KU LEUVEN

Linear Models

Project

Dec 2021

Contents

Introduction

The Dwarf Squeaker (Arthroleptis xenodactyloides) is a species of frog in the Arthroleptidae family which lives in cloud forests of the Eastern Arc Mountains in Africa. The size of the frog is quite small which is just about the size of a human fingernail. They can thrive in many habitats such as lowland montane forests and are believed to be a generalist. With accelerating deforestation and increasing critical forest habitat destruction, the frogs are facing habitat loss. Understanding the ecology of the frogs is important for protecting their survival. We want to investigate which factors can influence the length of the frogs. It is told that the species benefits from holes in the cloud forest's canopy or shrub layers. Because the holes allow sunlight to reach and warm up the ground floor of the forests which is helpful for this particular species. These holes, especially those in the canopy, can exist naturally but are appearing in increasing numbers due to human activities.
A biologist studies this species in Kenya and has investigated a large number of sites in three forests (Ngangao South, Ngangao North, and Chawia). She searched in randomly selected patches and examined the found one in her field lab to get related variables. The goal of this project is to build a model of $Length$ as a function of the other variables she collected in the dataset. We expect to find the model which can predict the *Length* well and satisfy the assumptions of the model.

Exploratory Analysis

In this part, we will try to have an exploratory analysis of the data. We split into dataset into two sets: a training set (160) and a validation set (160). The training set is used to develop the model and the validation set is used to evaluate the predictive ability of the selected model.

Variable	Description
Length	Body length (cm)
Sex	male or female
Size	Size of patch (m^2)
Canopy	Proportion of patch covered by canopy
Shrub	Proportion of patch covered by shrubs
Effort	The time it took to find the individual (minutes)
Natural	Indicator whether the site was naturally intact, i.e., unspoilt by any human activity (yes or no)
Forest	*Ngangao N, Ngangao S, or Chawia*

Tab 1 Variables in the Dataset

As shown in Tab 1, the response variable is *Length* which is the body length of the frog. There are 7 independent variables in the dataset among which are 3 categorical variables. Canopy and shrub covers are based on image analysis of a picture of the respective canopy and shrub layers at each patch. *Effort* is the time it took to find the individual and large individuals are assumed to be found more easily. Firstly, the categorical variables will be converted into dichotomous or polytomous predictor variables. We will define

$$Sex^{(1)} = \begin{cases} 1 & if\ individual\ i\ is\ male\ frog \\ 0 & if\ individual\ i\ is\ female\ frog \end{cases}$$

$$Natural^{(1)} = \begin{cases} 1 & if\ found\ in\ \ the\ site\ which\ was\ naturally\ intact \\ 0 & if\ found\ in\ the\ site\ which\ was\ not\ naturally\ intact \end{cases}$$

$$Forest^{(1)} = \begin{cases} 1 & if\ in\ \ Ngangao\ North \\ 0 & if\ not\ in\ Ngangao\ North \end{cases}$$
$$Forest^{(2)} = \begin{cases} 1 & if\ in\ \ Ngangao\ South \\ 0 & if\ not\ in\ Ngangao\ South \end{cases}$$

```
    Sex          Canopy           Shrub            Effort           Size           Length       Natural       Forest
 female:85   Min.   :0.2100   Min.   :0.1000   Min.   : 5.00   Min.   : 5.00   Min.   :1.370   no :72   Chawia    :55
 male  :75   1st Qu.:0.4300   1st Qu.:0.2900   1st Qu.:10.00   1st Qu.:12.00   1st Qu.:1.637   yes:88   Ngangao N:46
             Median :0.5950   Median :0.5100   Median :17.00   Median :21.00   Median :1.925            Ngangao S:59
             Mean   :0.5978   Mean   :0.5158   Mean   :16.88   Mean   :21.38   Mean   :1.917
             3rd Qu.:0.7900   3rd Qu.:0.7600   3rd Qu.:23.00   3rd Qu.:30.00   3rd Qu.:2.150
             Max.   :0.9500   Max.   :0.9500   Max.   :30.00   Max.   :40.00   Max.   :2.760
```

Fig 1 Summary of Variables

The summary of variables in the training set is shown in Fig 1. Among the 160 observations in the training dataset, there are 85 female frogs and 78 male frogs. 72 of the frogs were found in the site which was naturally intact while 88 of them were found in the site which was not naturally intact. To get an intuitive understanding, boxplots of length versus these variables are shown in Fig 2. Clearly, the length of the female frogs is significantly larger than male ones. There might be some outlying observations. Whether the site was naturally intact or not has a quite small influence on the median length of the frogs. As for the forests, the dwarf frogs in *Chawia* tends to be longer than the other two forests.

