Homework7

Question 1

Import data

```
library('car')
```

```
## Loading required package: carData
```

```
library("ggplot2")
library(MASS)
library(MuMIn)


ds = readxl::read_excel('/Users/kanoalindiwe/Downloads/Projects/playground/R/Quantitative Ecology/Datas

# Response variable
ABUND <- ds$ABUND

# Predictor variables
AREA <- ds$AREA
YR.ISOL <- ds$YR.ISOL
DIST <- ds$DIST
LDIST <- ds$LDIST
GRAZE <- ds$GRAZE
ALT <- ds$ALT
```

Research question

What are the most important factors that influence bird abundance in forest patches?

Statistical method

mutli linear regression

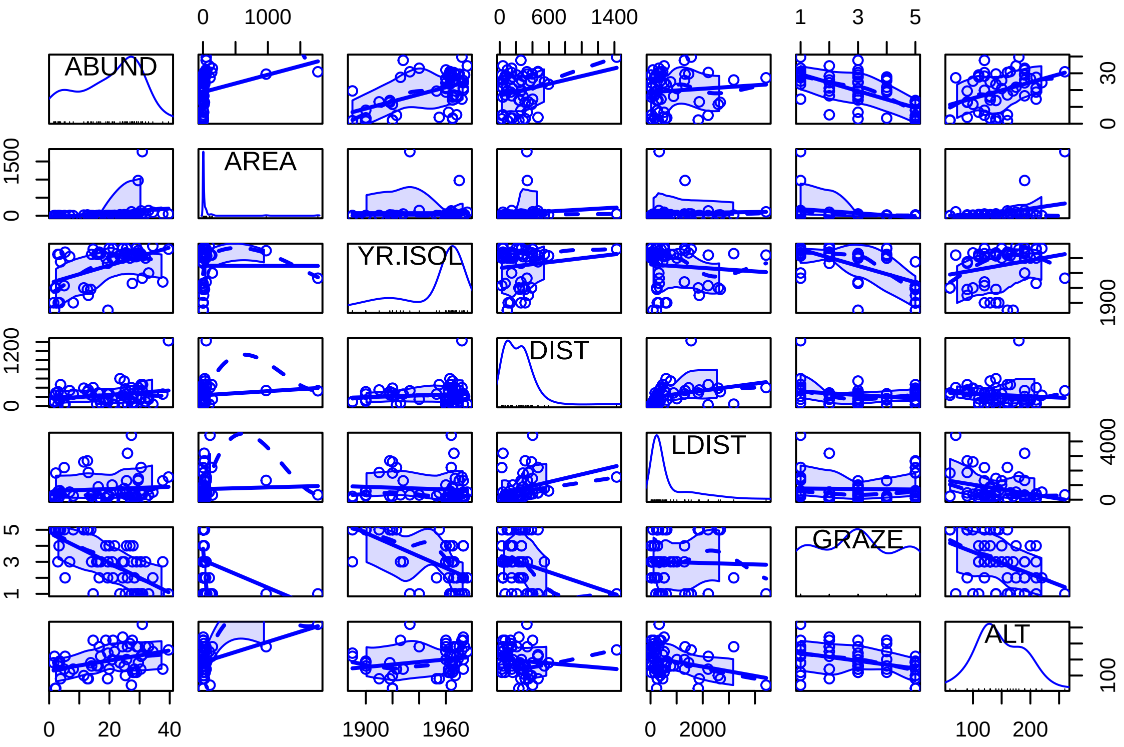
Hypothesis statements

H0: There is no significant relationship between abundance of forest birds and any of these variables: patch area, number of years of isolation, distance to the nearest patch and larger patch, grazing intensity and altitude.

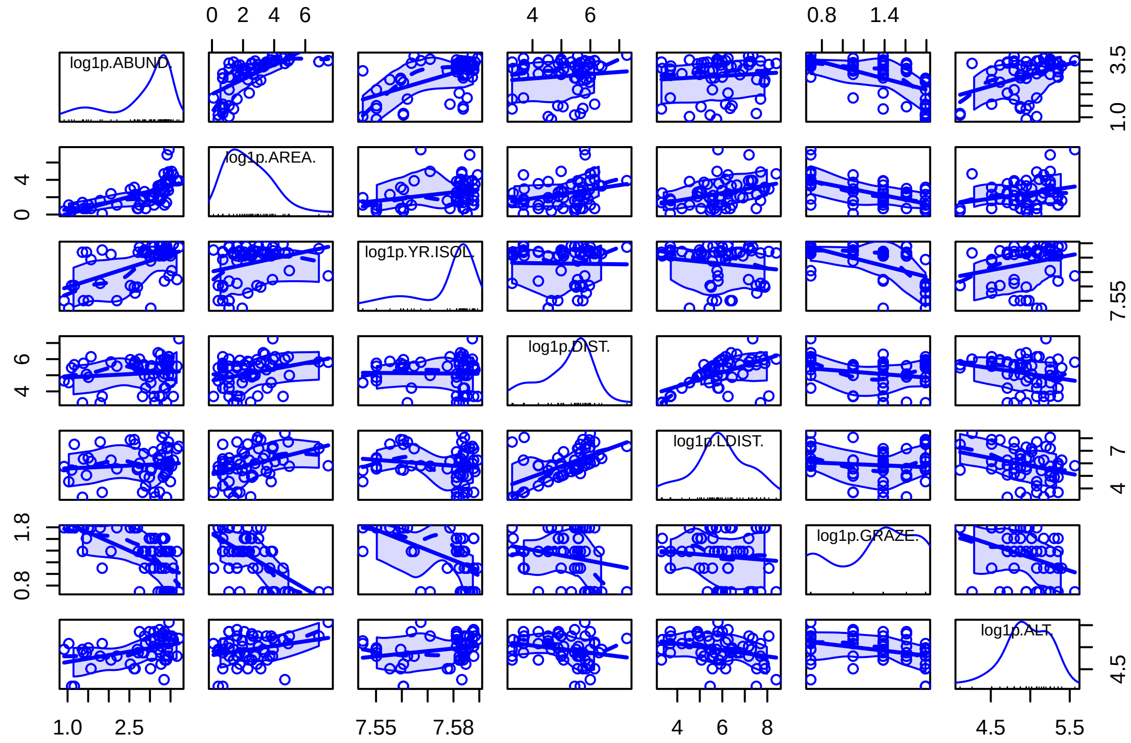
H1: At least one of these variables has a significant relationship with forest bird abundance.

