An Exploration of Statistical Modelling Methods on Simulation Data

Case Study: Biomechanical Predator–Prey Simulations

by

Christian Seto

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2018 by the
Graduate Supervisory Committee:

Theodore Pavlic, Chair
Jing Li
Hao Yan

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

Modern, advanced statistical tools from data mining and machine learning have become commonplace in molecular biology in large part because of the "big data" demands of various kinds of "-omics" (e.g., genomics, transcriptomics, metabolomics, etc.). However, in other fields of biology where empirical data sets are conventionally smaller, more traditional statistical methods of inference are still very effective and widely used. Nevertheless, with the decrease in cost of high-performance computing, these fields are starting to employ simulation models to generate insights into questions that have been elusive in the laboratory and field. Although these computational models allow for exquisite control over large numbers of parameters, they also generate data at a qualitatively different scale than most experts in these fields are accustomed to. Thus, more sophisticated methods from big-data statistics have an opportunity to better facilitate the often-forgotten area of bioinformatics that might be called "*in-silico*mics".

As a case study, this thesis develops methods for the analysis of large amounts of data generated from a simulated ecosystem designed to understand how mammalian biomechanics interact with environmental complexity to modulate the outcomes of predator–prey interactions. These simulations investigate how other biomechanical parameters relating to the agility of animals in predator–prey pairs are better predictors of pursuit outcomes. Traditional modelling techniques such as forward, backward, and stepwise variable selection are initially used to study these data, but the number of parameters and potentially relevant interaction effects render these methods impractical. Consequently, new modelling techniques such as LASSO regularization are used and compared to the traditional techniques in terms of accuracy and computational

complexity. Finally, the splitting rules and instances in the leaves of classification trees provide the basis for future simulation with an economical number of additional runs. In general, this thesis shows the increased utility of these sophisticated statistical techniques with simulated ecological data compared to the approaches traditionally used in these fields. These techniques combined with methods from industrial Design of Experiments will help ecologists extract novel insights from simulations that combine habitat complexity, population structure, and biomechanics.

DEDICATION

For my parents who have given me love, provided mental (and financial) support, and have encouraged me to pursue my goals and aspirations.

ACKNOWLEDGMENTS

First, I would like to thank Dr. Theodore Pavlic, thesis committee chair, for providing ideas, knowledge and mental support throughout the thesis process. I would also like to thank Dr. Rebecca Wheatley, a postdoctoral researcher at the University of Queensland, for providing the simulation data as well as answering questions about the inner workings of her NetLogo predation simulation. Finally, I would like to thank all the other committee members, Dr. Jing Li and Dr. Hao Yan, for their time and involvement in the thesis process.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

INTRODUCTION

Due to the declining cost of data storage as well as the increase of data generated from smartphones and social media, new modelling techniques have been developed in the fields of data mining and machine learning. Many scientific fields are utilizing these new techniques, particularly the social sciences like biology and ecology. One area of interest that has not been examined with these new techniques is agent-based simulations, particularly ecological simulations of predation. Predation is a major biological factor that influences animal behavior, pack structure, and ecosystems. One instance of predation is pursuit predation, which is when a single predator or group of predators chases and attempts to catch fleeing prey. In the past, success of the predator was assumed to be predicted solely by the difference in the top speeds of the predator and the prey it pursues. However, there is growing evidence that speed–agility tradeoffs play more of a role and suggest that agile prey can escape a faster predator if prey can force a predator to run at lower speeds (Wilson, et al., 2018). To understand how this may occur in natural scenarios, an agent-based simulation of predator–prey pursuits in habitats of varying complexity was developed. This thesis focuses on exploring and applying different statistical techniques to data generated from this simulation for the case of a single predator pursuing a single prey over different biomechanical and environmental parameter values. The goals of this thesis include understanding which variables result in a predator success (predator catches prey) or a prey success (prey escapes predator) as well as developing an iterative process to narrow down the variable ranges for future simulations. Specifically, traditional modelling and model selection techniques were first applied, and their performance were compared to more sophisticated techniques that are less computationally complex. Then, a classification tree approach was developed to assist in the economical design of future experiments with the simulator. It was shown that these more sophisticated techniques can significantly improve the analysis pipeline for complex ecological simulations.

Background on Simulation and Data

The biomechanical predator/prey interaction was modelled in a currently unpublished NetLogo program developed by collaborator Rebecca Wheatley. Figure 1 is a screenshot of the graphical user interface for this program. It includes sliders to manipulate the initial variables, a graph of the velocities of the prey and predator over time, and a visual representation of the predation chase on the right side (the white spider represents the predator, and the orange mouse represents the prey). In addition, the various brown shapes represent obstacles, and the green shapes represent safe zones for the prey.

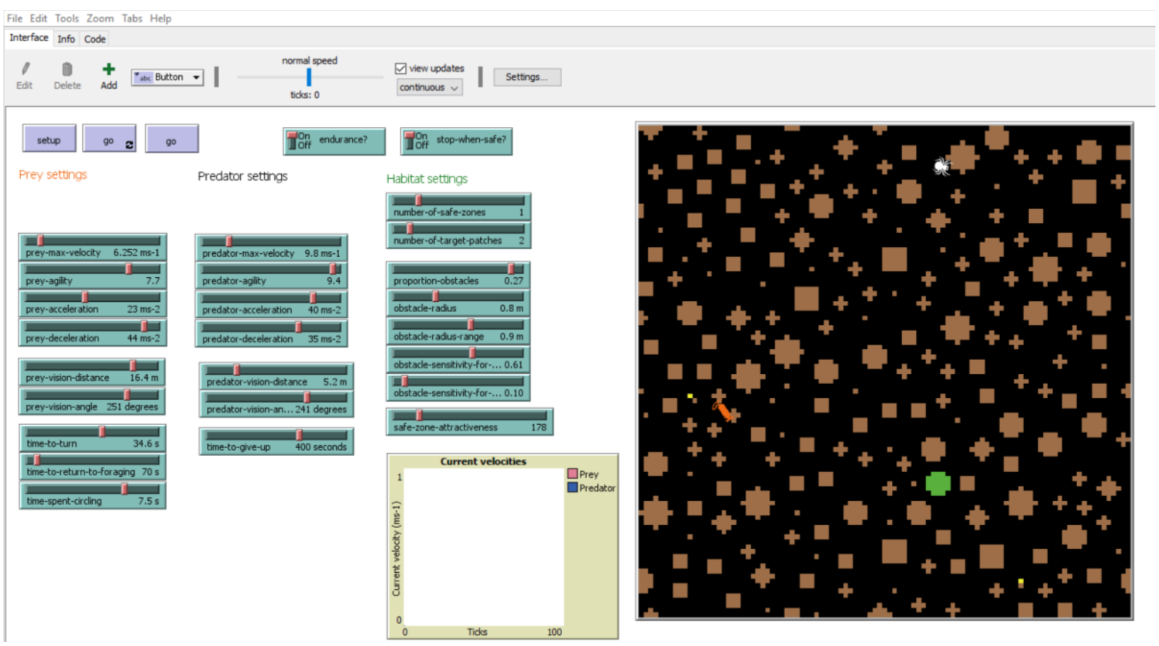

Figure 1: Screenshot of the NetLogo Simulation Program Interface

As the number of sliders indicate, the program includes 24 different variables that set the initial conditions for the simulation. These variables included parameters characterizing the biomechanics and sensory capabilities of the predator and prey as well as parameters relating to the complexity of the surrounding habitat. The names of these variables, which are chosen to describe what they represent, include:

- prey-max-velocity
- prey-agility
- prey-acceleration
- prey-deceleration
- prey-vision-distance
- prey-vision-angle
- time-to-turn
- time-to-return-to-foraging
- time-spent-circling
- predator-max-velocity
- predator-agility
- predator-acceleration
- predator-deceleration
- predator-vision-distance
- predator-vision-angle
- time-to-give-up
- proportion-obstacles
- obstacle-radius
- obstacle-radius-range
- obstacle-sensitivity-for-prey
- obstacle-sensitivity-for-predators
- safe-zone-attractiveness
- number-of-safe-zones
- number-of-target-patches

All these variables are continuous except for the number of safe zones and the number of target patches. The initial experimental design for this simulation model was a Latin hypercube sampling using the Latin Hypercube Sampling function (lhs) from the Treed Gaussian Process Model Package (tgp) in R. Each simulation with a specific set of initial conditions was run 10 times with different seeds. The output data of these models included the 24 input variables and five output variables—whether the prey survived the run (prey-win), whether the predator succeeded in attacking the prey during the run (predator-win), the length of the run (time), a measure of the tortuosity of the prey trajectory (prey-curviness), and a measure of the tortuosity of the predator trajectory (predator-curviness). Since prey-win and predator-win are simply negations of each other, only prey-win was investigated.

METHODS

Since prey wins and predator wins are two different classes to be investigated, logistic regression, specifically binomial regression, was the main model method used to conduct sensitivity analysis on these factors. Logistic regression takes the form of the following equation $f(\boldsymbol{x}) = \frac{1}{1+\exp(-\boldsymbol{x}^T\boldsymbol{\beta})}$ where $\boldsymbol{x}$ is the vector of 24 input variables that establish the initial conditions, $f(\boldsymbol{x}) = 0$ when the **predator wins**, and $f(\boldsymbol{x}) = 1$ when the **prey wins**. The data included 160,037 predator wins and 200,960 prey wins. In both modelling methodologies, the data was separated into training and testing data, with 80% of the data devoted to training and 20% devoted to testing. The analysis was run in the statistical software package R on an Intel(R) Core(TM) i7-4650 CPU @ 1.70GHz 2.30 GHZ processor with 8.00 GB of RAM.