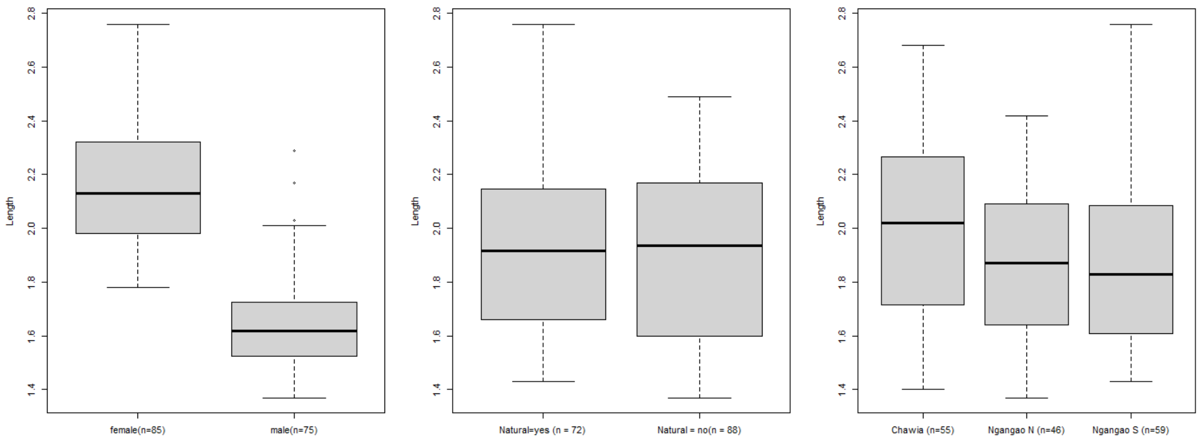

Fig 2 Boxplot of *Length* versus *Sex, Natural, Forest*

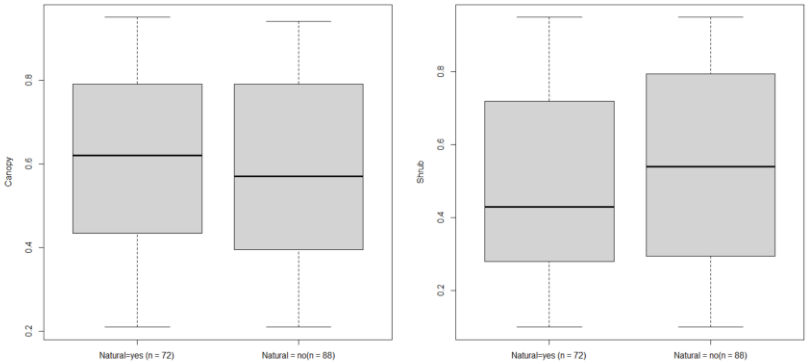

Fig 3 Boxplot of *Canopy/Shrub* versus *Natural*

As mentioned before, the frogs benefits from holes in the cloud forest's canopy or shrub layers. The median of *Canopy* in the site which was naturally intact seems larger than *Canopy* in non-intact ones. The median proportion of *Shrub* in the natural sites is less than ones in non-intact sites. Whether the site is naturally intact or not seems to have different influence on the proportions of canopy and shrub. As shown in Fig 4, we can learn about the histograms of variables.