Assumptions

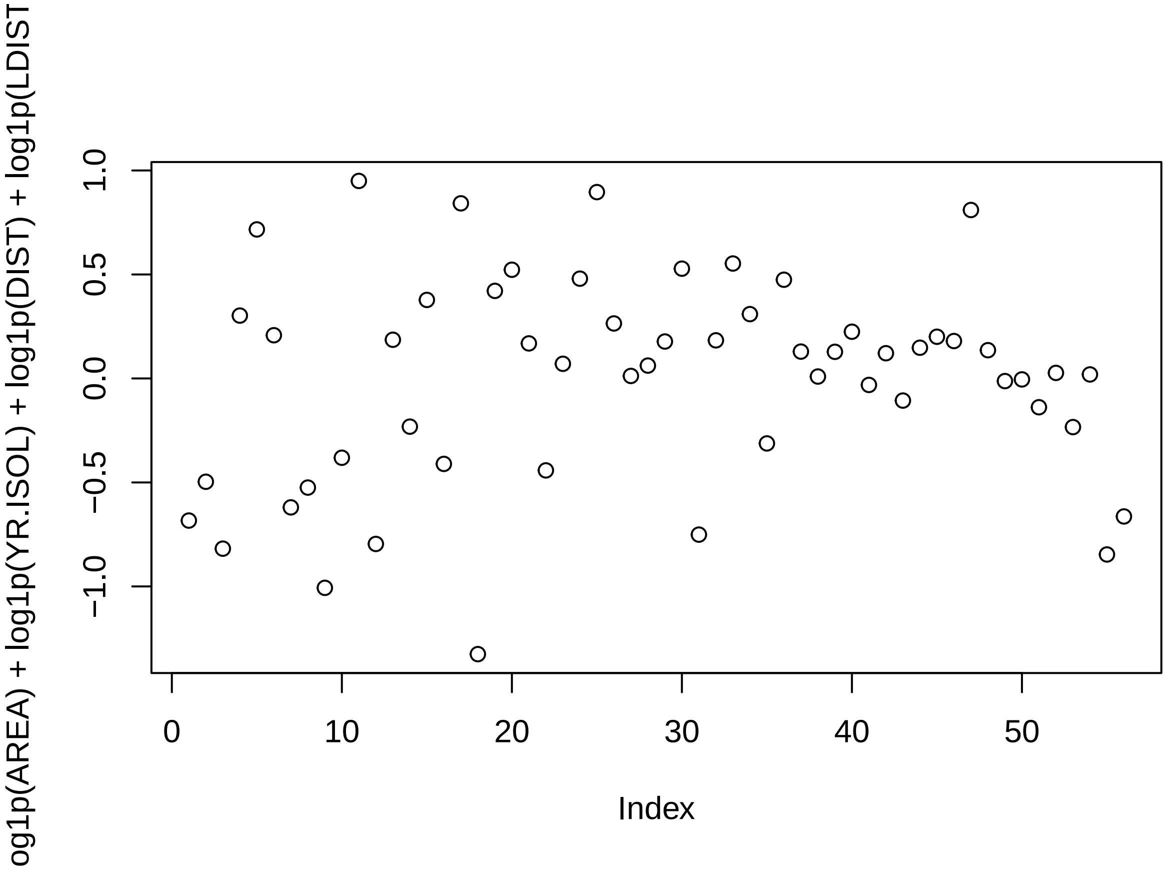
```
# Linearity - The relationship between predictors and the response variable is linear.
scatterplotMatrix(~ ABUND+AREA+YR.ISOL+DIST+LDIST+GRAZE+ALT)
```



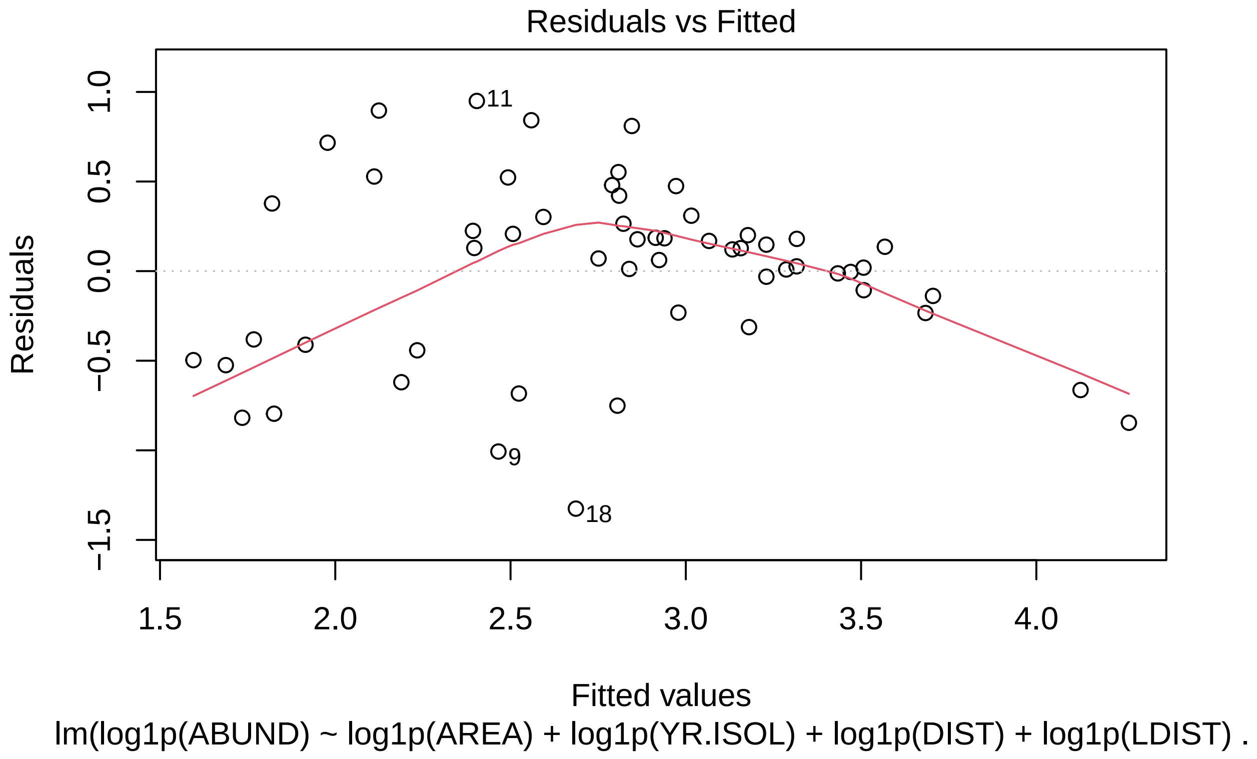
```
# Test if log will improve assumption
scatterplotMatrix(~ log1p(ABUND)+log1p(AREA)+log1p(YR.ISOL)+log1p(DIST)+log1p(LDIST)+log1p(GRAZE)+log1p
```



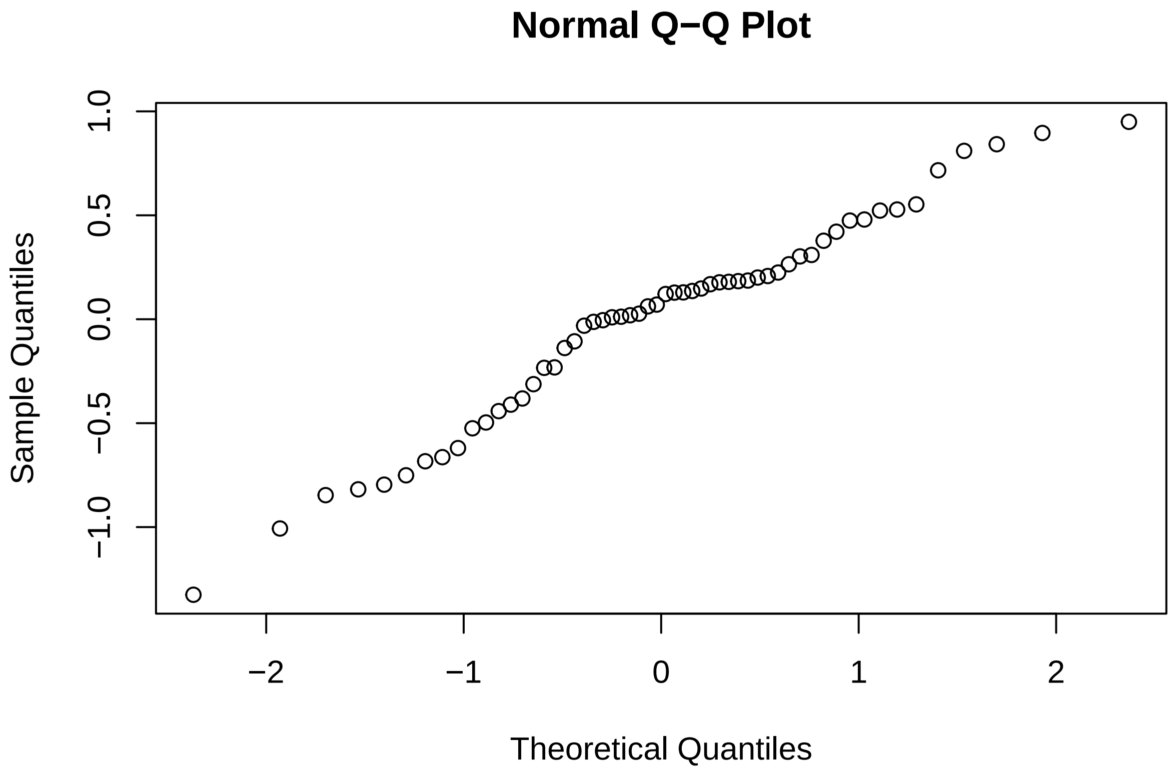
```
# Pass, logging variables improve normality and therefor linear relationship. Use a multiplicative mode

# Independence - Observations (and residuals) are independent of each other.
plot(residuals(lm(log1p(ABUND) ~ log1p(AREA)+log1p(YR.ISOL)+log1p(DIST)+log1p(LDIST)+log1p(GRAZE)+log1p
```



```r
# Pass, residuals look good

# Homoscedasticity - The variance of residuals is constant across all fitted values.
plot((lm(log1p(ABUND) ~ log1p(AREA)+log1p(YR.ISOL)+log1p(DIST)+log1p(LDIST)+log1p(GRAZE)+log1p(ALT))),
```



