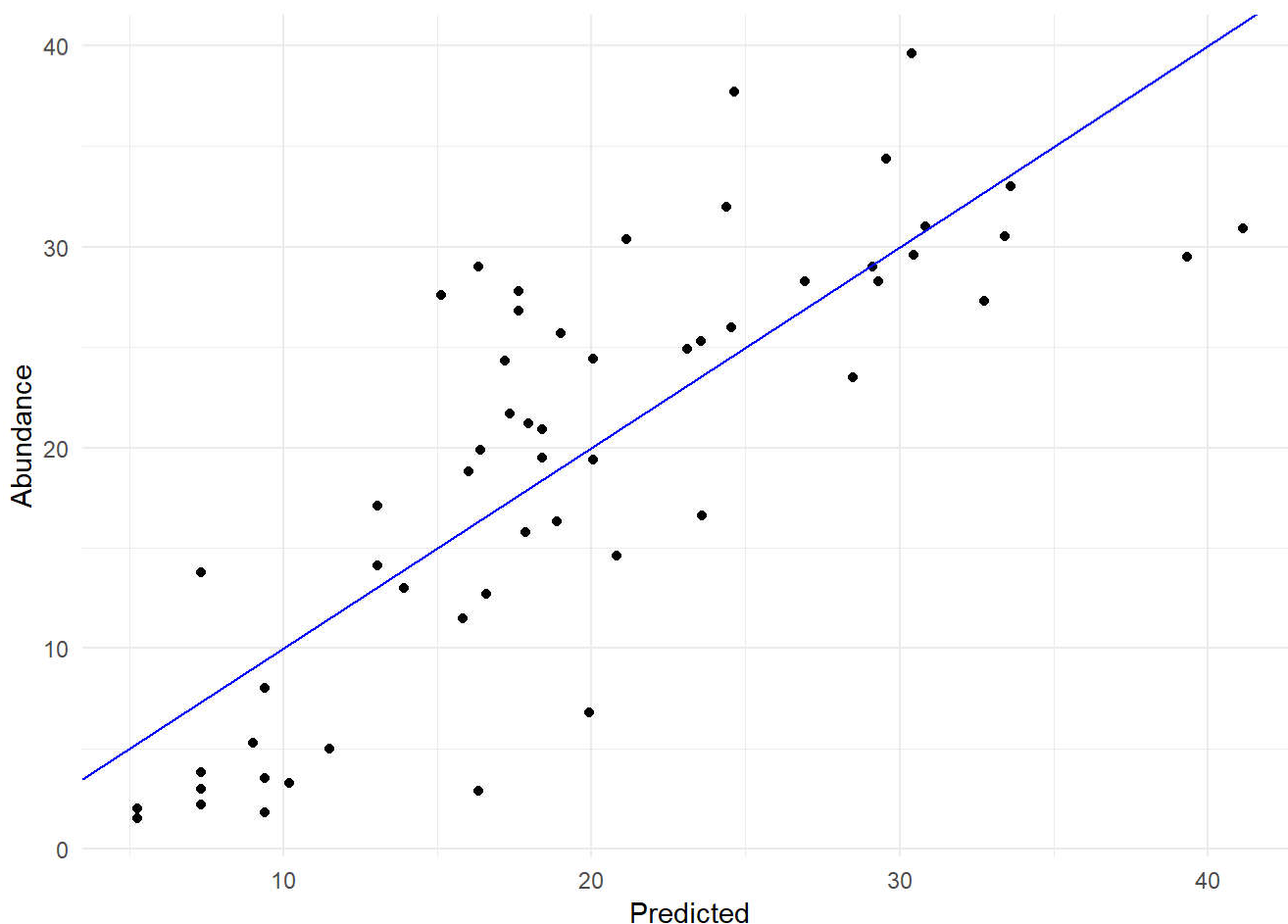



Figure 1. Abundance values vs abundance values predicted by the MAM of $\text{Abundance} \sim \log\text{Area} + \text{grazing}$. Blue line is of formula $y=1x+0$, and represents a perfect match between predicted and actual abundance values.