Traditional Modelling Methodology

Three initial models were developed using binomial regression. The first model (**model #1**) was a simple main-effects model that excluded all interactions. The second model (**model #2**) included the main effects and two-way interactions that included pairs of prey–predator pairs. For instance, the interaction between prey-max-velocity and predator-max-velocity ($prey.max.velocity \times predator.max.velocity$) was added to this model. Finally, the third model (**model #3**) included main effects and all two-way interactions. Unfortunately, due to time constraints, this model was not able to be fully analyzed using traditional variable selection. It was, however, fully analyzed using a regression method called least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996). Each initial model underwent backward, forward, and stepwise variable selection using the R function step(), and the processed model with the smallest Akaike Information Criterion (AIC) was chosen. Then, each model was diagnostically checked using normal probability plots of the deviance and Pearson residuals, plots of deviance and Pearson residuals versus the estimated probabilities, and histograms of the deviance and Pearson residuals. These plots are useful in checking the fit of the model as well as checking for possible outliers (Sarkar, Midi, & Rana, 2011) (Montgomery, Peck, & Vining, 2012). In addition to the residual plots previously mentioned, several numerical values were outputted. These include: the computational times for each type of variable selection; the deviance, null deviance, and their corresponding degrees of freedom; the coefficients and their corresponding statistics like standard error and p-value; and the confusion matrices from the testing data for different thresholds between 0.01 and 1.

The deviance, null deviance, and their corresponding degrees of freedom were based on the training data and were used to measure goodness-of-fit. This was accomplished by comparing the difference between the null deviance and the deviance and the chi-squared statistic with degrees of freedom as the difference between the null deviance and deviance degrees of freedom (Montgomery, Peck, & Vining, 2012). Specifically, let $D(\boldsymbol{\beta})$ be the deviance of the model, $D(\boldsymbol{\beta}_0)$ be the null deviance, and $\chi^2_{\alpha,r}$ be the chi-squared statistic with r degrees of freedom and a type I error rate of α. If

$D(\boldsymbol{\beta}) - D(\boldsymbol{\beta}_0) \geq \chi^2_{\alpha,r}$, then the model is statistically better than the null model. If $D(\boldsymbol{\beta}) - D(\boldsymbol{\beta}_0) < \chi^2_{\alpha,r}$, then the model is no better than the null model.

After developing each model, sensitivity analysis was performed on the coefficients to determine the effect of changing the simulation variables have on changing the prey/predator success. This works well in the traditional modelling methodology since the model coefficients have errors associated with them. Thus, using the odds ratio ($\hat{O}_R$), the estimated increase in the probability of prey success (or the decrease of predator success) can have bounds associated for each coefficient for each variable. For instance, let $\beta_i^{(LB)}$ and $\beta_i^{(UB)}$ be the lower bound and upper bound respectively for a given coefficient β_i with a desired type-I error rate (say $\alpha = 0.05$). Then, using the odds ratio, $\exp\left(\beta_i^{(LB)}\right) \leq \hat{O}_R \leq \exp\left(\beta_i^{(UB)}\right)$ represents the 95% confidence interval for an estimated increase in the probability of prey success associated with a one-unit increase in the value of x_i (Montgomery, Peck, & Vining, 2012). If $\hat{O}_R > 1$, then x_i has a positive effect on prey success (or negative effect on predator success). If $\hat{O}_R < 1$, then x_i has a negative effect on prey success (or positive effect on predator success).

Finally, receiving operating characteristic (ROC) curves were plotted for each model using the data collected from the confusion matrices generated from the testing data. ROC curves are graphical plots that illustrate the diagnostic ability of a binary classifier at different discrimination thresholds. They specifically plot the false positive rates on the horizontal axis and the true positive rates on the vertical axis based on various thresholds between 0 and 1. A threshold of 0.5 would be used as the default threshold for accuracy of the model. For instance, if a data point has a predicted probability that is greater than 0.5, it is sorted as a prey win, while if it is less than 0.5, it is sorted as a predator win. In addition to plotting the ROC curves, the area under the curve (AUC) was calculated for each of these plots. The AUC was then used to compare models with each other. Despite negative criticisms of the AUC as a metric for model comparison (Hanczar, et al., 2010) (Hand, 2009) (Lobo, Jimenez-Valverde, & Real, 2007), the AUC has been vindicated as a measure of aggregated classification performance in terms of a uniform rate distribution (Ferri, Hernandez-Orallo, & Flach, 2011).

New Modelling Methodology (LASSO)

Least absolute shrinkage and selection operator, or LASSO, is a regression technique that performs both variable selection and regularization to improve the prediction accuracy and interpretability of the model produced. Like other regularization methods, LASSO adds a constraint to penalize adding coefficients to the model. This in turn decreases variance drastically by increasing the bias exploiting the bias-variance trade-off. In general, regularization constraints are usually expressed as $\left||\boldsymbol{\beta}|\right|_k \leq t$ where $\left||\boldsymbol{\beta}|\right|_k = \left(\sum_{i=1}^{p} \beta_i^k\right)^{1/k}$, $k \geq 0$ and $t \geq 0$. When $k \geq 1$, the constraint region is convex and thus computationally efficient. In addition, when $0 \leq k \leq 1$, the concavity of the feasible region ensures that optimal solutions will activate boundary constraints; in other words, some elements of the solutions will be zero, which is useful for model selection (Fonti & Belitster, 2017). LASSO uses ℓ_1-regularization, thus making it computationally efficient and allowing it to perform variable selection.

The same three models were run using LASSO regression just like the traditional modelling methodology. Before each model was run, the regularization parameter lambda (λ) was found by performing cross-validation over different values of lambda. Then, the best[1] lambda was chosen based on the cross-validation error. Then, all the residual plots used in the tradition methodology were also used on the LASSO regression models. In addition to the residual plots, the same numerical outputs were calculated and outputted except for the coefficient statistics like standard error and p-value. This is because there is no consensus on standard error or confidence intervals for LASSO coefficients (Kyung, Gill, Ghosh, & Casella, 2010). Even if there was a consensus, standard error and confidence intervals for LASSO coefficients are misleading since LASSO and other ℓ_k-regularization techniques are strongly biased (Goeman, Meijer, & Chaturvedi, 2016). Despite this, the odds ratio was still performed on the selected coefficients to examine their effect on prey success. The ROC curves for each model were also plotted, and the area under the curve (AUC) was also calculated.

Comparison between Modelling Methodologies

Upon reflection and further research of the traditional modelling methods, research has found that the variable selection methods of backward, forward, and stepwise selection are biased (Wilkinson & Dallal, 1981), may have incorrect degrees of freedom (Hurvich & Tsai, 1990), and are prone to over-simplification of the real models of the data (Roecker, 1991). As a result, LASSO regression applied to logistic regression

[1] The glmnet package offers two lambda values: the lambda with the smallest cross-validation error and the lambda that is one standard error away larger from the lambda with the smallest cross-validation error. This thesis explores both lambdas.

was used for variable selection. For this thesis, the package glmnet in R was used, which optimizes the following objective for logistic regression:

$$\min_{\boldsymbol{\beta}}\left\{-\frac{1}{N}\sum_{i=1}^{N}(y_i^T x_i^T \boldsymbol{\beta} - \ln(1 + exp(x_i^T \boldsymbol{\beta}))) + \lambda\left|\left|\boldsymbol{\beta}\right|\right|_1\right\} \tag{1}$$

Unfortunately, unlike ridge regression, LASSO regression does not have a closed form solution. As a result, optimization algorithms must be employed to find the minimizing solution. In the case of glmnet, cyclical coordinated descent is used, which has a complexity of $O(np)$ (Gordan & Tibshirani, 2015).

There are many advantages that LASSO has compared to backward, forward, or stepwise selection. One major advantage LASSO has over the various stepwise regressions is that it performs covariate selection as well as reduces overfitting by shrinking large regression coefficients. Another advantage is that LASSO optimized with cyclical coordinated descent is computationally more efficient with a time complexity of $O(np)$ (Gordan & Tibshirani, 2015) versus $O(np^2)$ for the variable selection techniques mentioned earlier (Landy, 2017). Finally, LASSO, like other regularization methods, improves prediction error by increasing bias, while backward, forward, or stepwise selection has no guarantee of improving prediction error. Despite these advantages that LASSO has over backward, forward, or stepwise selection, one large downside with LASSO regularization is that there is no consensus on standard error or confidence intervals for LASSO coefficients (Kyung, Gill, Ghosh, & Casella, 2010). As a result, it is difficult to estimate the range of influence a variable has on outcome data, in this case prey/predator success. This makes sensitivity analysis difficult, though not impossible. One method would be bootstrapping techniques on the data and estimate the error of the coefficients (Goeman, Meijer, & Chaturvedi, 2016). However, for the purposes of this thesis, the odds ratio of the estimated coefficients for LASSO regression were sufficient.

Both modelling methods can focus on changing the selected variables for future simulations. The remaining variables can be made into fixed variables or simply eliminated from the simulation entirely. Although both modelling techniques help which variables to focus on for future simulations, they do not help narrow their range of values. As a result, alternative models and methods were explored for narrowing variable ranges for future simulations. One method that this thesis explored was classification trees.

Variable Range Reduction (Classification Tree)

Classification trees can be used for variance reduction for future simulations in two ways. On the one hand, the classification tree could also help set variables involved in splitting rules to be fixed variables. For instance, if a splitting rule was $predator.vision.distance \geq 9.747$, this variable can be fixed at 9.747 to balance the representation of outcomes in either branch following the split. This would allow other variables to be run with more granularity for the same simulation budget. This also reduces the overall variance since the variance of these variables are essentially eliminated. On the other hand, the classification tree would sort the data into each terminating node, which provides a subset of the data. If, for instance, the classification error of a terminating node seems large, this subset can be simulated for further investigation. Again, because each subset has reduced variable ranges based on the splitting rules, the simulations run on these subsets also have reduced overall variance.