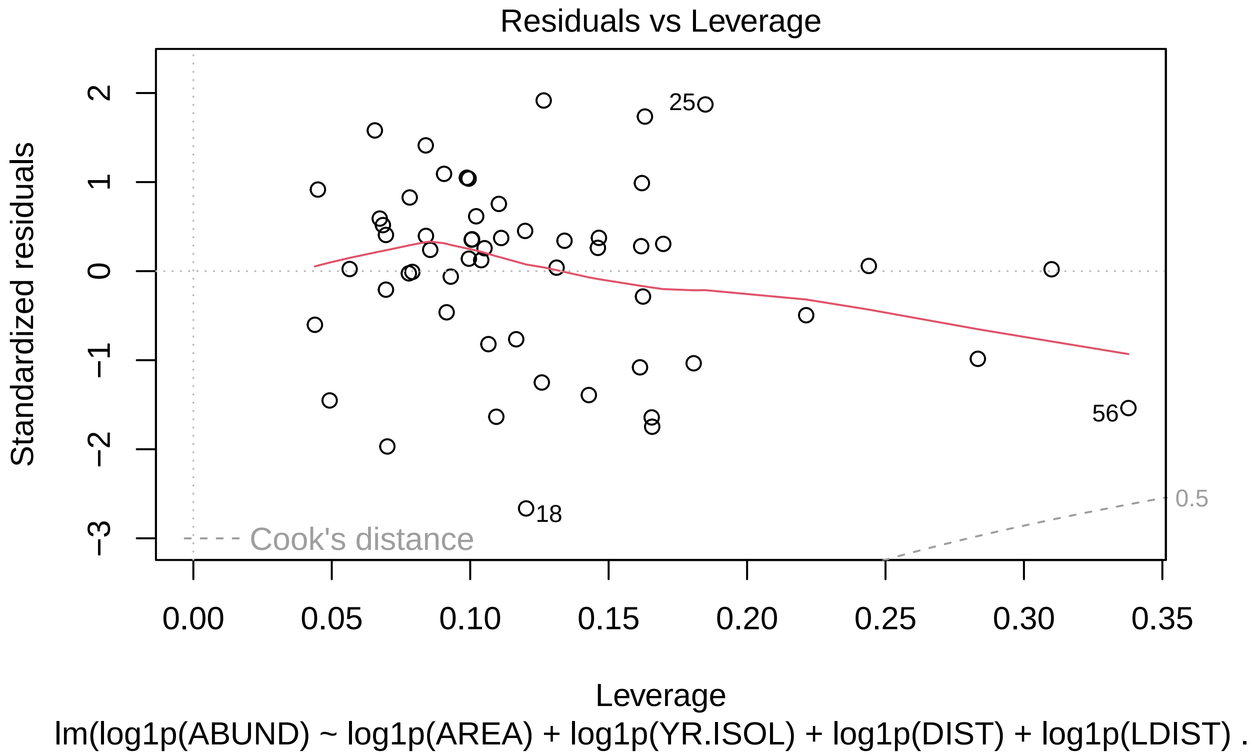
```
# Pass, it looks okay

# Normality of Errors - Residuals are approximately normally distributed.
qqnorm(residuals(lm(log1p(ABUND) ~ log1p(AREA)+log1p(YR.ISOL)+log1p(DIST)+log1p(LDIST)+log1p(GRAZE)+log
```



```r
# Pass, they look normal

# No Multicollinearity - Predictor variables are not highly correlated with each other.
cor(cbind(log1p(ABUND),log1p(AREA),log1p(YR.ISOL),log1p(DIST),log1p(LDIST),log1p(GRAZE),log1p(ALT)))
```

```
##             [,1]       [,2]        [,3]        [,4]        [,5]       [,6]
## [1,]  1.00000000  0.6511503  0.54825275  0.11317392  0.09348866 -0.6150857
## [2,]  0.65115032  1.0000000  0.27530073  0.30362431  0.37280755 -0.6118485
## [3,]  0.54825275  0.2753007  1.00000000 -0.01824663 -0.15993943 -0.5688038
## [4,]  0.11317392  0.3036243 -0.01824663  1.00000000  0.60266234 -0.2197997
## [5,]  0.09348866  0.3728076 -0.15993943  0.60266234  1.00000000 -0.1063894
## [6,] -0.61508573 -0.6118485 -0.56880380 -0.21979973 -0.10638938  1.0000000
## [7,]  0.39186560  0.2609036  0.24790126 -0.24628612 -0.29805676 -0.3541861
##            [,7]
## [1,]  0.3918656
## [2,]  0.2609036
## [3,]  0.2479013
## [4,] -0.2462861
## [5,] -0.2980568
## [6,] -0.3541861
## [7,]  1.0000000
```

```r
# Pass, all below 80%

# No Significant Outliers or Influential Points - Extreme values do not unduly affect the model.
plot((lm(log1p(ABUND) ~ log1p(AREA)+log1p(YR.ISOL)+log1p(DIST)+log1p(LDIST)+log1p(GRAZE)+log1p(ALT))), 
```



```
# Pass, does not overlap cooks distance
```

Check reduced models

```
# All variables
model1 <- lm(log1p(ABUND) ~ log1p(AREA)+log1p(YR.ISOL)+log1p(DIST)+log1p(LDIST)+log1p(GRAZE)+log1p(ALT))

# Add interaction between area nd years sense grazing
modal2 <- lm(log1p(ABUND) ~ log1p(AREA)+log1p(YR.ISOL)+log1p(DIST)+log1p(LDIST)+log1p(GRAZE)+log1p(ALT)+

# Only included limited predictors
model3 <- lm(log1p(ABUND) ~ log1p(AREA)+log1p(YR.ISOL)+log1p(GRAZE))

AIC(model1, modal2, model3)
```

```
##        df      AIC
## model1  8 96.39492
## modal2  9 96.47208
## model3  5 93.37612
```

```
# Model 3, with limited predictors, performs the best.

# Check multiple models using stepAIC
stepAIC(model1)
```