7. Like backwards elimination, we begin model selection using AICc with a full model. Use the dredge() function to run all possible models and, for each, compute the log-likelihood, corrected AIC, DAIC, Akaike weight and r^2 . Express the regression parameters as standardized partial regression coefficients.

```
options(na.action = "na.fail")
dd <- MuMIn::dredge(full, extra = "R^2", beta = "sd")
```

```
## Fixed term is "(Intercept)"
```

```
print(dd, abbrev.names = F)
```

```
## Global model call: lm(formula = abundance ~ grazing + altitude_masl + logArea +
##      logLarge + logNear + nyrs_isol, data = birdDat)
## ---
## Model selection table
##      (Intercept) altitude_masl grazing logArea logLarge  logNear nyrs_isol
## 39              0             -0.2606  0.5423              -0.1867
## 40              0      0.10480 -0.2194  0.5355              -0.1904
## 7               0             -0.3910  0.5214
## 47              0             -0.2524  0.5866 -0.08816         -0.1654
## 55              0             -0.2656  0.5666      -0.078940    -0.1753
## 15              0             -0.3603  0.5835 -0.11730
## 38              0      0.14690      0.6168              -0.2974
## 8               0      0.10010 -0.3541  0.5145
## 23              0             -0.3873  0.5523      -0.095420
## 37              0             0.6503              -0.3223
## 56              0      0.08697 -0.2296  0.5521      -0.050240    -0.1825
## 48              0      0.08439 -0.2225  0.5636 -0.05341         -0.1768
## 16              0      0.06549 -0.3428  0.5655 -0.09187
## 45              0             0.6973 -0.10140         -0.2929
## 63              0             -0.2582  0.5870 -0.05925 -0.048440    -0.1654
## 31              0             -0.3660  0.5838 -0.08828 -0.048590
## 24              0      0.07502 -0.3606  0.5393      -0.071250
## 46              0      0.13000      0.6417 -0.04542         -0.2871
## 54              0      0.13840      0.6277      -0.026840    -0.2959
## 53              0             0.6739      -0.070750    -0.3144
## 64              0      0.07910 -0.2285  0.5654 -0.03472 -0.034960    -0.1760
## 61              0             0.6989 -0.08534 -0.027160    -0.2945
## 32              0      0.05977 -0.3489  0.5674 -0.07117 -0.038420
## 62              0      0.12890      0.6427 -0.04046 -0.009173    -0.2877
## 6               0      0.19710      0.6858
## 13              0             0.8139 -0.19320
## 14              0      0.14420      0.7498 -0.12910
## 22              0      0.18210      0.7039      -0.046080
## 5               0             0.7400
## 29              0             0.8143 -0.19090 -0.004067
## 30              0      0.14590      0.7477 -0.13730  0.015690
## 21              0             0.7723      -0.106600
## 3               0             -0.6825
## 4               0      0.12940 -0.6298
## 12              0      0.18010 -0.6042      0.14690
## 11              0             -0.6793      0.09504
## 35              0             -0.6083              -0.1167
## 44              0      0.19920 -0.4792      0.18580         -0.1824
## 19              0             -0.6782      0.029980
## 36              0      0.13320 -0.5504              -0.1226
## 43              0             -0.5801      0.12330         -0.1545
## 20              0      0.15380 -0.6094      0.073480
## 28              0      0.17810 -0.6068      0.15500 -0.014440
## 27              0             -0.6847      0.12170 -0.044420
## 51              0             -0.5969      0.044030    -0.1249
```

| | | | | | | | |
|----|----|---|---------|---------|---------|-----------|---------|
| ## | 52 | 0 | 0.16450 | -0.5128 | | 0.092370 | -0.1410 |
| ## | 60 | 0 | 0.19770 | -0.4813 | 0.19190 | -0.010950 | -0.1822 |
| ## | 59 | 0 | | -0.5856 | 0.14980 | -0.044260 | -0.1545 |
| ## | 42 | 0 | 0.35690 | | 0.29120 | | -0.4672 |
| ## | 58 | 0 | 0.36160 | | 0.25460 | 0.061200 | -0.4614 |
| ## | 50 | 0 | 0.33120 | | | 0.207700 | -0.4304 |
| ## | 34 | 0 | 0.28410 | | | | -0.4372 |
| ## | 41 | 0 | | | 0.20450 | | -0.5363 |
| ## | 33 | 0 | | | | | -0.5034 |
| ## | 49 | 0 | | | | 0.136600 | -0.5060 |
| ## | 57 | 0 | | | 0.19110 | 0.021790 | -0.5346 |
| ## | 10 | 0 | 0.45220 | | 0.24200 | | |
| ## | 18 | 0 | 0.43440 | | | 0.221900 | |
| ## | 2 | 0 | 0.38580 | | | | |
| ## | 26 | 0 | 0.45940 | | 0.16800 | 0.125900 | |
| ## | 1 | 0 | | | | | |
| ## | 17 | 0 | | | | 0.126700 | |
| ## | 9 | 0 | | | 0.11810 | | |
| ## | 25 | 0 | | | 0.06548 | 0.087180 | |