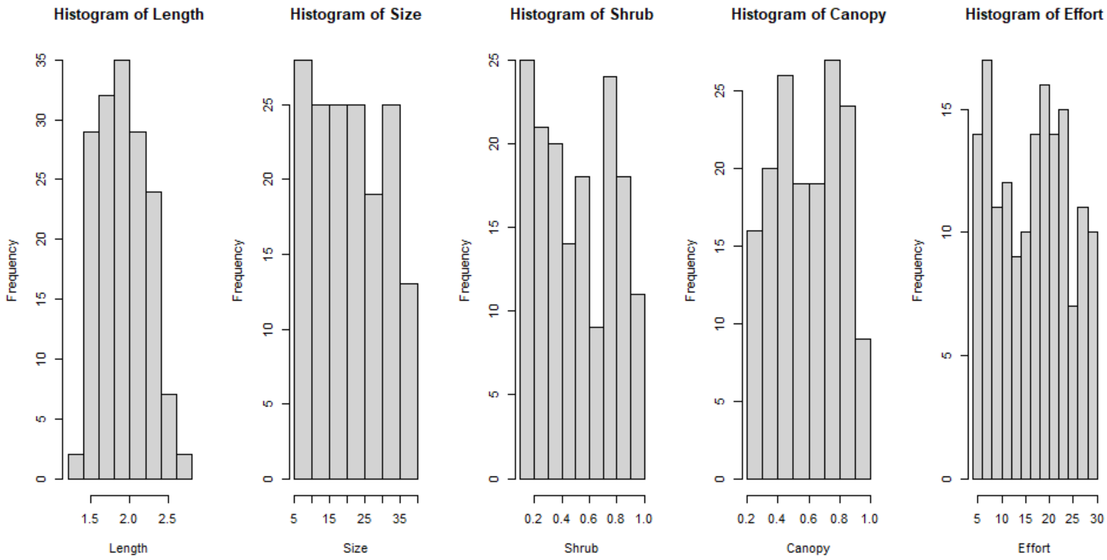

Fig 4 Histogram of Variables

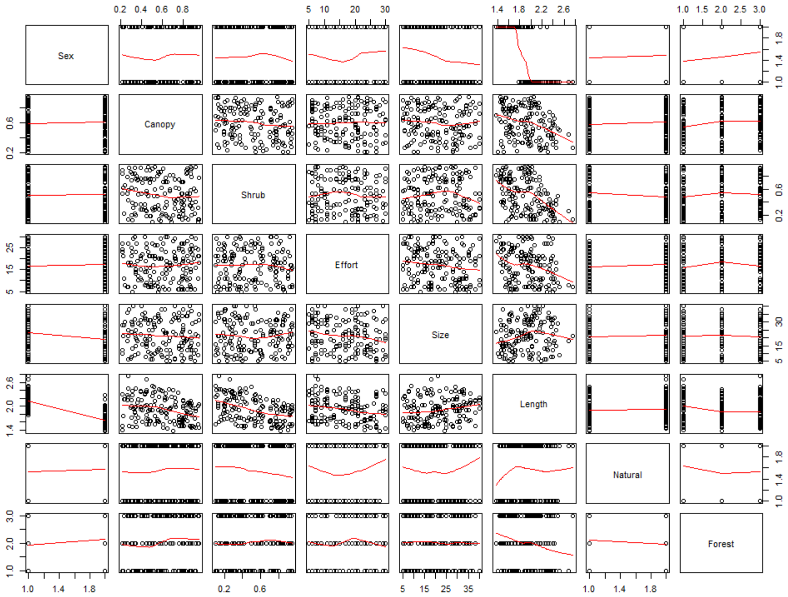

Fig 5 Scatter Plot

The scatter plot of variables is shown in Fig 5. *Canopy/Shrub/Effort* seems to be negatively correlated with *Length* while *Length* versus *Size* has positive trends. The relationships need to be checked in the model.

	Canopy	Shrub	Effort	Size
Canopy	1.00000000	-0.13452026	0.03051447	-0.07199762
Shrub	-0.13452026	1.00000000	-0.05462527	0.00714414
Effort	0.03051447	-0.05462527	1.00000000	-0.14189613
Size	-0.07199762	0.00714414	-0.14189613	1.00000000

Tab 2 Correlation Between Variables

Variable	Canopy	Shrub	Effort	Size
VIF	1.023875	1.021184	1.023735	1.025443

Tab 3 VIF of Variables

Tab 2 shows the correlations between continuous variables are rather low, which does not suggest a high possibility of multicollinearity between them. At the same time, variance inflation factors in Tab 3 also suggest there seems to be no multicollinearity between those variables since all the VIFs are close to 1.

Regression Models

In this part, we will build a set of models to investigate the relationship between the response and explanatory variables. We want to predict the body length of the dwarf squeaker by the variables in the dataset. First, we fit all the variables in the dataset to get an overview of the full model.

```
Call:
lm(formula = Length ~ ., data = data.training)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.188036 -0.055843 -0.004967  0.042915  0.257776 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      2.8690118  0.0363306  78.970   <2e-16 ***
Sexmale         -0.4739086  0.0136449 -34.732   <2e-16 ***
Canopy          -0.4745835  0.0321166 -14.777   <2e-16 ***
Shrub           -0.5566277  0.0259808 -21.425   <2e-16 ***
Effort          -0.0089393  0.0009200  -9.716   <2e-16 ***
Size            -0.0007954  0.0006987  -1.138   0.2567    
Naturalyes       0.0317716  0.0135992   2.336   0.0208 *  
ForestNgangao N -0.0204215  0.0172218  -1.186   0.2376    
ForestNgangao S -0.0081697  0.0160867  -0.508   0.6123    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0839 on 151 degrees of freedom
Multiple R-squared:  0.9319,    Adjusted R-squared:  0.9283 
F-statistic: 258.3 on 8 and 151 DF,  p-value: < 2.2e-16
```

Fig 6 Summary of full model

The adjusted R^2 is 0.9283, which is not bad. Cleary, the parameters of *Size and Forest* are not significant while other parameters of variables are significant.

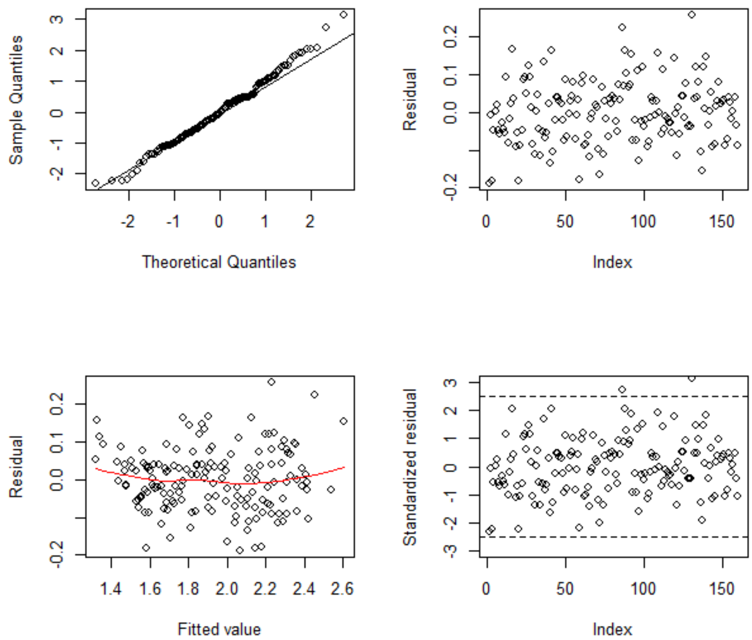

Fig 7 Plot of Full Model

The residuals plots are helpful to check the validity of model assumptions. As shown in Fig 7, norm quantile plot tells that the error terms are not quite normally distributed. There is not clear pattern in the plot of residual versus index. The plot of residuals versus fitted values shows a non-horizontal band which indicates that the linear model is a little defective. Also, there are probably 2 outliers told by the plot of standardized residuals versus index.