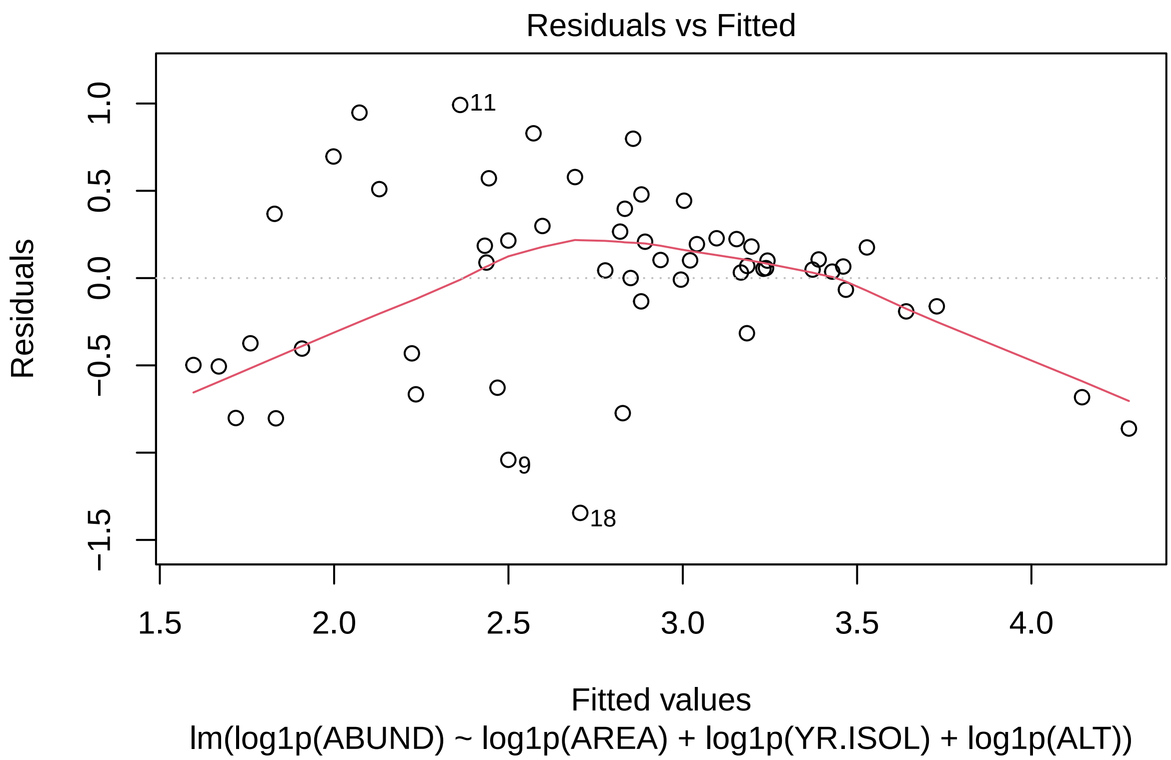
```
## Start:  AIC=-64.53
## log1p(ABUND) ~ log1p(AREA) + log1p(YR.ISOL) + log1p(DIST) + log1p(LDIST) +
##     log1p(GRAZE) + log1p(ALT)
##
##                  Df Sum of Sq    RSS     AIC
## - log1p(DIST)     1    0.0034 13.782 -66.512
## - log1p(LDIST)    1    0.0093 13.788 -66.488
## - log1p(GRAZE)    1    0.1067 13.885 -66.094
## <none>                        13.778 -64.526
## - log1p(ALT)      1    0.6259 14.404 -64.038
## - log1p(YR.ISOL)  1    2.4389 16.217 -57.399
## - log1p(AREA)     1    3.6356 17.414 -53.413
##
## Step:  AIC=-66.51
## log1p(ABUND) ~ log1p(AREA) + log1p(YR.ISOL) + log1p(LDIST) +
##     log1p(GRAZE) + log1p(ALT)
##
##                  Df Sum of Sq    RSS     AIC
## - log1p(LDIST)    1    0.0061 13.788 -68.488
## - log1p(GRAZE)    1    0.1032 13.885 -68.094
## <none>                        13.782 -66.512
## - log1p(ALT)      1    0.6672 14.449 -65.865
## - log1p(YR.ISOL)  1    2.4470 16.229 -59.360
## - log1p(AREA)     1    3.6322 17.414 -55.412
##
## Step:  AIC=-68.49
## log1p(ABUND) ~ log1p(AREA) + log1p(YR.ISOL) + log1p(GRAZE) +
##     log1p(ALT)
##
##                  Df Sum of Sq    RSS     AIC
## - log1p(GRAZE)    1    0.1073 13.895 -70.053
## <none>                        13.788 -68.488
## - log1p(ALT)      1    0.7439 14.532 -67.545
## - log1p(YR.ISOL)  1    2.5362 16.324 -61.032
## - log1p(AREA)     1    4.6194 18.407 -54.306
##
## Step:  AIC=-70.05
## log1p(ABUND) ~ log1p(AREA) + log1p(YR.ISOL) + log1p(ALT)
##
##                  Df Sum of Sq    RSS     AIC
## <none>                        13.895 -70.053
## - log1p(ALT)      1    0.8828 14.778 -68.604
## - log1p(YR.ISOL)  1    4.1387 18.034 -57.454
## - log1p(AREA)     1    7.8096 21.705 -47.078
```

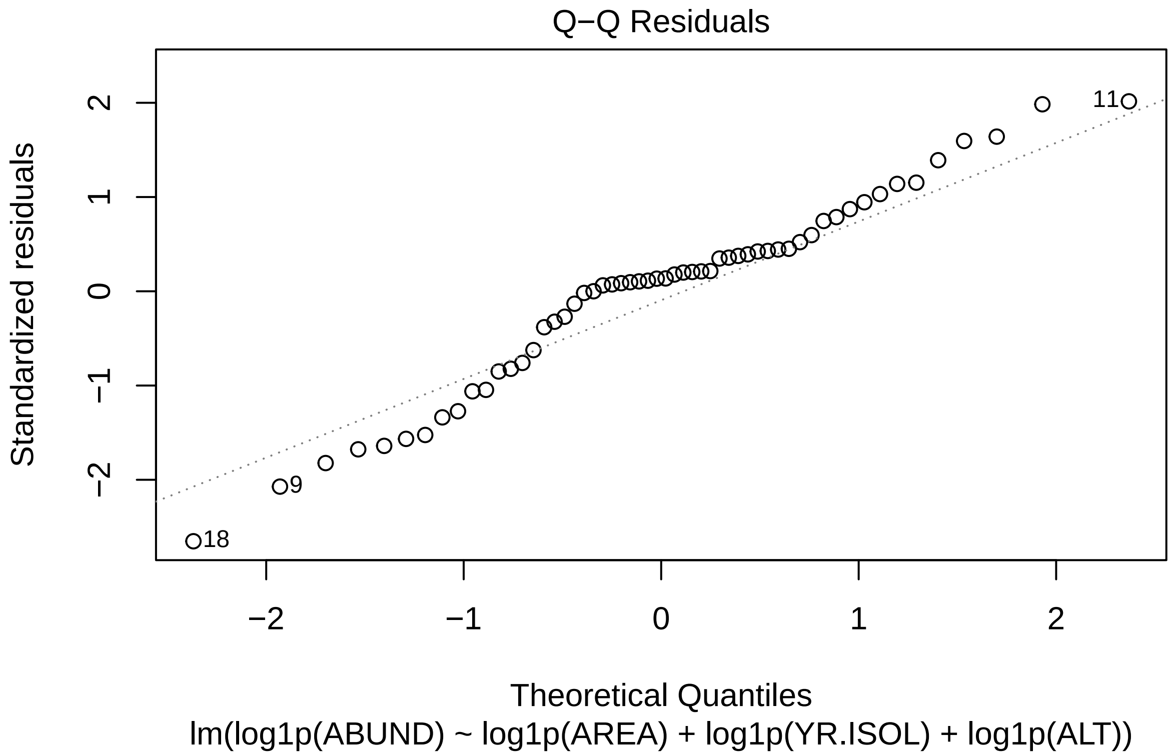
```
##
## Call:
## lm(formula = log1p(ABUND) ~ log1p(AREA) + log1p(YR.ISOL) + log1p(ALT))
##
## Coefficients:
##    (Intercept)     log1p(AREA)  log1p(YR.ISOL)      log1p(ALT)
##      -166.6040          0.2530         22.0051          0.4199
```