However, classification trees have many parameters that allow an infinite number of trees to be made. For instance, the R package rpart, which this thesis utilized, has three main parameters: a split type parameter which includes using either an information or the Gini index (denoted as *split*), a minimum number of instances parameter that is required for splitting (denoted as *minsplit*), and a complexity parameter that penalizes the number of terminal nodes (denoted as *cp*). The complexity parameter can be summarized as the regularized cost $C_\alpha(T) = C(T) + \alpha|T|$, where $C(T)$ is the cost of the tree, α is the complexity parameter, and $|T|$ is the number of terminal nodes (Therneau & Atkinson, 1997). A method was developed to choose the best classification tree by choosing the optimal set of parameters. This was accomplished by performing cross-validation on a wide variety of these parameters. Specifically, 25 repeated measures of unique parameters *split*, *minsplit*, and *cp* were run under 10-fold cross-validation, and the accuracy was measured each instance. After performing an analysis of variance (ANOVA) on the data, Tukey's honest significant difference (HSD) test was performed to find the set of best classification trees based on their accuracy. The tree with the highest accuracy of this set was chosen.

RESULTS AND ANALYSIS

Outliers and Diagnostic Checks

The main assumptions the diagnostic plots verify are normality of the residuals in the Q-Q normal probability plots and random distribution in the residuals versus estimated probabilities plots. Both plots are also used to find outliers. For the Q-Q normal probability plots, the Pearson residuals plot appears more normal than the deviance residuals plot across all the models. This is because the residuals plotted against the normal quantile values appears more linear for the Pearson residuals compared to the deviance residuals, as shown in the example Q-Q plots for the main effects model with traditional variable selection methods (see Figure 2 below).

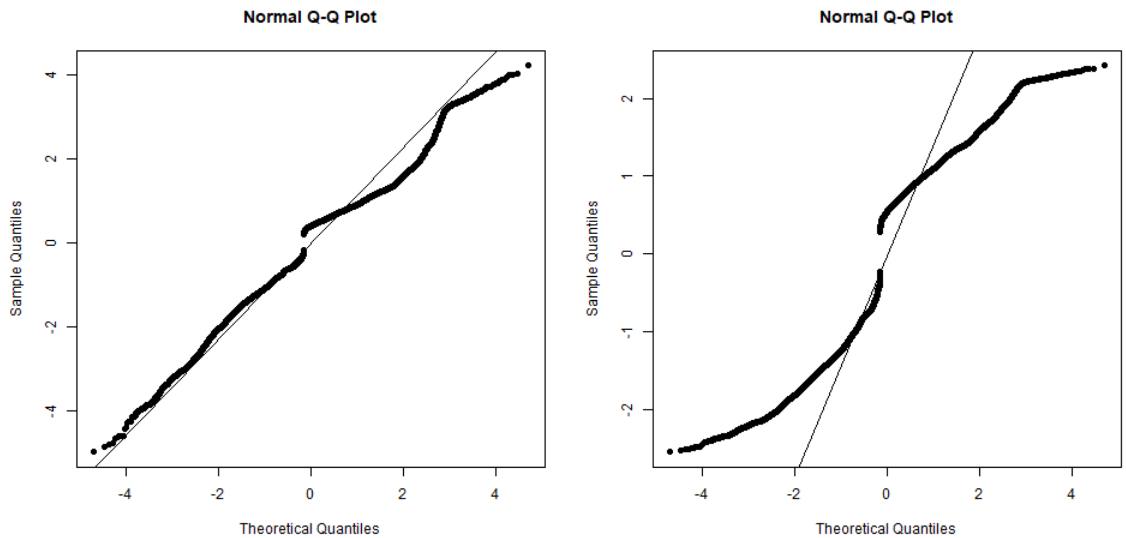

Figure 2. Q-Q Normality Plots for Pearson and Deviance Residuals Respectively

The deviance residuals in the Q-Q normal probability plots appears to be shaped as an inverted S-curve, suggesting a distribution with short tails. Despite this discrepancy in the deviance residuals, the normality assumption does not seem to be totally violated. This is because the number of instances is very large (training set had 288,797 instances), and thus the central limit theorem would apply in this situation.

For the residuals versus estimated probabilities plots, according to Sarkar, Midi, and Rana (2011), if a logistic regression model is correct and contains no outliers, “the plot of the residuals against the estimated logistic probability or linear predictor should result approximately in a horizontal line with zero intercept” (Sarkar, Midi, & Rana, 2011). The Pearson and deviance residuals versus estimated probabilities plots for all the models appear to have these characteristics, as illustrated in the example estimated probabilities plots for the main effects model with traditional variable selection methods in Figure 3.

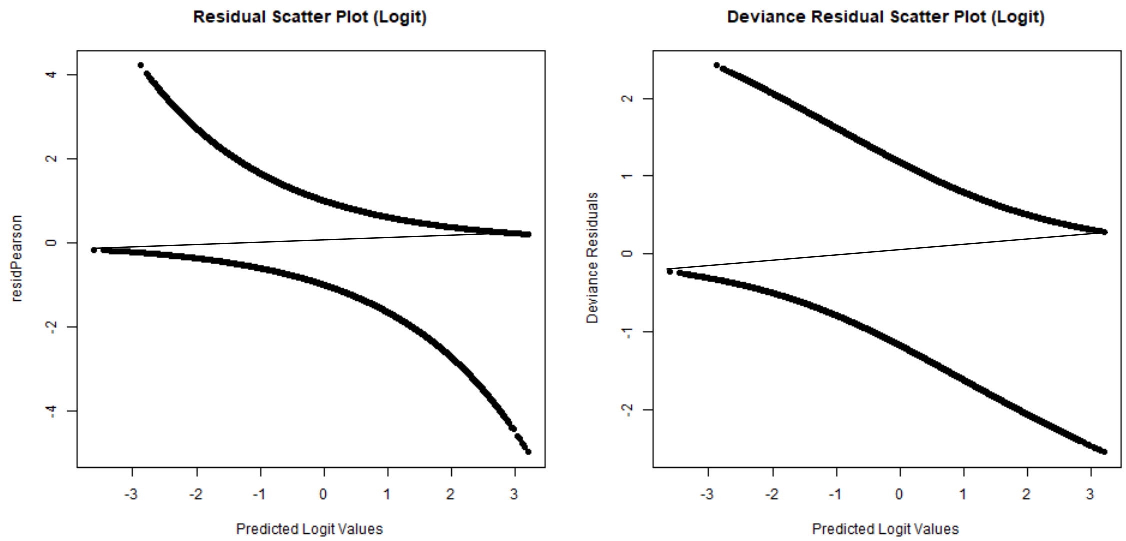

Figure 3. Estimated Probabilities Plots Versus Pearson and Deviance Residuals Respectively

Finally, the histograms of the residuals did not contain a long tail in one direction, which would indicate skewness, nor do they contain a bar that was far away from the other bars, which would indicate an outlier. The following illustrates these distinctions for the main effects model with traditional variable selection methods in Figure 4.

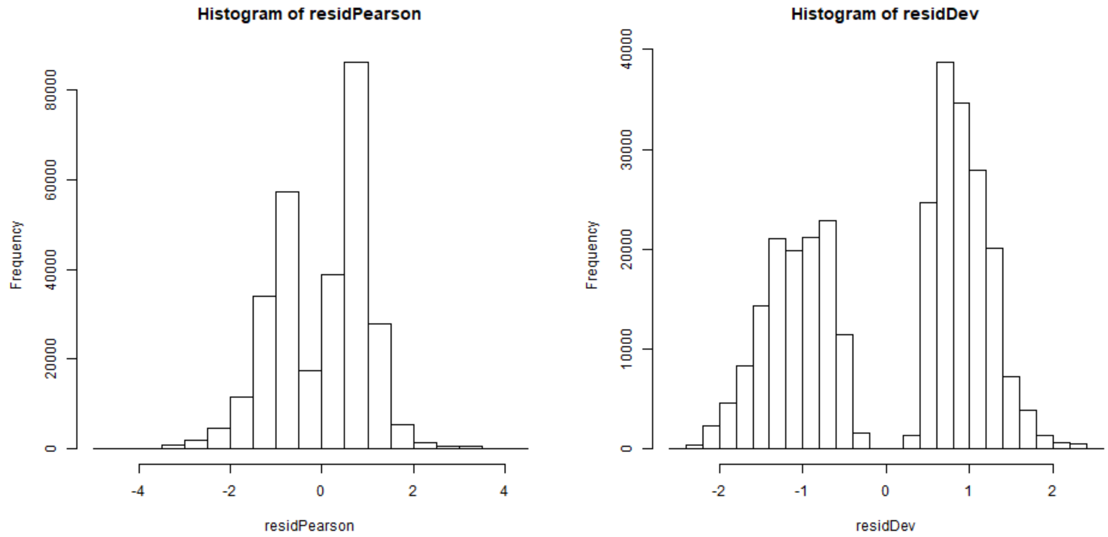

Figure 4. Residual Histograms (Pearson and Deviance Residuals Respectively) for Main Effects Model

For further details concerning the other diagnostic plots, see the APPENDIX A for the normality plots, APPENDIX B for the predicted probability plots, and APPENDIX C for the histograms.