| ## | | R^2 | df | logLik | AICc | delta | weight |
|----|----|---------|----|----------|-------|-------|--------|
| ## | 39 | 0.67320 | 5 | -180.555 | 372.3 | 0.00 | 0.147 |
| ## | 40 | 0.68230 | 6 | -179.761 | 373.2 | 0.93 | 0.092 |
| ## | 7 | 0.65270 | 4 | -182.257 | 373.3 | 0.99 | 0.090 |
| ## | 47 | 0.67920 | 6 | -180.036 | 373.8 | 1.48 | 0.070 |
| ## | 55 | 0.67880 | 6 | -180.072 | 373.9 | 1.55 | 0.068 |
| ## | 15 | 0.66380 | 5 | -181.347 | 373.9 | 1.59 | 0.066 |
| ## | 38 | 0.66290 | 5 | -181.428 | 374.1 | 1.75 | 0.061 |
| ## | 8 | 0.66110 | 5 | -181.577 | 374.4 | 2.05 | 0.053 |
| ## | 23 | 0.66100 | 5 | -181.582 | 374.4 | 2.06 | 0.053 |
| ## | 37 | 0.64350 | 4 | -182.993 | 374.8 | 2.46 | 0.043 |
| ## | 56 | 0.68430 | 7 | -179.585 | 375.5 | 3.19 | 0.030 |
| ## | 48 | 0.68420 | 7 | -179.597 | 375.5 | 3.22 | 0.029 |
| ## | 16 | 0.66690 | 6 | -181.093 | 375.9 | 3.59 | 0.024 |
| ## | 45 | 0.65150 | 5 | -182.360 | 375.9 | 3.61 | 0.024 |
| ## | 63 | 0.68070 | 7 | -179.908 | 376.1 | 3.84 | 0.022 |
| ## | 31 | 0.66530 | 6 | -181.225 | 376.2 | 3.86 | 0.021 |
| ## | 24 | 0.66510 | 6 | -181.238 | 376.2 | 3.88 | 0.021 |
| ## | 46 | 0.66420 | 6 | -181.316 | 376.3 | 4.04 | 0.020 |
| ## | 54 | 0.66340 | 6 | -181.380 | 376.5 | 4.16 | 0.018 |
| ## | 53 | 0.64800 | 5 | -182.638 | 376.5 | 4.17 | 0.018 |
| ## | 64 | 0.68490 | 8 | -179.532 | 378.1 | 5.82 | 0.008 |
| ## | 61 | 0.65190 | 6 | -182.322 | 378.4 | 6.05 | 0.007 |
| ## | 32 | 0.66780 | 7 | -181.018 | 378.4 | 6.06 | 0.007 |
| ## | 62 | 0.66430 | 7 | -181.312 | 379.0 | 6.65 | 0.005 |
| ## | 6 | 0.58360 | 4 | -187.342 | 383.5 | 11.16 | 0.001 |
| ## | 13 | 0.57950 | 4 | -187.614 | 384.0 | 11.70 | 0.000 |
| ## | 14 | 0.59520 | 5 | -186.548 | 384.3 | 11.99 | 0.000 |
| ## | 22 | 0.58530 | 5 | -187.226 | 385.7 | 13.34 | 0.000 |
| ## | 5 | 0.54770 | 3 | -189.659 | 385.8 | 13.47 | 0.000 |
| ## | 29 | 0.57950 | 5 | -187.613 | 386.4 | 14.12 | 0.000 |
| ## | 30 | 0.59540 | 6 | -186.537 | 386.8 | 14.48 | 0.000 |

```
## 21 0.55800 4 -189.012 386.8 14.50 0.000
## 3 0.46580 3 -194.315 395.1 22.78 0.000
## 4 0.47980 4 -193.573 395.9 23.62 0.000
## 12 0.49920 5 -192.507 396.2 23.91 0.000
## 11 0.47480 4 -193.838 396.5 24.15 0.000
## 35 0.47390 4 -193.886 396.6 24.25 0.000
## 44 0.51770 6 -191.457 396.6 24.32 0.000
## 19 0.46670 4 -194.268 397.3 25.01 0.000
## 36 0.48870 5 -193.087 397.4 25.07 0.000
## 43 0.48830 5 -193.112 397.4 25.11 0.000
## 20 0.48460 5 -193.314 397.8 25.52 0.000
## 28 0.49930 6 -192.500 398.7 26.41 0.000
## 27 0.47610 5 -193.772 398.7 26.44 0.000
## 51 0.47580 5 -193.786 398.8 26.46 0.000
## 52 0.49610 6 -192.681 399.1 26.77 0.000
## 60 0.51770 7 -191.452 399.2 26.93 0.000
## 59 0.48950 6 -193.046 399.8 27.50 0.000
## 42 0.40730 5 -197.226 405.7 33.34 0.000
## 58 0.40960 6 -197.115 407.9 35.64 0.000
## 50 0.37070 5 -198.902 409.0 36.70 0.000
## 34 0.32970 4 -200.670 410.1 37.82 0.000
## 41 0.29410 4 -202.119 413.0 40.71 0.000
## 33 0.25340 3 -203.690 413.8 41.53 0.000
## 49 0.27200 4 -202.981 414.7 42.44 0.000
## 57 0.29440 5 -202.107 415.4 43.10 0.000
## 10 0.20310 4 -205.516 419.8 47.51 0.000
## 18 0.19570 4 -205.772 420.3 48.02 0.000
## 2 0.14890 3 -207.358 421.2 48.87 0.000
## 26 0.21310 5 -205.162 421.5 49.21 0.000
## 1 0.00000 2 -211.871 428.0 55.66 0.000
## 17 0.01606 3 -211.418 429.3 56.99 0.000
## 9 0.01395 3 -211.478 429.4 57.11 0.000
## 25 0.01878 4 -211.340 431.5 59.16 0.000
## Models ranked by AICc(x)
```