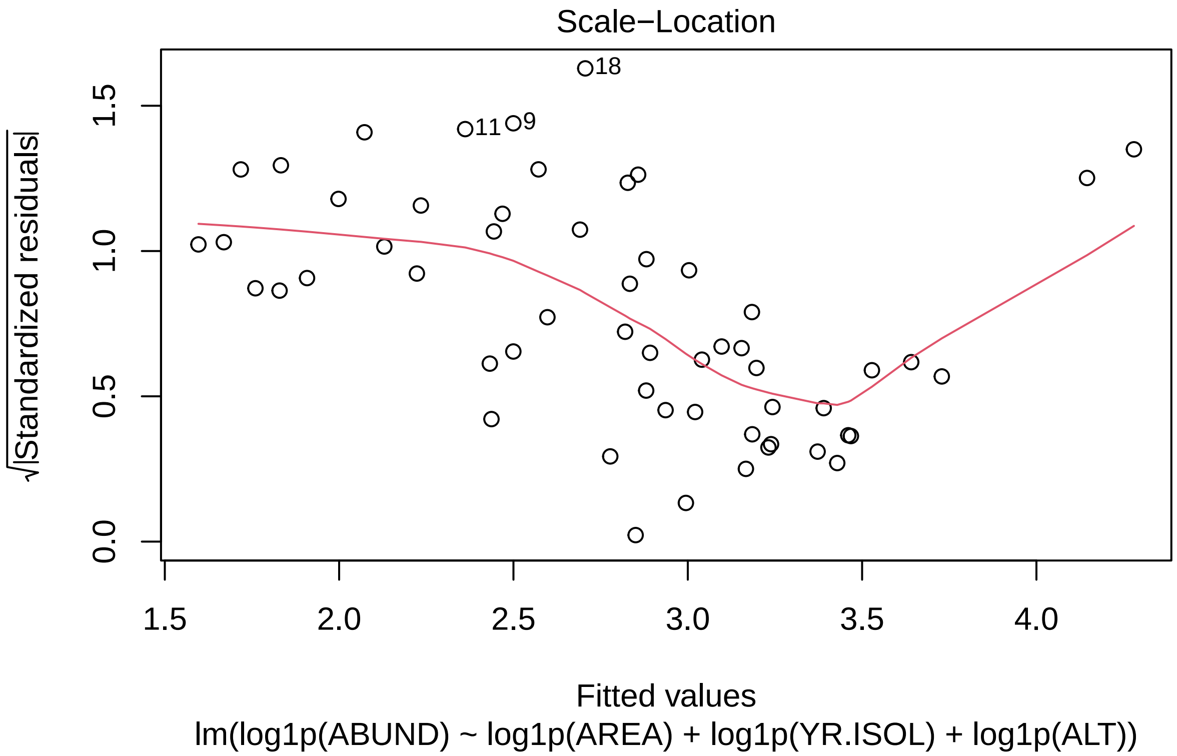
```
# Best model found as: log1p(ABUND) ~ log1p(AREA) + log1p(YR.ISOL) + log1p(ALT)
```

Test

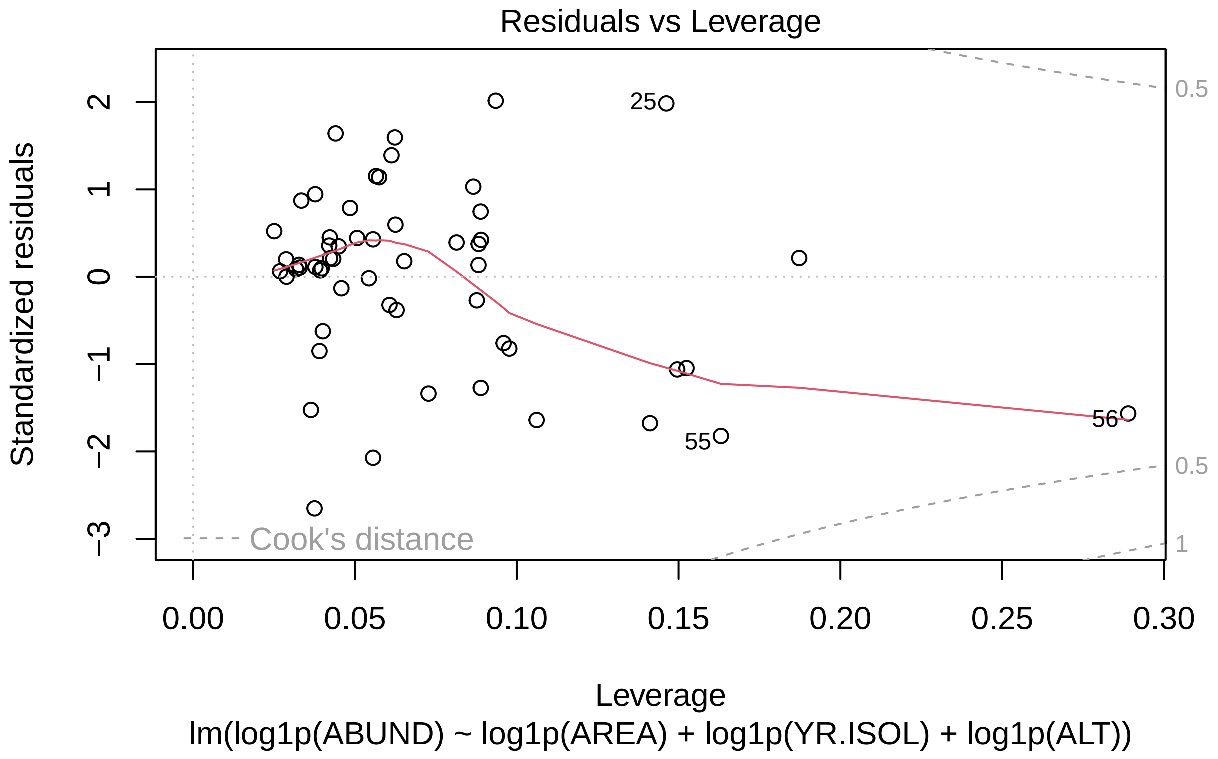
```
# Best model
model_best <- lm(log1p(ABUND) ~ log1p(AREA) + log1p(YR.ISOL) + log1p(ALT))
plot(model_best)
```




Q−Q Residuals
Standardized residuals
2
1
0
−1
−2
−2
−1
0
1
2
18
9
11
Theoretical Quantiles
lm(log1p(ABUND) ~ log1p(AREA) + log1p(YR.ISOL) + log1p(ALT))



Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
18
11
9
1.5
2.0
2.5
3.0
3.5
4.0
Fitted values
lm(log1p(ABUND) ~ log1p(AREA) + log1p(YR.ISOL) + log1p(ALT))



```r
summary(model_best)
```

```
##
## Call:
## lm(formula = log1p(ABUND) ~ log1p(AREA) + log1p(YR.ISOL) + log1p(ALT))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.34499 -0.33042  0.06774  0.23755  0.99187
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -166.6040    42.1321  -3.954 0.000233 ***
## log1p(AREA)       0.2530     0.0468   5.406 1.64e-06 ***
## log1p(YR.ISOL)   22.0051     5.5914   3.935 0.000248 ***
## log1p(ALT)        0.4199     0.2310   1.818 0.074883 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5169 on 52 degrees of freedom
## Multiple R-squared:  0.5969, Adjusted R-squared:  0.5737
## F-statistic: 25.67 on 3 and 52 DF,  p-value: 2.509e-10
```

Results

StepAIC was used to find the best variables to predict bird abundance among these variables: patch area, number of years of isolation, distance to the nearest patch and larger patch, grazing intensity and altitude. The best model was found to be log(ABUND) = -166.6040 + 0.2530 × log(AREA) + 22.0051 × log(YR.ISOL) + 0.4199 × log(ALT). It accounted for 57% of variability of bird abundance (R^2 = 0.5969, adjusted R^2 = 0.5737, F(3, 52) = 25.67, p = 2.509e-10). Area and years sense isolation were significant and altitude was not: log1p(AREA) = 0.2530 ± 0.0468, t(52) = 5.406, p = 1.64e-06; log1p(YR.ISOL) = 22.0051 ± 5.5914, t(52) = 3.935, p = 0.000248; log1p(ALT) = 0.4199 ± 0.2310, t(52) = 1.818, p = 0.074883. We reject our null hypothesis and accept that these three variables significantly predict bird abundance.