Computation Times

Table 1 summarizes the computation time for each type of variable selection:

Table 1: Table of Computational Times for Each Variable Selection Method and Model Type

		Computation Times (Seconds)		
Variable Selection		**Model #1**	**Model #2**	**Model #3**
Traditional Methods	**Backward**	273.2878	502.0415	*NA*
	Forward	612.3457	942.9345	*NA*
	Stepwise	1423.0500	2014.0444	*NA*
Lasso (1se)	**LassoCV**	184.1133	310.8239	3276.0220
	LassoFinal	2.6939	4.8715	228.3983
Lasso (Min)	**LassoCV**	171.1997	314.5494	2980.3747
	LassoFinal	3.4525	5.9412	362.0748

LassoCV refers to running 10-fold cross-validation on LASSO regression to find the regularization value, and **LassoFinal** refers to taking the best regularization value and rerunning LASSO regression on the full dataset. **Lasso (Min)** refers to the LASSO regression that chooses the regularization value with the smallest validation error, and **Lasso (1se)** refers to the LASSO regression that chooses the regularization value one standard error away from the regularization value previously mentioned. **Backward**, **Forward**, and **Stepwise** refer to the traditional variable selection methods.

As Table 1 illustrates, the more variables introduced to the model (the number of parameters, p, increases for each model), the longer it takes to find the final model. The data also confirm the theoretical result that the traditional methods, which have a polynomial time complexity (in terms of p), are much slower than LASSO regression, which has a linear time complexity, despite cross-validation being run with LASSO to find the optimal regularization value. One instance that highlights this fact is the computational times for **model #2** in which LASSO took about five minutes to run while the traditional method took almost an hour.

Goodness of Fit via Deviance

Table 2 shows the deviances and degrees of freedom for each type of variable selection as well as the null deviance, which is the same for all models.

Table 2: Table of Deviances and Their Degrees of Freedom for the Null Model and Each Variable Selection Method and Model Type

Null Deviance	df
396634.1590	288796

	Model #1		Model #2		Model #3	
Variable Selection	**Deviance**	**df**	**Deviance**	**df**	**Deviance**	**df**
Traditional Method	334069.0204	288774	333790.8480	288769	*NA*	*NA*
Lasso (1se)	334511.9208	288771	334209.8172	288764	319170.3373	288495
Lasso (Min)	334090.6003	288771	333829.4779	288764	318760.9575	288495

All the models compared to the null deviance model were all significant, meaning that each model performed better than the null model (p-values were all $\ll 0.0001$). This further confirms that the above modelling techniques result in models with good fits.

Significant Variables via Coefficients and Odds Ratios

For the dependent variable $prey.win$, $f(\boldsymbol{x}) = 0$ when the predator wins and $f(\boldsymbol{x}) = 1$ when the prey wins. Thus, the smallest coefficients for each model correspond to contributing to predator success, and the largest coefficients contribute to prey success. Also, the odds ratio can be interpreted as the estimated increase in the probability of prey success associated with a one-unit increase in the corresponding variable. Table 3 shows the smallest and largest coefficients for each type of variable selection in each model as well as their odds ratio.

Table 3: Table of Significant Values for Predator and Prey Success

	Variable Selection	Aids:	Variable	Coefficient	Odds Ratio
Model #1	**Traditional Methods**	**Predator**	**proportion.obstacles**	-3.5177	0.0297
		Prey	**number.of.safe.zones**	0.7129	2.0398
	Lasso (1se)	**Predator**	**proportion.obstacles**	-2.9859	0.0505
		Prey	**number.of.safe.zones**	0.6894	1.9925
	Lasso (Min)	**Predator**	**proportion.obstacles**	-3.3734	0.0343
		Prey	**number.of.safe.zones**	0.7078	2.0295
Model #2	**Traditional Methods**	**Predator**	**proportion.obstacles**	-3.5763	0.0280
		Prey	**obstacle.sensitivity.for.predators**	0.9794	2.6629
	Lasso (1se)	**Predator**	**proportion.obstacles**	-3.1760	0.0418
		Prey	**number.of.safe.zones**	0.6978	2.0093
	Lasso (Min)	**Predator**	**proportion.obstacles**	-3.5022	0.0301
		Prey	**obstacle.sensitivity.for.predators**	0.8058	2.2384
Model #3	**Traditional Methods**	**Predator**	NA	NA	NA
		Prey	NA	NA	NA
	Lasso (1se)	**Predator**	**proportion.obstacles**	-6.1638	0.0021
		Prey	**proportion.obstacles: obstacle.radius**	3.7708	43.4129
	Lasso (Min)	**Predator**	**proportion.obstacles**	-7.0051	0.0009
		Prey	**proportion.obstacles: obstacle.radius**	3.9172	50.2591

Based on the above results, environmental variables like number of safe zones, proportion of the obstacles, and obstacle radius had the largest effects on prey and predator success. Even the predator-specific variable $obstacle.sensitivity.for.predators$ was still environmentally linked. If other variables besides the environmental ones are to be further examined, the environmental variables could be fixed for future simulations. See APPENDIX E for details of the other coefficients.

Model Comparison via ROC Curves, AUC, and Accuracy

Table 4 contains the AUC for the ROC curve as well as the accuracy on the testing data ($threshold = 0.5$) for all models.

Table 4: Table of AUC and Model Accuracy for Each Variable Selection Method and Model Type

	Model #1		**Model #2**		**Model #3**	
Variable Selection	**AUC**	**Accuracy**	**AUC**	**Accuracy**	**AUC**	**Accuracy**
Traditional Methods	0.7566	0.6907	0.7568	0.6904	*NA*	*NA*
Lasso (1se)	0.7560	0.6894	0.7654	0.6902	0.7795	0.7083
Lasso (Min)	0.7566	0.6900	0.7569	0.6906	0.7798	0.7093

As the above table shows, since all the AUCs are greater than 0.5, indicating that the models performed better than random. This is also evident in the ROC curves, as illustrated in the ROC curve for the main effects model for traditional methods in Figure 5.

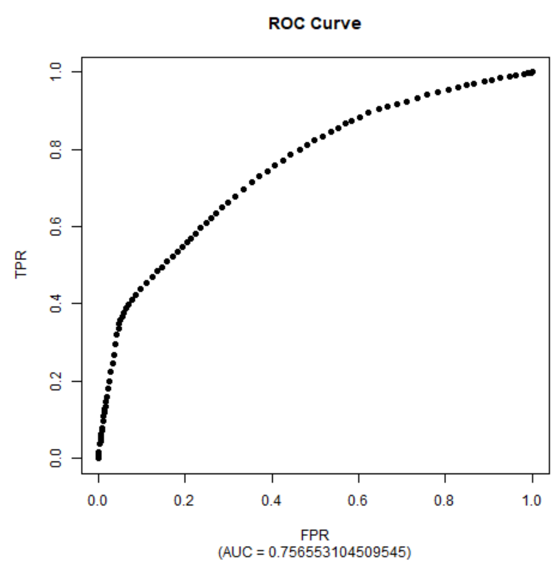

Figure 5. ROC Curve for Traditional Selection Methods on Main Effects Model

The results in Table 4 show that there is not much discrepancy between the traditional methods and LASSO regression within each model. Between models, it appears that **model #3** performs slightly better than **model #1** and **model #2** in terms model accuracy, though it can be difficult to determine since the accuracy is only one sample. This could be mitigated by running cross-validation, though this would be impractical for the traditional methods due to the large computational time complexity. For further details concerning the other ROC curves, see APPENDIX D for the remaining ROC curves.

<u>Variance Reduction via Classification Tree</u>

The best tree was found to use the information index as its split type, have a complexity parameter of 0.01, and the minimum number of instances required for splitting was negligible, so this parameter was set to 0 (the smallest number of instances for a terminating node was 22,360, approximately 6% of the total number of instances). Figure 6 shows the final decision tree run on the whole data set based on ANOVA and Tukey's HSD test.

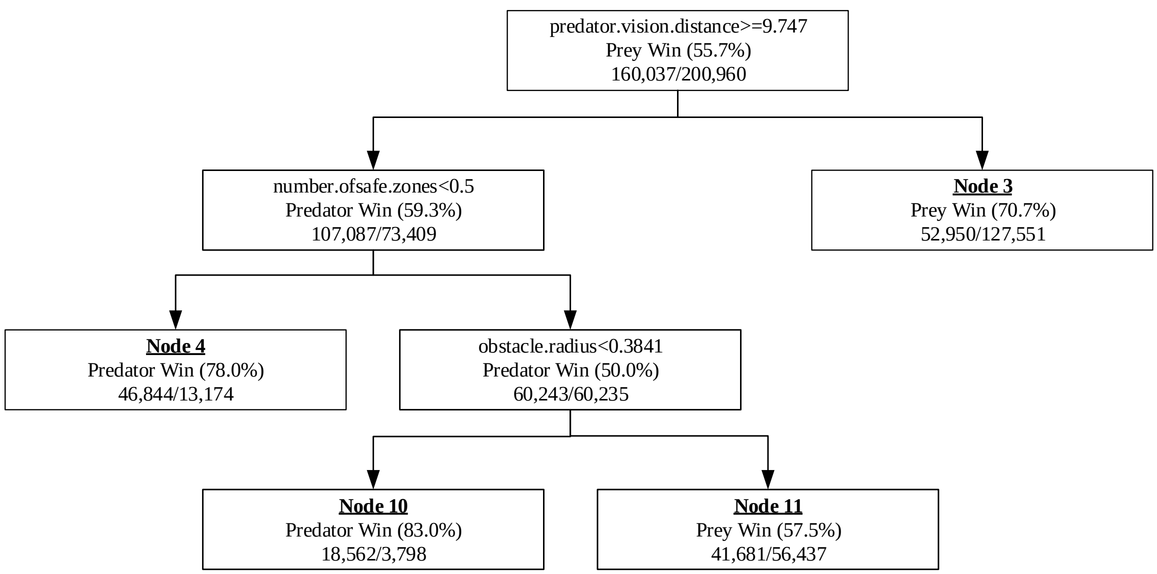

Figure 6. Final Decision Tree Using ANOVA and Tukey's HSD Test

As the final model suggests, the significant variables seemed to be predator vision distance, number of safe zones, and obstacle radius. From a qualitative perspective, the decision tree confirms intuitions about single predation. As mentioned before, the specific split rules could help set variables to be fixed variables for future simulations. Based on the final decision tree, these variables would be $obstacle\ radius =\ 0.3841$

and $predator\ vision\ distance\ =\ 9.747$. Because the number of safe zones is a discrete value, half of the simulation runs would have the number of safe zones fixed as 0 and the other half fixed as 1. Making these variables fixed would help either decrease the number of necessary simulations or increase the granularity of other variables of interest.