8. What predictors are included in the “top model” and does this set of predictors differ from the MAM that you arrived at using backwards and forwards selection?

The top model includes Grazing, logArea and nyrs_Isol, a different set of predictors from our forwards and backwards selection MAM.

9. A common convention is to consider all models with $DAIC \leq 2$ to be statistically indistinguishable. Let's consider how useful this criterion is in this case. Use ggplot to make a scatterplot with the rank of each model in terms of AICc on the x-axis (ranked from lowest AICc to highest) and DAIC on the y-axis. Plot a red horizontal line at $DAIC = 2$ to indicate how many other models you should consider along with the top model. This graph requires a figure caption, etc. In the text of your R notebook, interpret this graph. Do you see any obvious breakpoints in DAIC that separate models into likely vs. unlikely groups? Do any of these break points correspond to $DAIC \leq 2$?

```

ddr <- data.frame(model = seq(1:64), AICc = dd$AICc, delta = dd$delta,
                  weight = dd$weight, evid.ratio = max(dd$weight)/dd$weight)

ggplot(ddr)+
  geom_point(aes(x = model, y = delta))+
  geom_hline(yintercept = 2, colour = "red")+
  theme_minimal() +
  scale_y_continuous (name = "ΔAICc") +
  scale_x_continuous (name = "Model")