Question 2

Import data

```r
library('car')
library("ggplot2")
library(MASS)

ds = readxl::read_excel('/Users/kanoalindiwe/Downloads/Projects/playground/R/Quantitative Ecology/Datas

# Response variable
growth <- ds$`growth rate`

# Predictor variables

plot <- ds$plot
species <- ds$species
dbh2 <- as.numeric(ds$dbh2)
comstat <- factor(ds$comstat)
```

Research question

Does tree size OR competitive status OR their interaction predict growth rate?

Statistical method

mutli linear regression

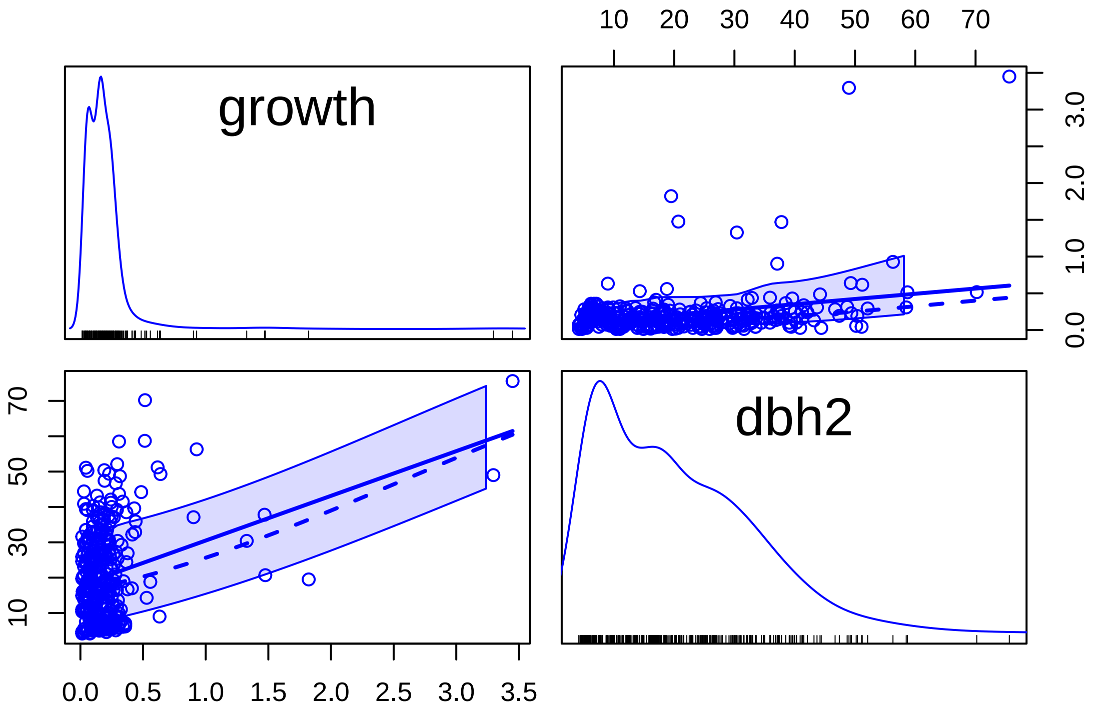
Hypothesis statements

H0: There is a significant relationship between growth rates and tree size, competitive status, or their interaction.

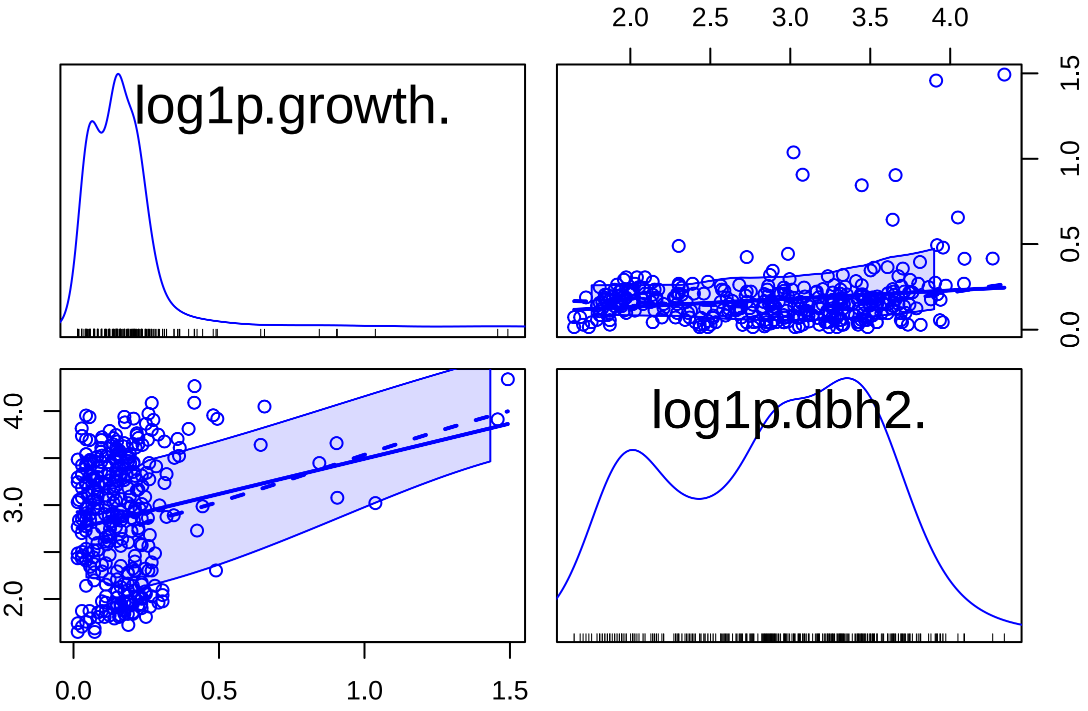
H1: There is not a significant relationship between growth rates and tree size, competitive status, or their interaction.