Table 5 shows prey and predator success for each node as well as the total number of instances in each terminating node.

Table 5: Table of Number of Prey/Predator Successes for Each Terminating Node

Node	Prey Wins	Predator Wins	Total
3	127,551	52,950	180,501
4	13,174	46,844	60,018
10	3,798	18,562	22,360
11	56,437	41,681	98,118

Of these terminating nodes, it appears that node 11 has a roughly equal split between prey wins ($\sim 58\%$) and predator wins ($\sim 42\%$), and so it may be prudent to run further simulations on the conditions for this node to confirm that the slight prey preference in this scenario, where $predator\ vision\ distance \geq 9.747$, $number\ of\ safe\ zones > 0.5$, and $obstacle\ radius \geq 0.3841$. Although not calculated, confidence intervals for the proportion of prey and predator wins may be a better indicator for choosing which subset to run future simulations.

DISCUSSION AND CONCLUSION

Traditional variable-selection techniques like backward, forward, and stepwise selection have large time complexity in terms of the number of variables ($O(np^2)$). Unfortunately, backward, forward, and stepwise selection are the default methods for many scientific fields, including biology and ecology. Although these techniques are adequate for a small number of variables, their time complexity is detrimental for analyzing complex simulations that have many parameters. As this thesis illustrated, running a simple stepwise regression on a dataset with 24 variables that included all main effects and two-way interactions did not finish, taking over a week before deciding to end it. In contrast, the LASSO regularization regression technique only took about an hour to run cross-validation to find the optimal regularization value parameter, and less than five minutes to run that specific model. Despite the significant decrease in computational time, LASSO regression produced statistical models with false positive rates and true positive rates on par with those of traditional modelling methods (as characterized by ROC curve, AUC, and overall accuracy). Both the traditional variable selection methods and LASSO regression appear to have selected the same variables and produced similar coefficients. For this dataset, the variables that were selected were mostly environmental variables like number of safe zones, proportion of obstacles, and obstacle radius, suggesting that environment factors play a heavier role in predation than prey or predator traits.

In addition, classification trees were explored to provide the basis for an iterative process to run future simulations. They could be used in two ways: setting the splitting rules as fixed variables and allow other variables to be run with more granularity or examining the terminating leaves and run simulations based on the instances in these nodes. The analysis revealed that predator vision distance (split at 9.747), number of safe zones (split at 0.5), and obstacle radius (split at 0.3841) were important splitting variables.

Implications for Ecology and Ecological Simulations

As the above analysis indicates, environmental variables seem to have a large influence on prey and predator success. The goal of the simulation study was to demonstrate how biomechanical couplings between speed and agility within an animal (either as a prey or predator) could interact with consistent features of a habitat to modulate the success of a prey in evading a predator. As expected, the conventional characteristics used to predict predator advantage (i.e., top speed) are poor predictors of predator performance in a realistic complex environment. Designing a real-world experiment to demonstrate this would not only be difficult but would likely be constrained by a low sample size and a high variability across sample results for the same experimental conditions. Thus, there is a compelling case for using computer simulation to answer these questions, and the statistical methods described here have provided a valuable perspective on the large quantities of data from such simulation studies and have confirmed that environmental parameters do have a strong effect on prey success evading predators.

As computer simulation of realistic habitats becomes more common place in studies of biomechanics and landscape ecology, variable-selection methods and classification trees like the ones applied in this thesis can realistically be used to discover important relationships. Moreover, computationally enhanced conservation biology may be a new bioinformatics application area for more Industrial Engineers to pursue.

WORKS CITED

Ferri, C., Hernandez-Orallo, J., & Flach, P. (2011). A Coherent Interpretation of AUC as a Measure of Aggregated Classification Performance. *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, (pp. 657-664). Bellevue, WA, USA.

Fonti, V., & Belitster, E. (2017). Feature Selection using LASSO.

Goeman, J., Meijer, R., & Chaturvedi, N. (2016, May 27). *L1 and L2 Penalized Regression Models*. Retrieved from R Foundation for Statistical Computing: https://cran. r-project. org/web/packages/penalized/vignettes/penalized.pdf

Gordan, G., & Tibshirani, R. (2015). Coordinate descent. *Optimization, 10*(36), 725.

Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., & Dougherty, E. R. (2010). Small-sample precision of ROC-related estimates. *Bioinformatics, 26*(6), 822-830.

Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine learning, 77*(1), 103-123.

Hurvich, C. M., & Tsai, C.-L. (1990). The Impact of Model Selection on Inference in Linear Regression. *The American Statistician, 44*(3), 214-217.

Kyung, M., Gill, J., Ghosh, M., & Casella, G. (2010). Penalized Regression, Standard Errors, and Bayesian Lassos. *Bayesian Analysis, 5*(2), 369-412.

Landy, J. (2017). Stepwise regression for unsupervised learning. *arXiv preprint arXiv:1706.03265*.

Lobo, J. M., Jimenez-Valverde, A., & Real, R. (2007). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography, 17*(2), 145-151.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (Vol. 821). Hoboken, New Jersey: John Wiley & Sons.

Roecker, E. B. (1991). Prediction Error and Its Estimation for Subset-Selected Models. *Technometrics, 33*(4), 459-468.

Sarkar, S., Midi, H., & Rana, S. (2011). Detection of Outliers and Influential Observations in Binary Logistic Regression: An Emirical Study. *Journal of Applied Sciences, 11*(1), 26-35.

Therneau, T. M., & Atkinson, E. J. (1997). *An Introduction to Recursive Partitioning Using the RPART Routines.*

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.

Wilkinson, L., & Dallal, G. E. (1981). Tests of Significance in Forward Selection Regression with an F-to-Enter Stopping Rule. *Technometrics, 23*(4), 377-380.

Wilson, A. M., Hubel, T. Y., Wilshi, S. D., Lowe, J. C., Lorenc, M., Dewhirst, O. P., . . . West, T. (2018). Biomechanics of predator–prey arms race in lion, zebra, cheetah and impala. *Nature, 554*(7691), 183-188.

APPENDIX

APPENDIX A: Q-Q normality plots of the Pearson and deviance residuals

APPENDIX B: Predicted probabilities versus the Pearson and deviance residuals

APPENDIX C: Histograms of the Pearson and deviance residuals

APPENDIX D: ROC curves and cross-validation plots for LASSO

APPENDIX E: Table of significant coefficients

APPENDIX A

Q-Q NORMALITY PLOTS OF PEARSON AND DEVIANCE RESIDUALS

The following are the Q-Q normality plots of the Pearson and deviance residuals for each model using the various variable selection methods.

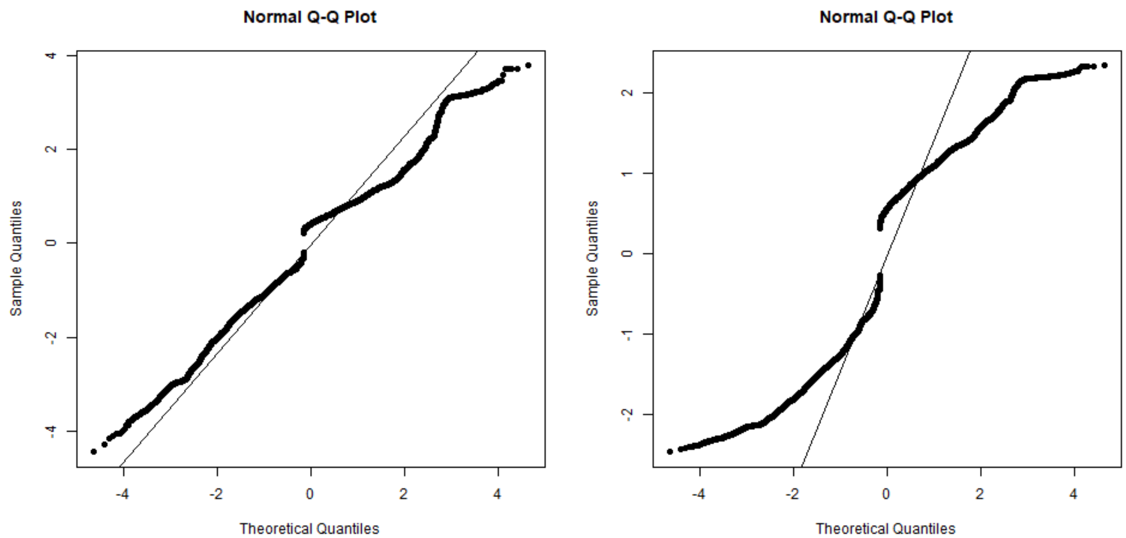

Figure 7. Q-Q Normality Plots for Pearson and Deviance Residuals Respectively for Main Effects Model Using Lasso (1se)