Variable selection

Stepwise regression procedures are used to develop a sequence of regression models, such that at each step an explanatory variable is added to or removed from the model. The selected model is

$$Length_i = \beta_0 + \beta_1 Sex_i^{(1)} + \beta_2 Canopy_i + \beta_3 Shrub_i + \beta_4 Effort_i + \beta_5 Natural_i^{(1)} + \varepsilon_i$$

```
lm(formula = Length ~ Sex + Canopy + Shrub + Effort + Natural,
    data = data.training)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.189135 -0.061750 -0.004084  0.044177  0.253417 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.844793   0.030831  92.270   <2e-16 ***
Sexmale     -0.471640   0.013328 -35.387   <2e-16 ***
Canopy      -0.476649   0.031637 -15.066   <2e-16 ***
Shrub       -0.559714   0.025836 -21.664   <2e-16 ***
Effort      -0.008944   0.000901  -9.927   <2e-16 ***
Naturalyes   0.032090   0.013464   2.383   0.0184 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08386 on 154 degrees of freedom
Multiple R-squared:  0.9306,	Adjusted R-squared:  0.9284 
F-statistic: 413.1 on 5 and 154 DF,  p-value: < 2.2e-16
```

Fig 8 Summary of Model After Variable Selection

As shown in Fig 8, all the parameters of the model are significant except that *Nature* is on the borderline significant. The R^2 is 0.9306 which means that 93.06% of the total variance in the response *Length* that is explained by the linear model. The adjusted R_a^2 which corrects for the number of variables is 0.9284. The coefficient of *Sex* which is consistent with the boxplot of *Length* versus *Sex.* The coefficients of *Canopy/Shrub/Effort* are all negative which is consistent with expectations. Next, we will investigate the model assumptions.

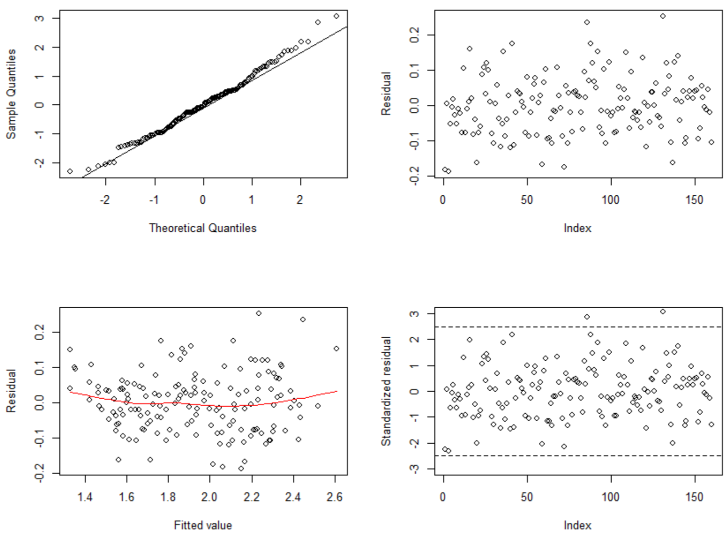

Fig 9 Plot of Model After Variable Selection

As shown in Fig 9, the up-left figure tells that there are some deviations from normal distribution at the tails which is the sign of violating the assumption of normality. The little curve of residuals versus fitted values suggests the linear model is not appropriate but we have to say the curve is not very obvious. It seems that variance of residuals increases with the level of the response variable. At same time, 2 data points have absolute standardized residual larger than 2.5 and can be pinpointed as outliers. In total, model assumptions are not satisfied well in this model and there is some space for building a better model.

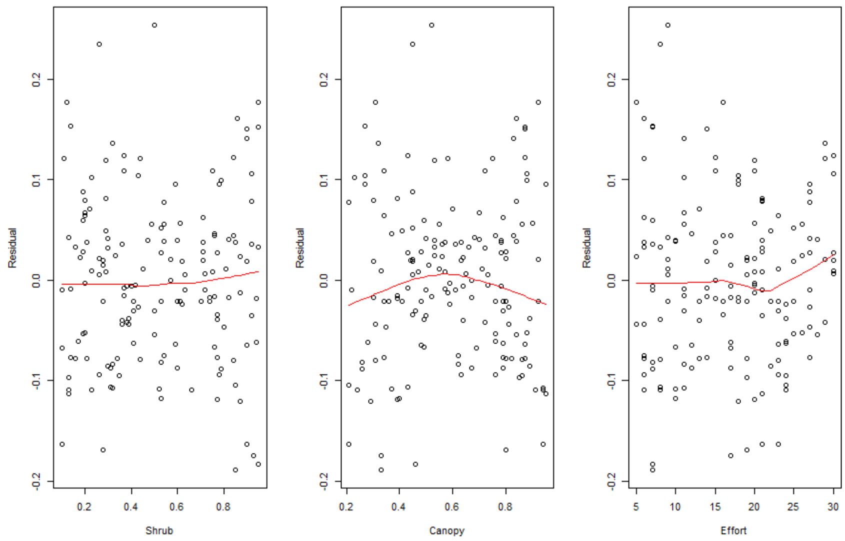

Fig 10 Plot of residual versus Variables

As shown in Fig 10, plots indicate that the linear model is defective, and we may add quadratic terms in next steps. At the same time, the errors are a little heteroscedastic.