Assumptions

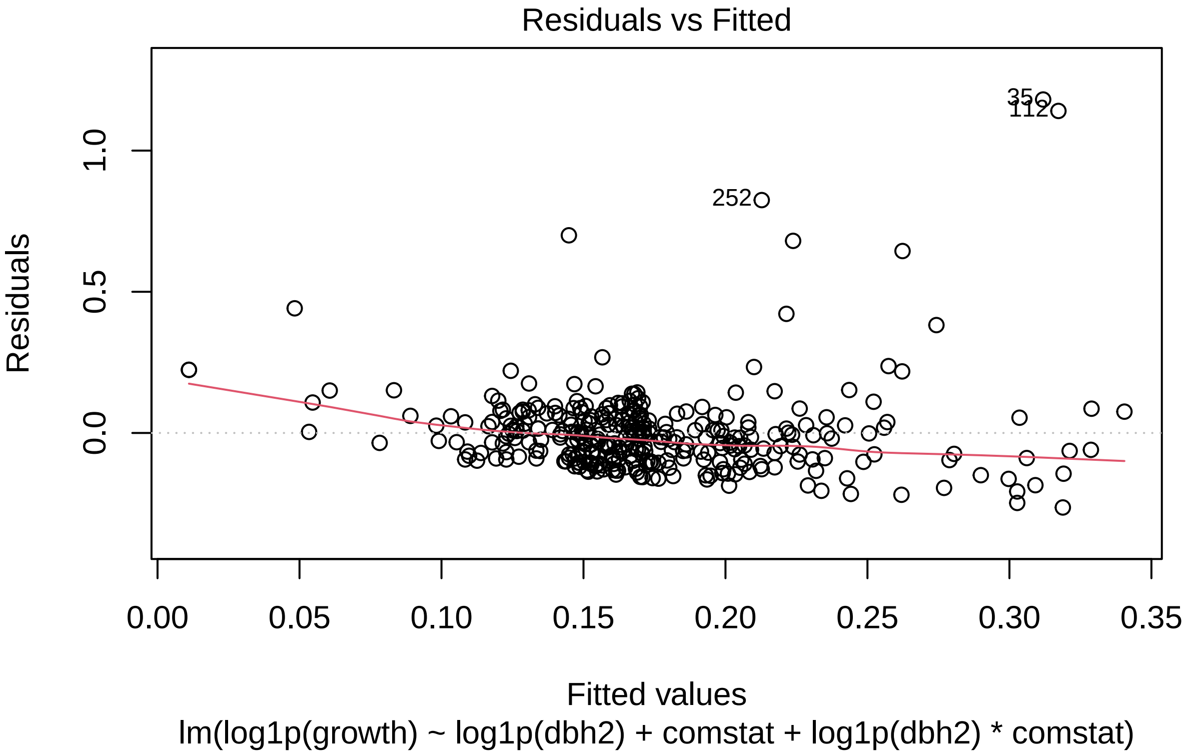
```r
# Test models
model1 <- lm(growth ~ dbh2 + comstat)
model2 <- lm(log1p(growth) ~ log1p(dbh2) + comstat + log1p(dbh2)*comstat)
model3 <- lm(scale(growth) ~ scale(dbh2) + comstat)
model4 <- lm(sqrt(growth) ~ sqrt(dbh2) + comstat)
growth.cube <- growth^(1/3)
dbh2.cube <- dbh2^(1/3)
model5 <- lm(growth.cube ~ dbh2.cube + comstat)
# plot(model1)
# plot(model2)
# plot(model3)
# plot(model4)
# plot(model5)

# Linearity - The relationship between predictors and the response variable is linear.
scatterplotMatrix(~ growth+dbh2)
```



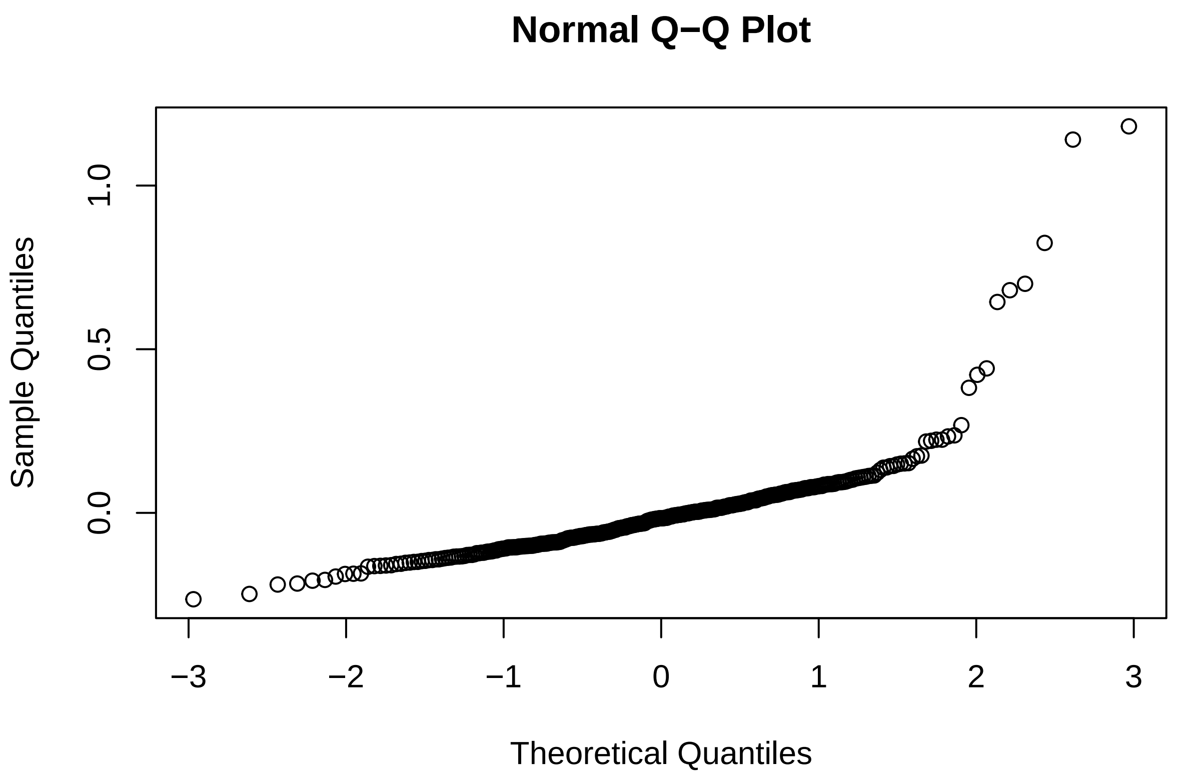
```
# Test if log will improve assumption
scatterplotMatrix(~ log1p(growth)+log1p(dbh2))
```



```
# Pass, logging variables improve normality and therefor linear relationship. Use a multiplicative mode

# Independence - Observations (and residuals) are independent of each other.
# Assumed to be true based on study design

# Homoscedasticity - The variance of residuals is constant across all fitted values.
plot(model2, which = 1)
```



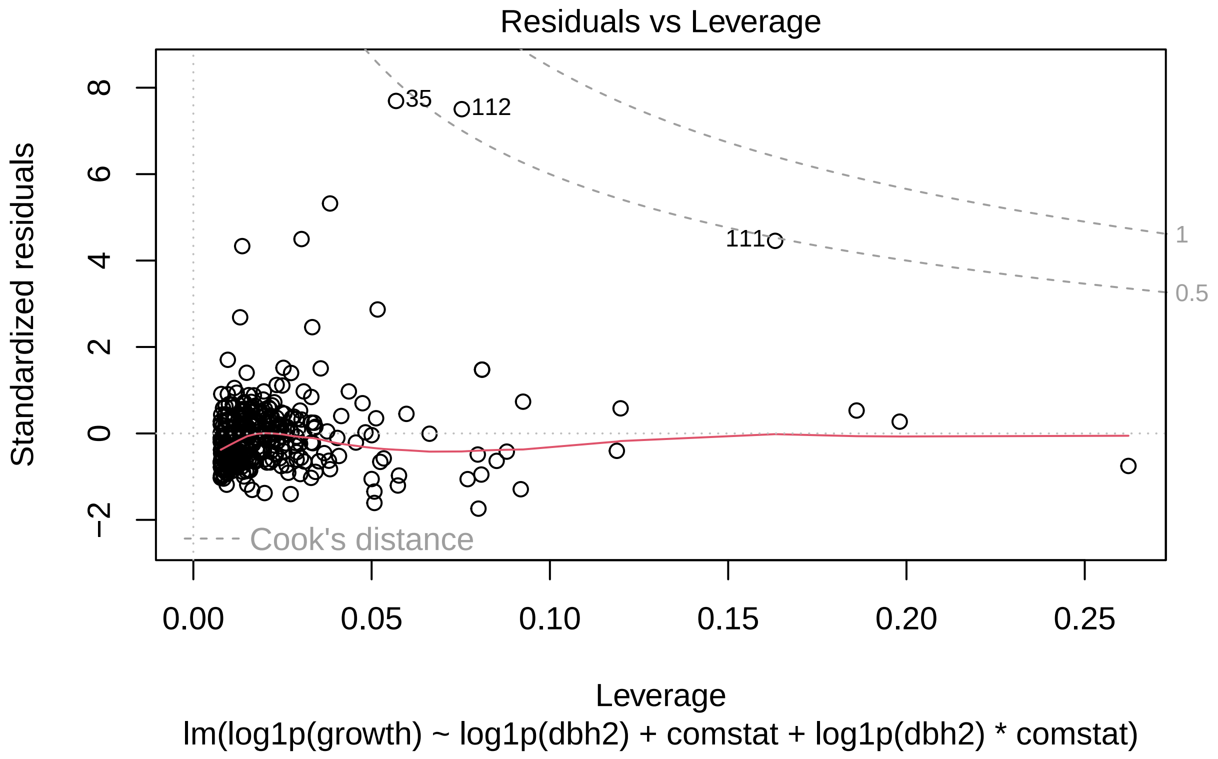
```r
# Pass, it looks okay

# Normality of Errors - Residuals are approximately normally distributed.
qqnorm(residuals(model2))
```



```
# Fail, as it seems to curve up at the end, and other transformations dont seem to help. We can proceed

# No Multicollinearity - Predictor variables are not highly correlated with each other.
cor(cbind(log1p(growth), log1p(dbh2)))
```

```
##           [,1]      [,2]
## [1,] 1.0000000 0.1903679
## [2,] 0.1903679 1.0000000
```

```
# Pass, all below 80%

# No Significant Outliers or Influential Points - Extreme values do not unduly affect the model.
plot(model2, which=5)
```



```r
# Pass, does not overlap cooks distance
```