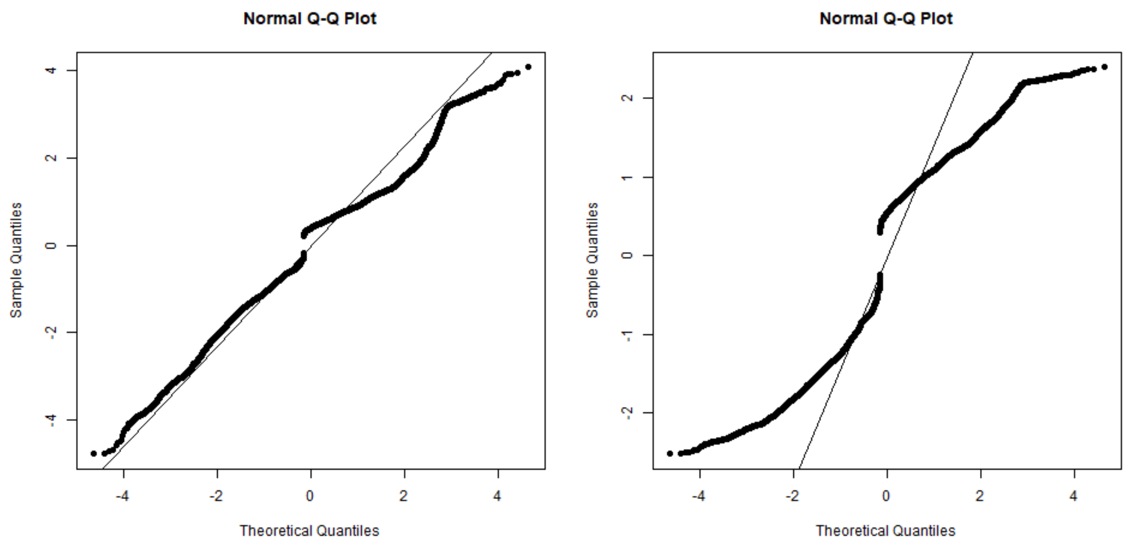

Figure 8. Q-Q Normality Plots for Pearson and Deviance Residuals Respectively for Main Effects Model Using Lasso (min)

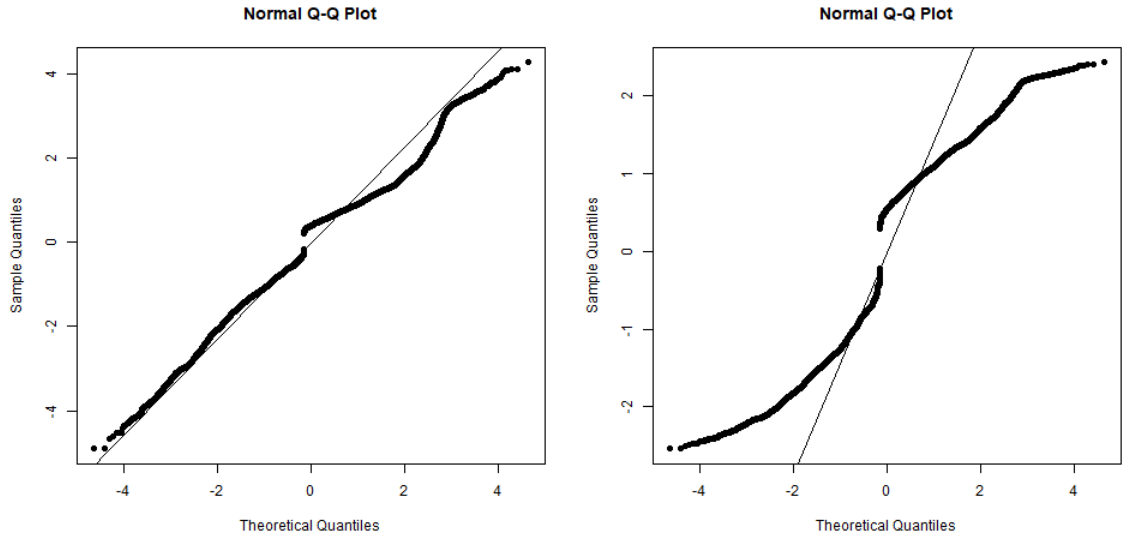

Figure 9. Q-Q Normality Plots for Pearson and Deviance Residuals Respectively for Main Effects and Prey-Predator Two-way Interactions Model Using Traditional Method

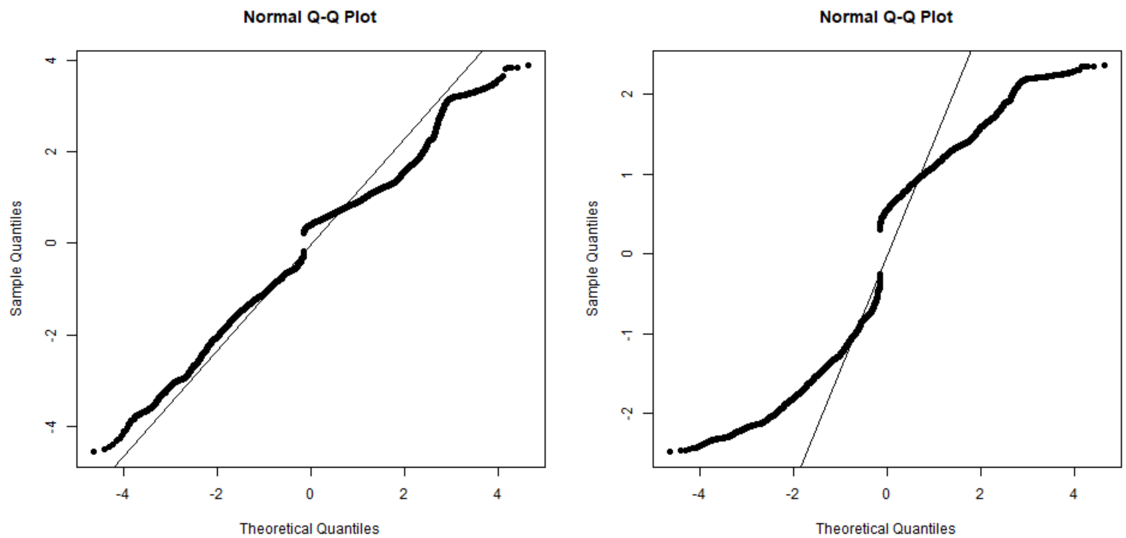

Figure 10. Q-Q Normality Plots for Pearson and Deviance Residuals Respectively for Main Effects and Prey-Predator Two-way Interactions Model Using Lasso (1se)

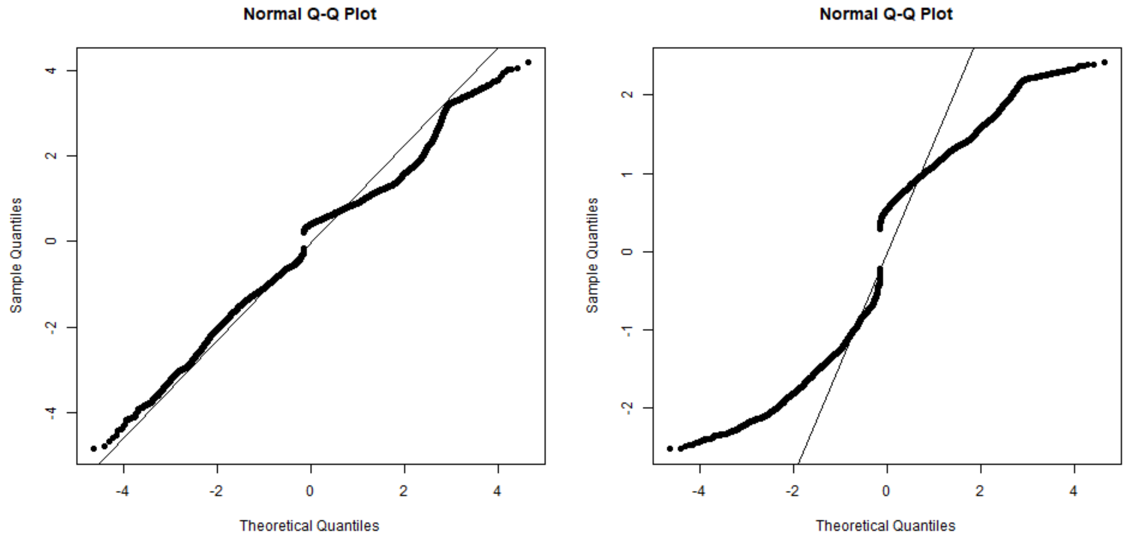

Figure 11. Q-Q Normality Plots for Pearson and Deviance Residuals Respectively for Main Effects and Prey-Predator Two-way Interactions Model Using Lasso (min)