Adding interaction terms

Since the little defection of the model after variable selection, we try to add interaction terms to the regression model. The holes in the cloud forest's canopy or shrub layers are important for the survival of the frogs. Different from other species, the frogs will benefit from the holes because these holes allow sunlight to reach. Canopy and shrub layers are in different heights and there may be an interaction between the two variables[1]. After several tries, we postulate the model

$$Length_i = \beta_0 + \beta_1 Sex_i^{(1)} + \beta_2 Canopy_i + \beta_3 Shrub_i + \beta_4 Effort_i + \beta_5 Natural_i^{(1)} + \beta_6 Shrub_i * Canopy_i + \varepsilon_i$$

[1] Armstrong, A.H, et al. "A Multi-Scaled Analysis of Forest Structure Using Individual-Based Modeling in a Costa Rican Rainforest." Ecological Modelling, vol. 433, 2020, p. 109226.

```
Call:
lm(formula = Length ~ Sex + Canopy + Shrub + Effort + Natural +
    Canopy:Shrub, data = data.training)

Residuals:
      Min        1Q    Median        3Q       Max
-0.144285 -0.029303 -0.003108  0.031333  0.246731

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.1933435  0.0335145  95.282  < 2e-16 ***
Sexmale      -0.4642187  0.0091040 -50.991  < 2e-16 ***
Canopy       -1.0595885  0.0486945 -21.760  < 2e-16 ***
Shrub        -1.2160268  0.0522123 -23.290  < 2e-16 ***
Effort       -0.0082946  0.0006162 -13.461  < 2e-16 ***
Naturalyes    0.0245021  0.0091970   2.664  0.00855 **
Canopy:Shrub  1.0906791  0.0816811  13.353  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05717 on 153 degrees of freedom
Multiple R-squared:  0.968,     Adjusted R-squared:  0.9667
F-statistic: 770.3 on 6 and 153 DF,  p-value: < 2.2e-16
```

Fig 11 Summary of Model with Interaction

As shown in Fig 11, all the parameters of the model are significant even *Natural* is more significant than before. The adjusted R_a^2 is 0.9667 which is improved a lot. In total, the results of the model with an interaction term seem better than before but we still need to verify the model assumptions.

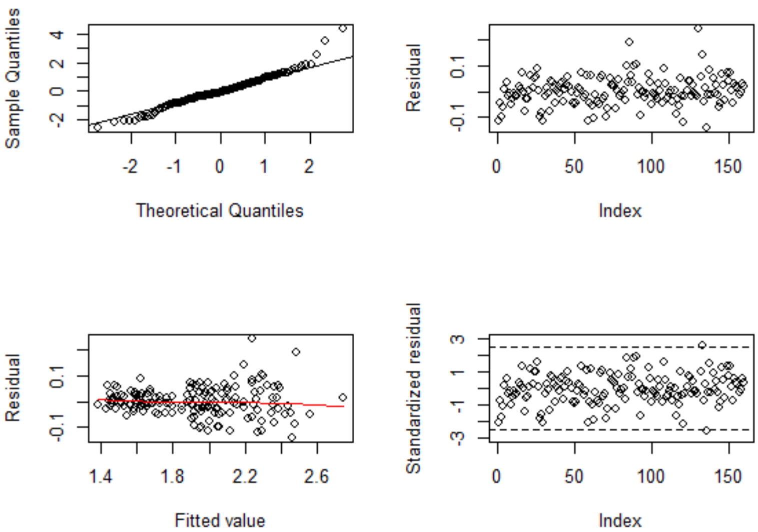

Fig 12 Plot of Model with Interaction

As shown in Fig 12, there are some deviations from normal distribution at right tails which is even worse. The plot of residual versus fitted values shows nonconstant error variance. There are still 2 outliers in the borderline but not so worse. In total, model assumptions seem to be satisfied better than the model without an interaction term, but the normality of error terms is still not satisfied well and there is little heteroscedasticity.

Weighted least square regression

Because the defection of the model with interaction, we thus apply weighted least square regression and try to get a better model. As shown in Fig 13, the results of the weighted least square regression model seem fine. All the parameters are significant and adjusted R^2 has improved compared with model (with interaction).

As shown in Fig 14, we can find that the majority of the data points lie around the line which means the normality has improves. The plot of weighted residual versus fitted values indicates that heteroscedasticity has improves than before.

```
Call:
lm(formula = Length ~ Sex + Canopy + Shrub + Effort + Natural +
    Canopy:Shrub, weights = w)

Weighted Residuals:
     Min       1Q   Median       3Q      Max
-0.20875 -0.04394  0.00086  0.04410  0.19256

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.2151272  0.0284040  113.19  < 2e-16 ***
Sexmale      -0.4684642  0.0077545  -60.41  < 2e-16 ***
Canopy       -1.0676307  0.0350000  -30.50  < 2e-16 ***
Shrub        -1.1971652  0.0417604  -28.67  < 2e-16 ***
Effort       -0.0091640  0.0004071  -22.51  < 2e-16 ***
Naturalyes    0.0158865  0.0059054    2.69  0.00793 **
Canopy:Shrub  1.0779880  0.0565505   19.06  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07148 on 153 degrees of freedom
Multiple R-squared:  0.9796,	Adjusted R-squared:  0.9788
F-statistic:  1225 on 6 and 153 DF,  p-value: < 2.2e-16
```