```

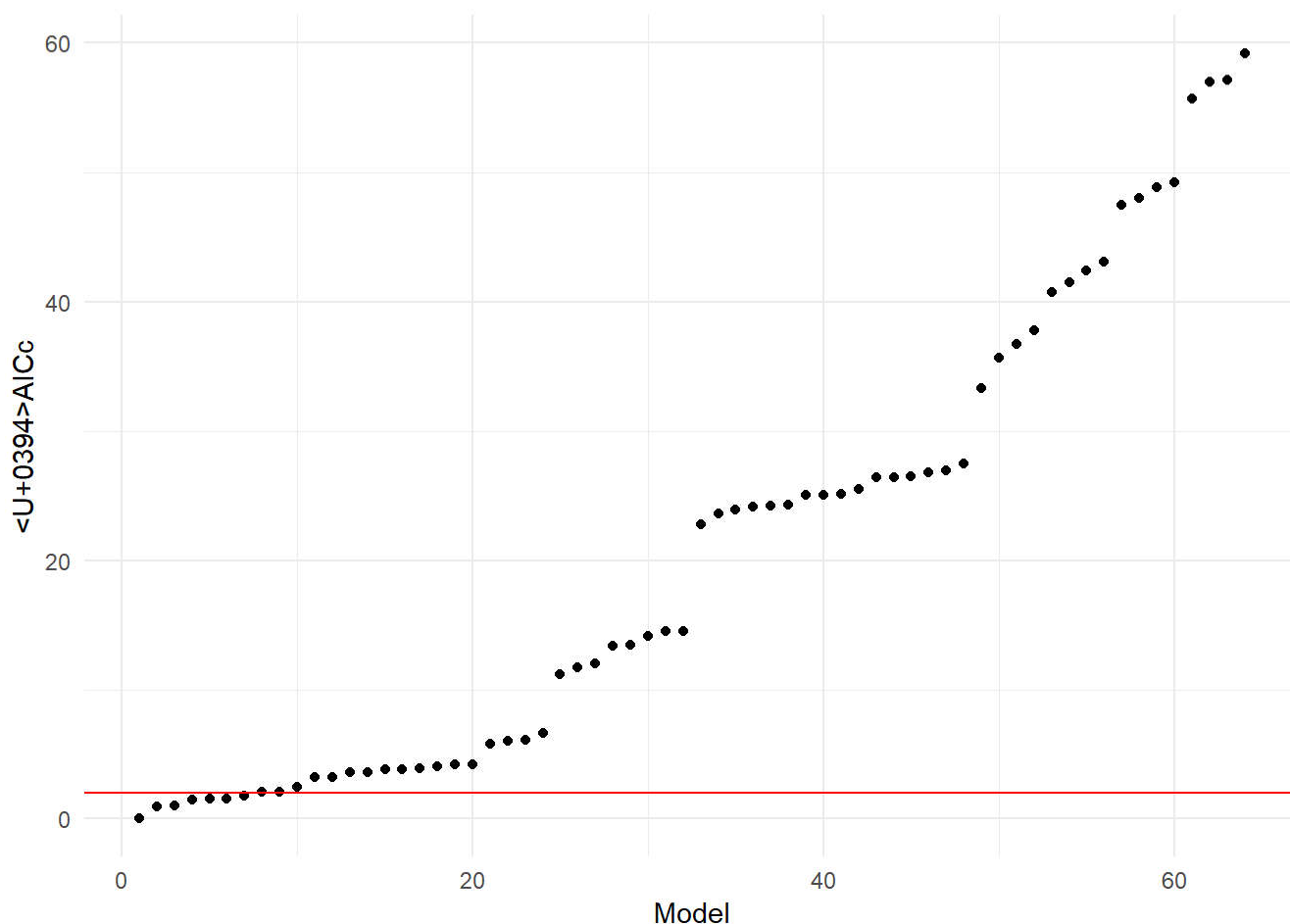


Figure 2. All possible linear models from the Birds in forest patches data set ranked from lowest (left) DAICc to highest (right), where the lowest is the most likely model to predict abundance. The red line is set to a DAICc threshold of 2, with all points below the line having explanatory power not significantly different from that of the “best” model.

This graph shows that while only the first 7 models are statistically identical, having DAICc values below 2, the first 20-24 are all relatively well grouped with no obvious rapid increases in AICc to separate them. The first debatably clear break in AICc is after model 20, while there is a definite clear break after model 24. This suggests that although we are only taking the models below the arbitrary value of 2 that is commonly accepted in stats, the first 20-24 are also worth keeping in mind

10. Given the $DAIC \leq 2$ criterion, how many other potential models should you consider along with the top model? Among this set of models, how frequently is each predictor included? Do the standardized partial regression coefficients of included predictors change much from model to model? Briefly describe the

relevant trends.

6 other models should be considered along with our MAM. Logarea appears in all 7 of these models, grazing in 6, nyrs_isol in 5, logLarge and altitude_masl appear in 2 and finally logNear appears in a very lonely 1 model. Standardized partial regression coefficients do not change much for most predictors between models. Altitude ranges between 0.1 and 0.14, area between 0.52 and 0.61, large between -0.088 and -0.117. However both nyrs isol and grazing have relatively large ranges between -0.1654 and 0.2974, and between -0.21 and -0.39 respectively.