Check reduced models

```r
model.global <- lm(log1p(growth) ~ log1p(dbh2) + comstat + log1p(dbh2)*comstat, na.action="na.fail")

model.drege<-dredge(model.global)
```

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

```r
head(model.drege, 5)
```

```
## Global model call: lm(formula = log1p(growth) ~ log1p(dbh2) + comstat + log1p(dbh2) *
##     comstat, na.action = "na.fail")
## ---
## Model selection table
##      (Int) cms lg1(db2) cms:lg1(db2) df  logLik   AICc delta weight
## 8 -0.24980   +  0.12950            +  9 146.706 -274.9  0.00  0.940
## 4  0.01543   +  0.04890               6 140.431 -268.6  6.25  0.041
## 3  0.03623      0.04829               3 136.361 -266.6  8.21  0.016
## 2  0.17640   +                        5 136.816 -263.4 11.41  0.003
## 1  0.17500                            2 130.178 -256.3 18.54  0.000
## Models ranked by AICc(x)
```

```
summary(model.drege)
```

```
##   (Intercept)       comstat   log1p(dbh2)      comstat:log1p(dbh2)       df
##  Min.   :-0.24979   +   :3   Min.   :0.04829   +   :1               Min.   :2
##  1st Qu.: 0.01543   NA's:2   1st Qu.:0.04859   NA's:4               1st Qu.:3
##  Median : 0.03623            Median :0.04890                        Median :5
##  Mean   : 0.03066            Mean   :0.07555                        Mean   :5
##  3rd Qu.: 0.17499            3rd Qu.:0.08918                        3rd Qu.:6
##  Max.   : 0.17643            Max.   :0.12946                        Max.   :9
##                              NA's   :2
##      logLik           AICc            delta            weight
##  Min.   :130.2   Min.   :-274.9   Min.   : 0.000   Min.   :0.0000886
##  1st Qu.:136.4   1st Qu.:-268.6   1st Qu.: 6.252   1st Qu.:0.0031300
##  Median :136.8   Median :-266.6   Median : 8.210   Median :0.0155035
##  Mean   :138.1   Mean   :-266.0   Mean   : 8.882   Mean   :0.2000000
##  3rd Qu.:140.4   3rd Qu.:-263.4   3rd Qu.:11.410   3rd Qu.:0.0412547
##  Max.   :146.7   Max.   :-256.3   Max.   :18.539   Max.   :0.9400232
##
```

```
best_model <- model.drege[1]
best_model
```

```
## Global model call: lm(formula = log1p(growth) ~ log1p(dbh2) + comstat + log1p(dbh2) *
##     comstat, na.action = "na.fail")
## ---
## Model selection table
##     (Int) cms lg1(db2) cms:lg1(db2) df  logLik   AICc delta weight
## 8 -0.2498   +   0.1295            +  9 146.706 -274.9     0      1
## Models ranked by AICc(x)
```

```
# Best model has 94% of the weight and second has more than 2 AIC seperation so no model averaging is r

# Model averaging
model.avging<-model.avg(model.drege, subset=TRUE)
summary(model.avging)
```

```
##
## Call:
## model.avg(object = model.drege, subset = TRUE)
##
## Component model call:
## lm(formula = log1p(growth) ~ <5 unique rhs>, na.action = na.fail)
##
## Component models:
##        df logLik    AICc delta weight
## 123     9 146.71 -274.86  0.00   0.94
## 12      6 140.43 -268.61  6.25   0.04
## 2       3 136.36 -266.65  8.21   0.02
## 1       5 136.82 -263.45 11.41   0.00
## (Null)  2 130.18 -256.32 18.54   0.00
##
## Term codes:
```

```
##             comstat         log1p(dbh2) comstat:log1p(dbh2)
##                   1                   2                   3
##
## Model-averaged coefficients:
## (full average)
##                      Estimate Std. Error Adjusted SE z value Pr(>|z|)
## (Intercept)          -0.23304    0.12755     0.12789   1.822  0.06843 .
## comstatD              0.29595    0.35474     0.35603   0.831  0.40583
## comstatI              0.42481    0.16469     0.16506   2.574  0.01006 *
## comstatS              0.21368    0.14955     0.15004   1.424  0.15440
## log1p(dbh2)           0.12446    0.03834     0.03844   3.237  0.00121 **
## comstatD:log1p(dbh2) -0.05987    0.09776     0.09811   0.610  0.54169
## comstatI:log1p(dbh2) -0.13711    0.05397     0.05408   2.535  0.01124 *
## comstatS:log1p(dbh2) -0.04807    0.05115     0.05133   0.937  0.34899
##
## (conditional average)
##                      Estimate Std. Error Adjusted SE z value Pr(>|z|)
## (Intercept)          -0.23304    0.12755     0.12789   1.822 0.068430 .
## comstatD              0.30064    0.35557     0.35687   0.842 0.399541
## comstatI              0.43154    0.15700     0.15739   2.742 0.006110 **
## comstatS              0.21707    0.14827     0.14878   1.459 0.144563
## log1p(dbh2)           0.12486    0.03774     0.03785   3.299 0.000971 ***
## comstatD:log1p(dbh2) -0.06369    0.09961     0.09998   0.637 0.524093
## comstatI:log1p(dbh2) -0.14586    0.04269     0.04284   3.404 0.000663 ***
## comstatS:log1p(dbh2) -0.05114    0.05125     0.05144   0.994 0.320135
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
sw(model.avging)
```

```
##                      log1p(dbh2) comstat comstat:log1p(dbh2)
## Sum of weights:      1.00        0.98    0.94
## N containing models:    3           3       1
```

The model selection table ranks all possible models by AIC, showing that the best-supported model (AIC=0, weight=0.94) includes log(dbh2), comstat, and their interaction as key predictors of growth.

The model summary table shows that the best model includes log(AREA), log(YR.ISOL), and log(ALT), explaining 57% of variation in bird abundance (adjusted R2 = 0.57), with area and years since isolation significant predictors.

The averaged model summary table shows that growth increases with dbh2 and varies by comstat, with significant positive effects for comstat and its interaction with dbh2, indicating that dbh2's influence on growth depends on comstat.

The relative importance table shows that log(dbh2) (1.00) and comstat (0.98) are the strongest predictors of growth, while their interaction (0.94) is slightly less influential but still well supported in the top models.

Question 3

The study used over 7000 tree-ring records from ~3000 plots across the US Mountain West to test whether declining growth resilience to drought predicts forest mortality. Most species showed decreasing drought resistance and resilience since the 1950s, which was linked to higher stand-level mortality. Growth and resilience together explained about half of mortality variation, showing that reduced resilience reflects physiological decline and can serve as an early warning of die-off. AIC was used for model selection to identify the strongest mortality predictors The authors included about 40 variables and selected the best models based on lowest AIC, favoring simpler models.