Fig 13 Summary of Model Using WLS

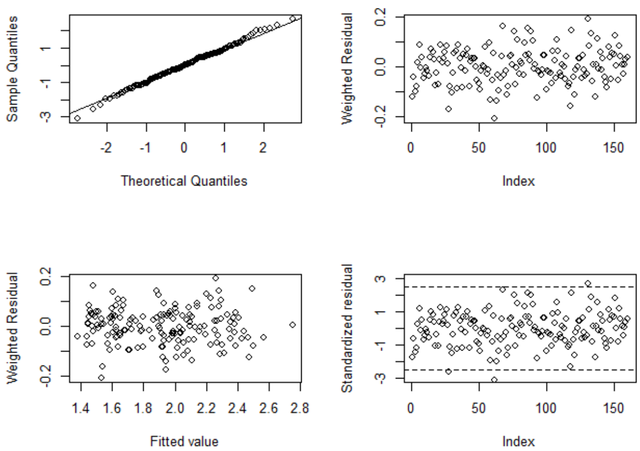

Fig 14 Plot of Model Using WLS

Box-Cox Transformation

The Box-Cox transformation can be used to normalize the error distribution, stabilize the error variance, and straighten the relation between *Y* and *X*. It will be useful to solve the problems in the models we built before. In this part, we will use it to find a better model. The transformation is defined as

$$y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda} & if\ \lambda \neq 0 \\ \ln y_i & if\ \lambda = 0 \end{cases}$$

Note that y_i should be positive, otherwise a constant should first be added. In this case, *Length* is positive so it will not be an issue.

We will build Box-Cox transformation based on model with interaction to see if the problems can be solved. As shown in Fig 15, we get the $\lambda = 0.38$. The proposed model is

$$Length_i^{(\lambda)} = \beta_0 + \beta_1 Sex_i^{(1)} + \beta_2 Canopy_i + \beta_3 Shrub_i + \beta_4 Effort_i + \beta_5 Natural_i^{(1)} + \beta_5 Canopy_i * Shrub_i + \varepsilon_i$$

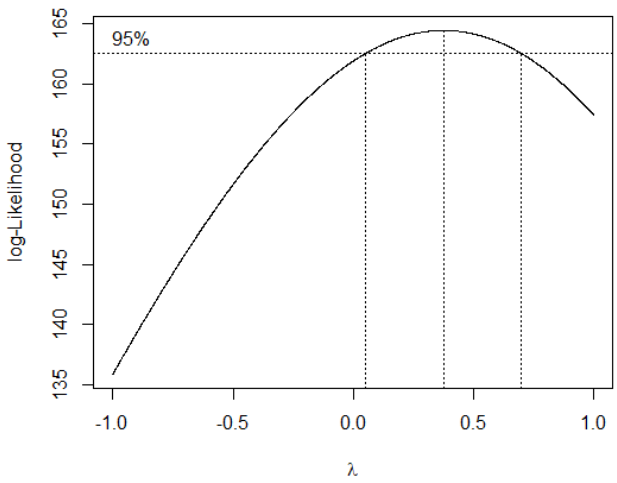

Fig 15 Maximized log likelihood versus λ

As shown in Fig 16, all the parameters of predictor variables are significant. It is worth noting that the parameter of *Natural* is more significant compared with the model without interaction terms. The positive or negative impact of each variable is same as in the weighted method. The adjusted R_a^2 also increases compared with the model without interaction terms.

```
Call:
lm(formula = ((Length)^lambda - 1)/lambda ~ Sex + Canopy + Shrub +
    Effort + Natural + Canopy * Shrub)

Residuals:
      Min        1Q    Median        3Q       Max
-0.099841 -0.022371 -0.002176  0.025250  0.145078

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.5607616  0.0215637  72.379  < 2e-16 ***
Sexmale      -0.3127486  0.0058576 -53.392  < 2e-16 ***
Canopy       -0.6839882  0.0313307 -21.831  < 2e-16 ***
Shrub        -0.7870867  0.0335941 -23.429  < 2e-16 ***
Effort       -0.0055039  0.0003965 -13.882  < 2e-16 ***
Naturalyes    0.0184393  0.0059175   3.116  0.00219 **
Canopy:Shrub  0.6989258  0.0525548  13.299  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03679 on 153 degrees of freedom
Multiple R-squared:   0.97,     Adjusted R-squared:  0.9688
F-statistic: 824.4 on 6 and 153 DF,  p-value: < 2.2e-16
```

Fig 16 Summary of Model Using Box-Cox Transformation

It is important to verify whether the proposed transformation can improve the appropriateness of the model assumption. Similarly, we can analysis this by using residual plots and normal quantile plots. As shown in Fig 17, the normality of error terms has improved but there is still a little deviation at the right tail. Plot of the residuals versus index does not show any pattern. The plot of the residuals versus fitted values shows curve and heteroscedasticity which means that the model is not appropriate. There are still several outliers in the model. Although the model has improved, this model cannot satisfy the model assumptions well.

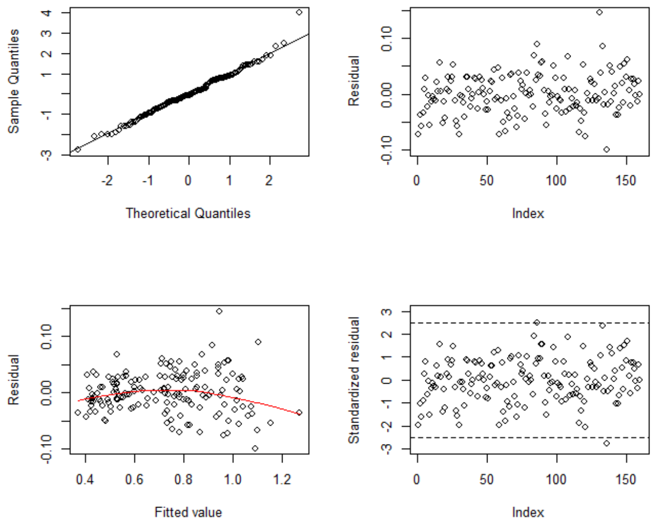

Fig 17 Plot of Model Using Box-Cox Transformation

Robust method

It is known that the least-squares estimator is very sensitive to vertical outliers and bad leverage points. We can find that the plot of standardized residuals always shows that there may be some outliers. It is worth using robust method to build a model and find if we get better models. Based on the model with an interaction term, the plots of detection methods are presented in Fig 18: standardized residual detects 2 outliers using 2.5 as a threshold; studentized residual detects 3 outliers; diagonal elements of hat matrix find 2 outliers; DFFITS detects 3 influential points. As for cook's distance, there are no outliers detected.