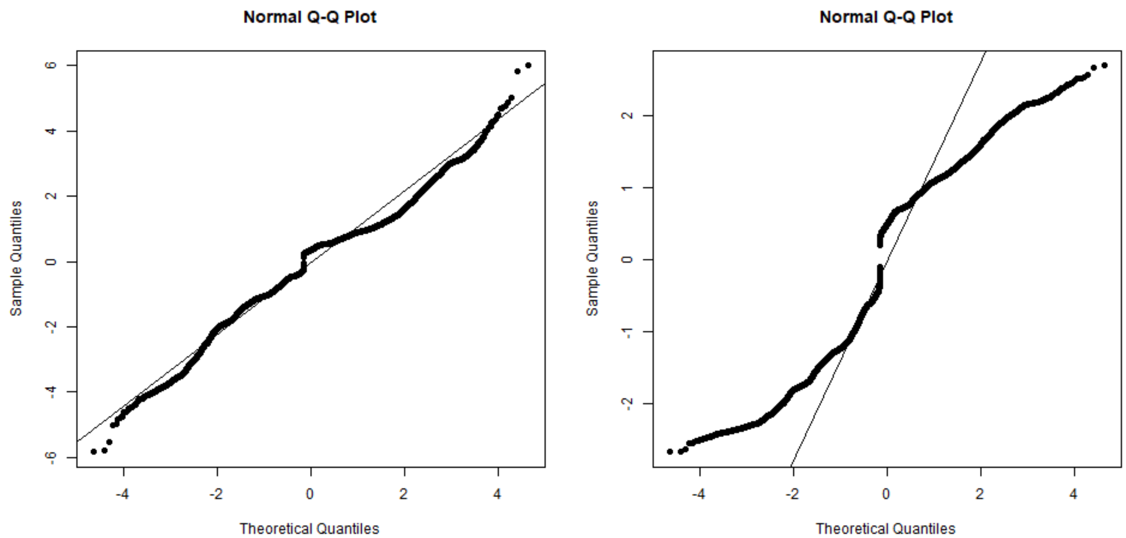

Figure 12. Q-Q Normality Plots for Pearson and Deviance Residuals Respectively for Main Effects and Full Two-way Interactions Model Using Lasso (1se)

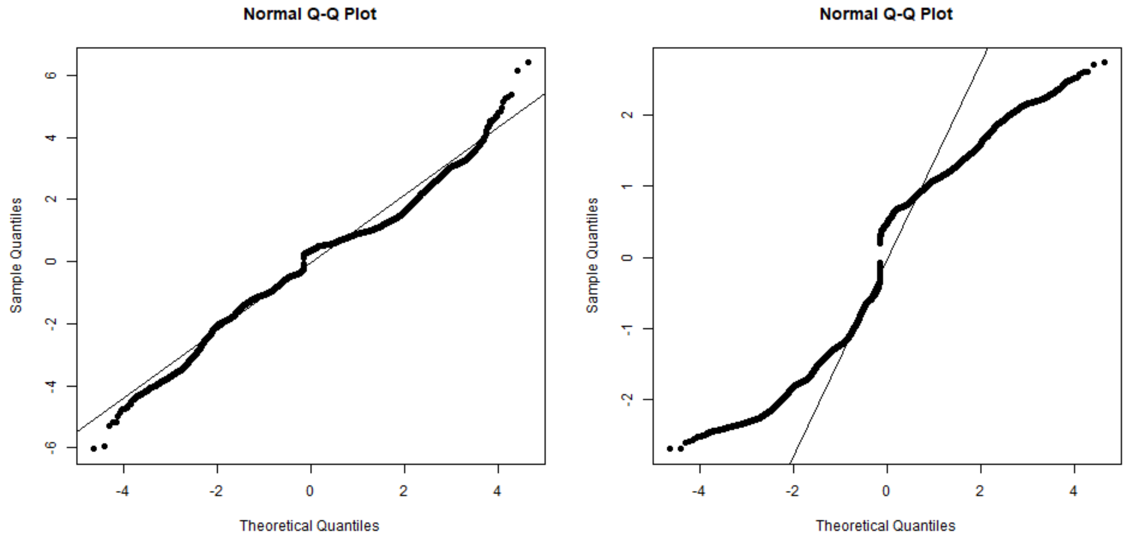

Figure 13. Q-Q Normality Plots for Pearson and Deviance Residuals Respectively for Main Effects and Full Two-way Interactions Model Using Lasso (min)

APPENDIX B

PREDICTED PROBABILITIES VERSUS PEARSON AND DEVIANCE RESIDUALS

The following are the predicted probabilities versus the Pearson and deviance residuals for each model using the various variable selection methods.

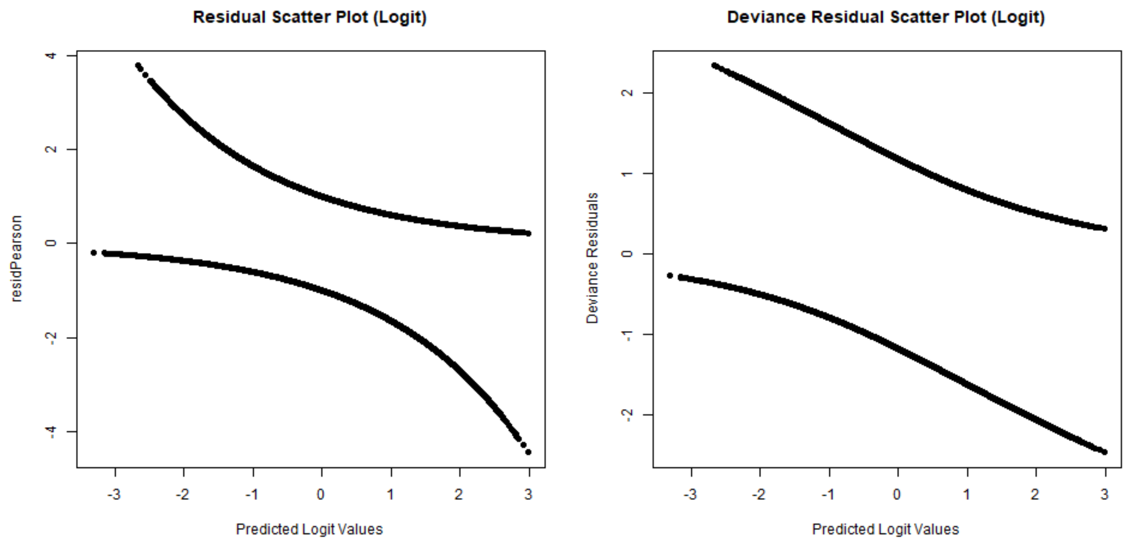

Figure 14. Pearson and Deviance Residuals (Respectively) Vs. Estimated Probabilities Plots for Main Effects Model Using Lasso (1se)

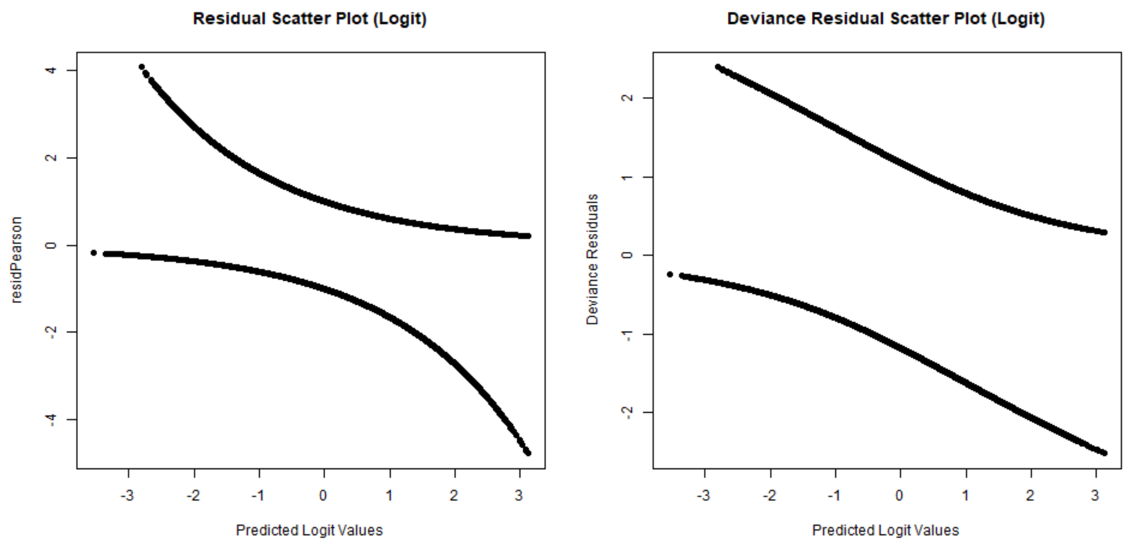

Figure 15. Pearson and Deviance Residuals (Respectively) Vs. Estimated Probabilities Plots for Main Effects Model

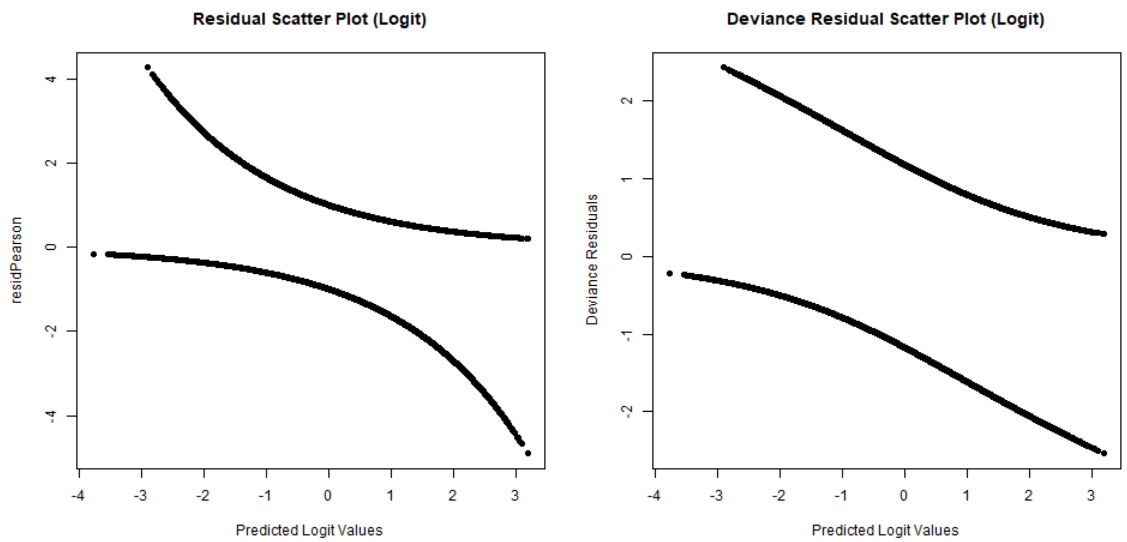

Figure 16. Pearson and Deviance Residuals (Respectively) Vs. Estimated Probabilities Plots for Main Effects and Prey-Predator Two-way Interactions Model Using Traditional Method