General trends for standardized partial regression coefficients are as follows: grazing seems to drop in explanatory power as more predictors are added to the model dropping from -0.39 in the grazing+logarea model down to -0.21 and -0.25 in both 4 predictor models. none of the other predictors seem to follow clear trends. nyrs seems to spike in explanatory power when altitude is added, and log area appears to on average increase in power when more predictors are added, but neither of these are clear cut.

11. Calculate evidence ratios for all potential models and interpret these values along with Akaike weights.
How likely is it that the top model is the best model?

```
(evid.ratio = max(dd$weight)/dd$weight)
```

```
## model weights
## [1] 1.000 1.590 1.639
## [4] 2.092 2.171 2.210
## [7] 2.396 2.781 2.795
## [10] 3.423 4.939 5.000
## [13] 6.022 6.079 6.823
## [16] 6.873 6.961 7.529
## [19] 8.024 8.032 18.342
## [22] 20.592 20.693 27.772
## [25] 264.899 347.658 400.703
## [28] 789.419 841.259 1162.548
## [31] 1393.829 1407.237 88498.179
## [34] 134632.449 155243.005 175469.431
## [37] 184085.377 190833.131 269933.699
## [40] 277276.004 284203.191 347664.423
## [43] 541894.737 550028.416 557844.962
## [46] 649101.506 703949.033 934737.143
## [49] 17391156.612 54734374.381 92975840.096
## [52] 162789218.246 692828638.857 1043973977.916
## [55] 1641726781.938 2291067421.053 20708202330.527
## [58] 26749354050.361 40889611688.644 48623579929.361
## [61] 1220030127616.871 2370595599663.339 2516806586802.062
## [64] 7006988638513.698
```

the top value has an Akaike weight of 0.147, suggesting a 14.7% chance of being the best model. In terms of evidence ratios, the top 7 models have evidence ratios of 1, 1.59, 1.64, 2.09, 2.17, 2.21, and 2.4. As the evidence ratio is the weight of the top model/the weight of the model you are looking at this means the top model is 1.59x, 1.64x, 2.09x, etc. more likely to be the best model than the other models in the top 7.

To summarize: the top model has an ~15% chance of being the best model and is ~1.6x more likely to be the best model than the 2nd best model.

12. Run this top model and use ggplot to create “added-variable” plots that show the partial relation between the response and each predictor. If there is more than one predictor, arrange all the plots together using `plot_grid()`. This is a formal graph requiring a figure caption, etc.

```
selBirdDat <- birdDat %>% select(abundance, logArea, nyrs_isol, grazing)
#y axis
A_y.g <- residuals(lm(abundance ~ nyrs_isol + grazing, data = selBirdDat))
A_a.g <- residuals(lm(abundance ~ logArea + grazing, data = selBirdDat))
A_a.y <- residuals(lm(abundance ~ logArea + nyrs_isol, data = selBirdDat))

#x axis residuals
a_y.g <- residuals(lm(logArea ~ nyrs_isol + grazing, data = selBirdDat))
y_a.g <- residuals(lm(nyrs_isol ~ logArea + grazing, data = selBirdDat))
g_a.y <- residuals(lm(grazing ~ logArea + nyrs_isol, data = selBirdDat))

datf <- data.frame(A_y.g,
                   A_a.g,
                   A_a.y,
                   a_y.g,
                   y_a.g,
                   g_a.y )

rm(A_y.g, A_a.g, A_a.y, a_y.g, y_a.g, g_a.y )

area <- ggplot(datf, aes(x=A_y.g, y=a_y.g)) +
  geom_point() +
  geom_smooth(method= "lm", se=T) +
  theme_minimal() +
  scale_y_continuous (name ="Abundance | Others") +
  scale_x_continuous (name = expression(paste("Fragment area (ha)", " (", log[10], ") | Others" )))

years <- ggplot(datf, aes(x=A_a.g, y=y_a.g)) +
  geom_point() +
  geom_smooth(method= "lm", se=T) +
  theme_minimal() +
  scale_y_continuous (name = "") +
  scale_x_continuous (name ="Years isolated | Others")

grazing <- ggplot(datf, aes(x=A_a.y, y=g_a.y)) +
  geom_point() +
  geom_smooth(method= "lm", se=T) +
  theme_minimal() +
  scale_y_continuous (name ="Abundance | Others") +
  scale_x_continuous (name ="Grazing | Others")

ggarrange(area, years, grazing, labels = c("A", "B", "C"))
```