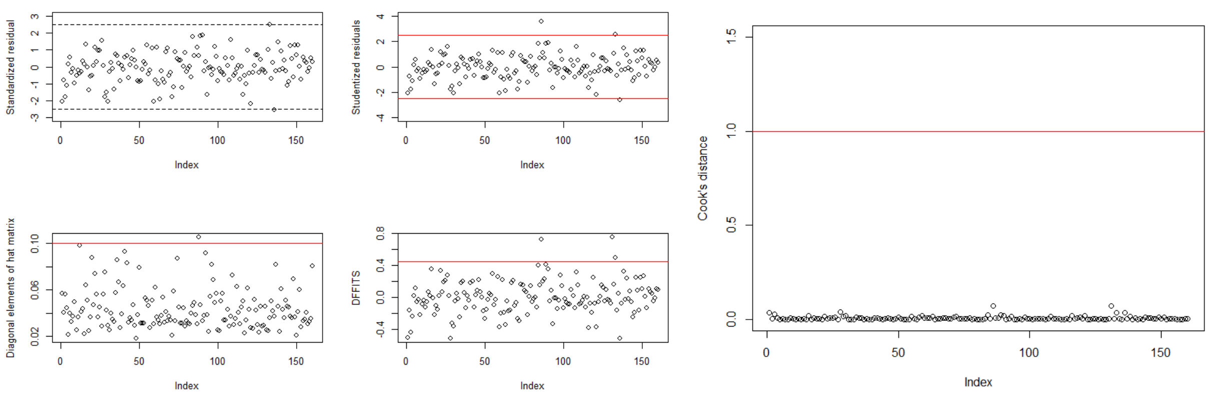

Fig 18 Detection of Outliers of Model with Interaction

The LTS estimators try to fit the majority of the residuals as small as possible. The breakdown value indicates how many of the observations need to be replaced before the estimate is carried away. Because there is a small proportion of outliers, we try to use 5% as breakdown value and built a robust model.

```
Call:
ltsReg.formula(formula = Length ~ Sex + Canopy + Shrub + Effort +
    Natural + Canopy * Shrub, data = data.training, alpha = 0.95)

Residuals (from reweighted LS):
      Min        1Q    Median        3Q       Max
-0.114357 -0.028561 -0.002366  0.029175  0.111777

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
Intercept     3.1787571  0.0289251 109.896  < 2e-16 ***
Sexmale      -0.4617954  0.0077701 -59.432  < 2e-16 ***
Canopy       -1.0549987  0.0413721 -25.500  < 2e-16 ***
Shrub        -1.1976639  0.0442779 -27.049  < 2e-16 ***
Effort       -0.0080467  0.0005314 -15.142  < 2e-16 ***
Naturalyes    0.0256432  0.0078673   3.259  0.00138 **
Canopy:Shrub  1.0740584  0.0690627  15.552  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04818 on 148 degrees of freedom
Multiple R-Squared: 0.9759,     Adjusted R-squared: 0.975
F-statistic:  1001 on 6 and 148 DF,  p-value: < 2.2e-16
```

Fig 19 Summary of Robust Model

As shown in Fig 19, all the coefficients are significant and the R^2 is relatively high. We obtain the diagnostic plot, on which the four types of observations can be distinguished as in Fig 20. We find that there are 4 vertical outliers and no bad leverage points.

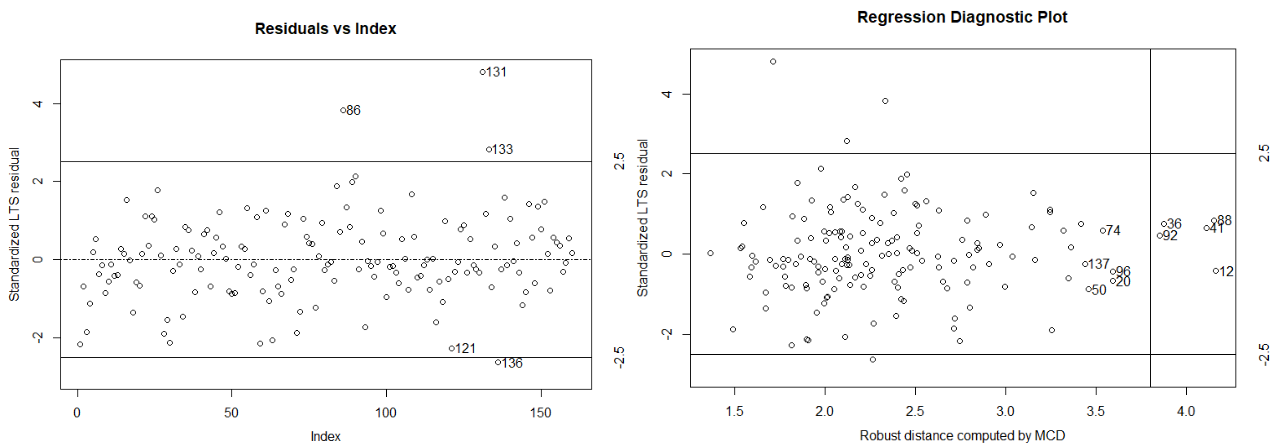

Fig 20 Detection of Outliers Using Robust Method

There will be a problem in predicting when we use lts.Reg and it suggests using lmrob which has a predict method. Actually, the two functions give simialr results of coefficeints, standard error, and p values.

```
Call:
lmrob(formula = Length ~ Sex + Canopy + Shrub + Effort + Natural + Canopy *
    Shrub, data = data.training)
 \--> method = "MM"
Residuals:
      Min        1Q    Median        3Q       Max
-0.134118 -0.029471 -0.002941  0.031163  0.254380

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.1726111  0.0318947  99.471  < 2e-16 ***
Sexmale      -0.4601950  0.0088293 -52.122  < 2e-16 ***
Canopy       -1.0510706  0.0412526 -25.479  < 2e-16 ***
Shrub        -1.1981194  0.0471935 -25.387  < 2e-16 ***
Effort       -0.0079735  0.0006338 -12.580  < 2e-16 ***
Naturalyes    0.0249196  0.0082486   3.021  0.00295 **
Canopy:Shrub  1.0784130  0.0663268  16.259  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Robust residual standard error: 0.04703
Multiple R-squared:  0.9742,    Adjusted R-squared:  0.9732
Convergence in 13 IRWLS iterations
```

Fig 21 Results of LMROB

Validation of models

In the regression model part, we build several models via the training set and check their model assumptions. Based on the results, we will use the 3 models to validate the models via the test dataset. We define

$$Length_i = \beta_0 + \beta_1 Sex_i^{(1)} + \beta_2 Canopy_i + \beta_3 Shrub_i + \beta_4 Effort_i + \beta_5 Natural_i^{(1)} + \beta_6 Shrub_i * Canopy_i + \varepsilon_i$$

Final Models:

(1) Final Model 1: the model using weighted least square regression.

(2) Final Model 2: the model using Box-Cox transformation which transform $Length$ into $(Length^{\lambda} - 1) / \lambda$.

(3) Final Model 3: the robust model using LTS.

MESP (mean squared predictor error) gives the mean of the square prediction errors of the new data. We can rely on *MSEP* as an indicator of how well the selected model will predict. The MSEP of model 1 is 0.003937; MSEP of model 2 is 0.001842; MSEP of model 3 is 0.004049. Clearly, model 2 has the best ability to predict in the test dataset. MSEP of model 1 is similar to the MSEP of model 3. Every model has its pros and cons. Final model 1 is the model using WLS which can give a good prediction of test dataset. About model assumptions, it can improve the normality of errors and heteroscedasticity. Also, the interpretation of the model is easier which is of importance in practice. Final Model 2 can give the best prediction and the model assumptions seem better than original model. But we have to say there is little heteroscedasticity and it has the disadvantage of changing the relation between the response and the independent variables. Final model 3 has a small MSEP. But it will drop some observations and the model will lose some information. Compared with final model 2, the MSEP of final model 3 is a little larger.

In sum, take model assumptions, prediction, model fitting and interpretation into consideration, we will choose the final model 1 which is the model using WLS.

Conclusion

Through the analysis of the model assumptions and ability of prediction in new data set, we define the final model

$$Length_i = 3.2151 - 0.4685Sex_i^{(1)} - 1.0676Canopy_i - 1.1972Shrub_i - 0.0092Effort_i + 0.0159Natural_i^{(1)} + 1.0780Canopy_i * Shrub_i + \varepsilon_i$$

In conclusion, the selected variables in the model have influence of the body length of the frogs.

If the frog is male and found in the site which was naturally intact,

$$Length_i = 2.7625 - 1.0676Canopy_i - 1.1972Shrub_i - 0.0092Effort_i + 1.0780Canopy_i * Shrub_i + \varepsilon_i$$

If the frog is female and found in the site which was naturally intact,

$$Length_i = 3.2310 - 1.0676Canopy_i - 1.1972Shrub_i - 0.0092Effort_i + 1.0780Canopy_i * Shrub_i + \varepsilon_i$$

If the frog is male and found in the site which was not naturally intact,

$$Length_i = 2.7466 - 1.0676Canopy_i - 1.1972Shrub_i - 0.0092Effort_i + 1.0780Canopy_i * Shrub_i + \varepsilon_i$$

If the frog is female and found in the site which was not naturally intact,

$$Length_i = 3.2151 - 1.0676Canopy_i - 1.1972Shrub_i - 0.0092Effort_i + 1.0780Canopy_i * Shrub_i + \varepsilon_i$$

If the frog is same-sex and found in the site which has the same situation of naturally intact, 10% increase in the proportion of patch covered by canopy results in an average *Length* change of - 0.10676 + 0.10780* $Shrub$ when all other variables are held constant. Similarly, 10% increase in the proportion of patch covered by shrub results in an average *Length* change of - 0.11972 + 0.10780*$Canopy$ when all other variables are held constant. Although *Effort* is not an ecological variable, it is associated with prevalence of the frogs. The parameter of *Effort* is negative which is consistent with the expectation that large individuals are assumed to be found easily.