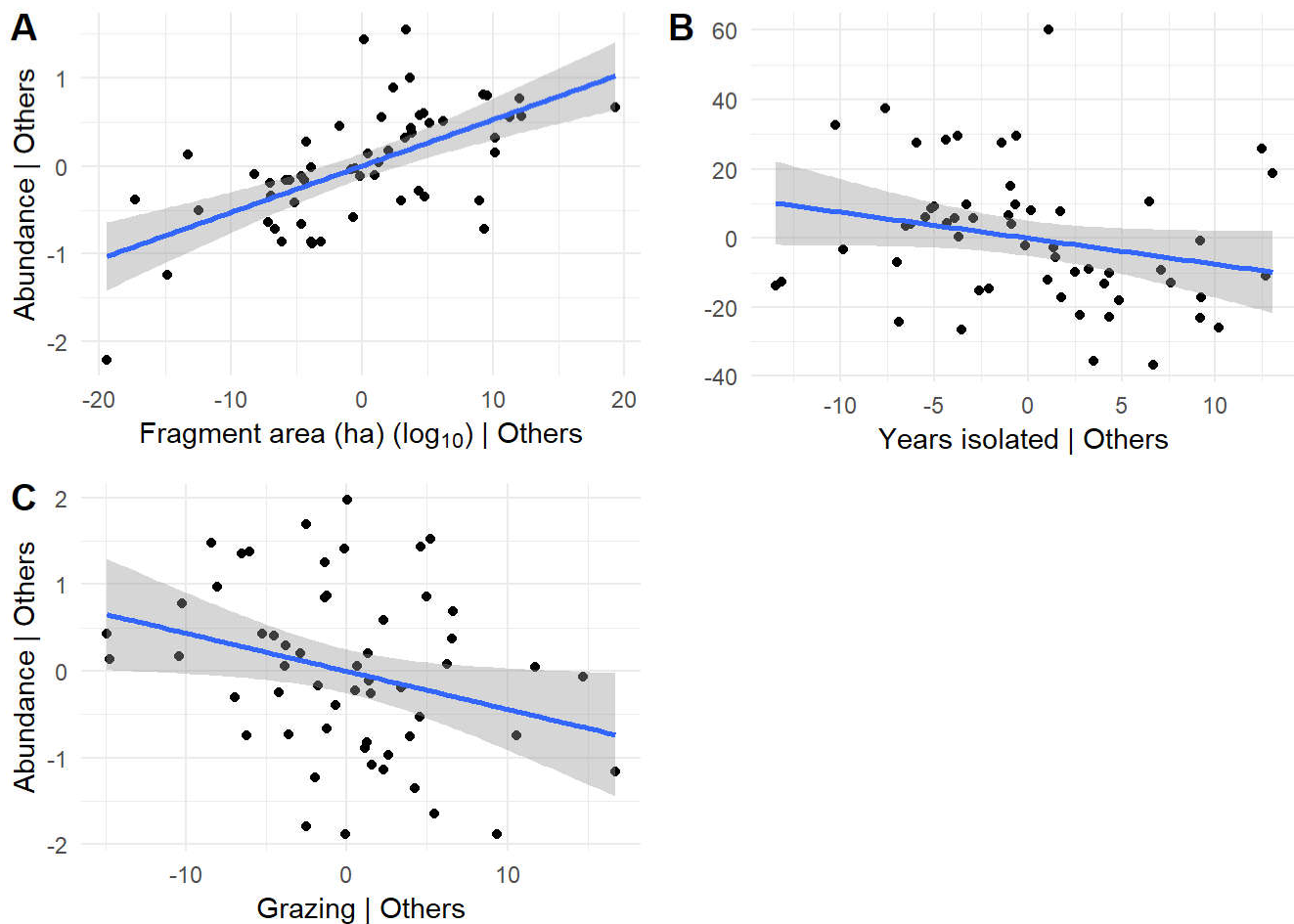


Figure 3. Added variable plots for all significant predictive variables. The three plots show the relationship between the response variable, abundance, and the predictive variables woodland fragment area in hectares (A), years since the fragment was isolated (B), and intensity of grazing in the fragment (C). Variables not being tested were controlled for.

13. Make sure that all the formal graphs asked for above are beautifully rendered with complete and informative figure legends. You should upload your file to the Assignment #6 OnQ dropbox by Saturday 7 March 2020 at 1159pm. As before, please submit a PDF version of your .html R notebook document called "StudentNumber_A6.pdf", where the file name starts with your student number.