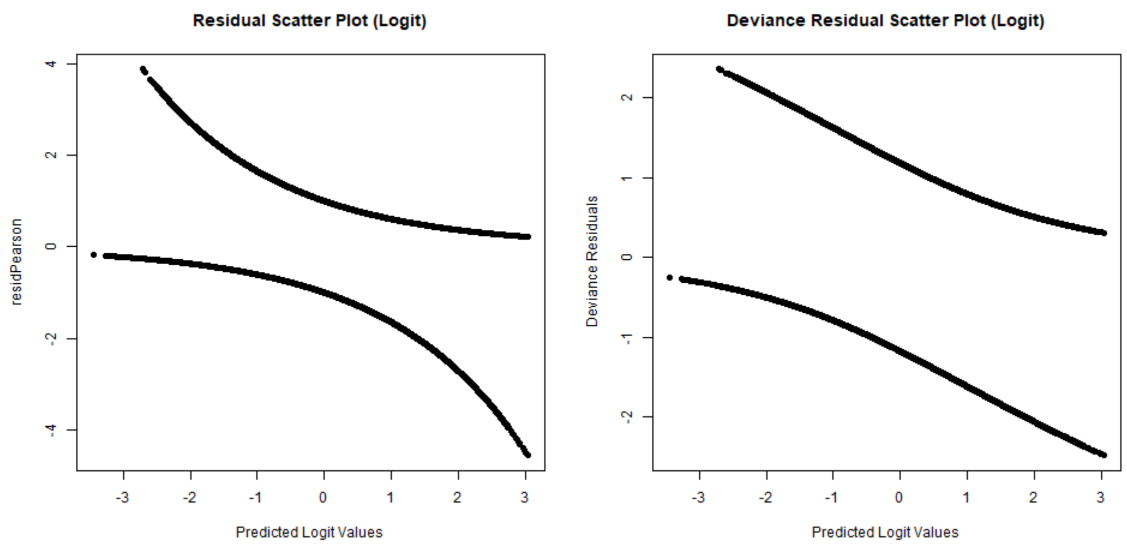

Figure 17. Pearson and Deviance Residuals (Respectively) Vs. Estimated Probabilities Plots for Main Effects and Prey-Predator Two-way Interactions Model Using Lasso (1se)

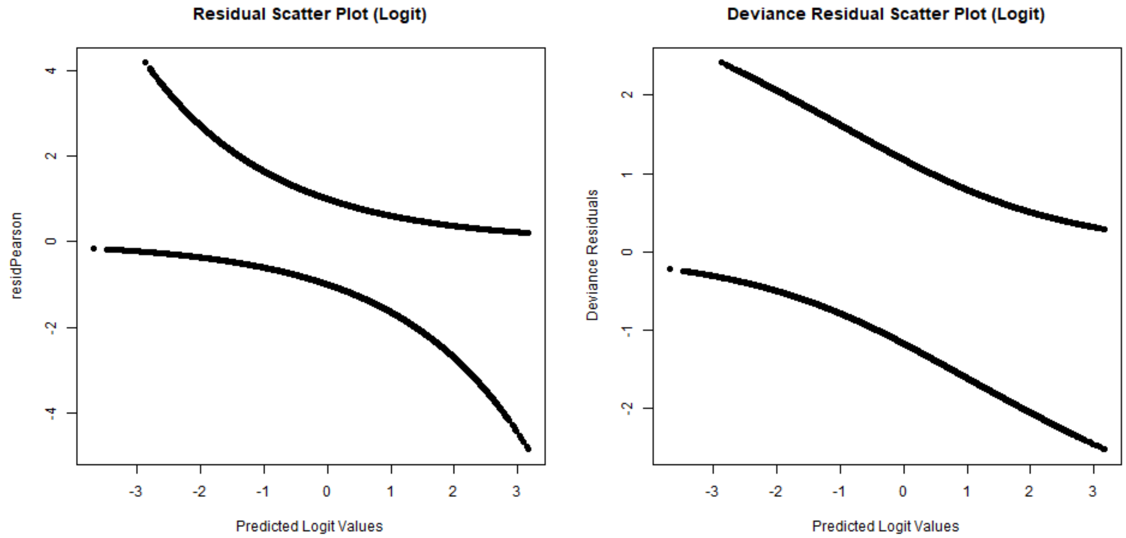

Figure 18. Pearson and Deviance Residuals (Respectively) Vs. Estimated Probabilities Plots for Main Effects and Prey-Predator Two-way Interactions Model Using Lasso (min)

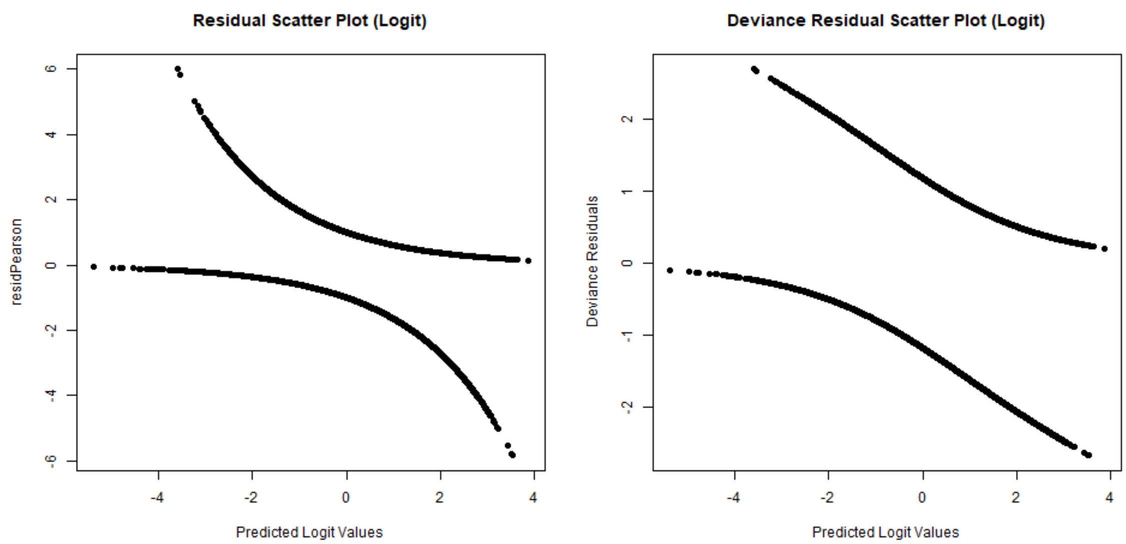

Figure 19. Pearson and Deviance Residuals (Respectively) Vs. Estimated Probabilities Plots for Main Effects and Full Two-way Interactions Model Using Lasso (1se)

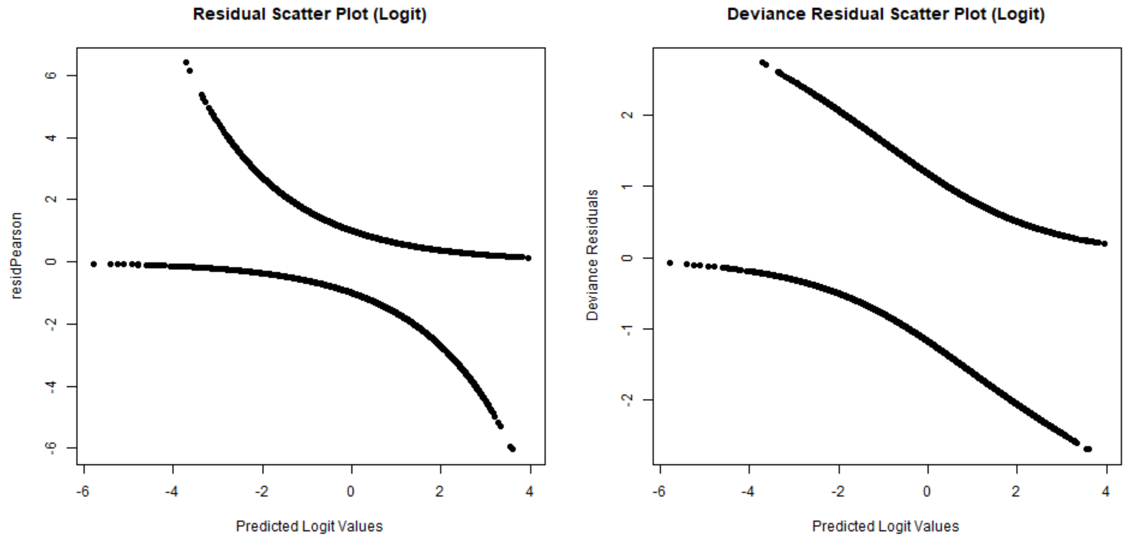

Figure 20. Pearson and Deviance Residuals (Respectively) Vs. Estimated Probabilities Plots for Main Effects and Full Two-way Interactions Model Using Lasso (min)

APPENDIX C

HISTOGRAMS OF PEARSON AND DEVIANCE RESIDUALS

The following are the histograms of the Pearson and deviance residuals for each model using the various variable selection methods.

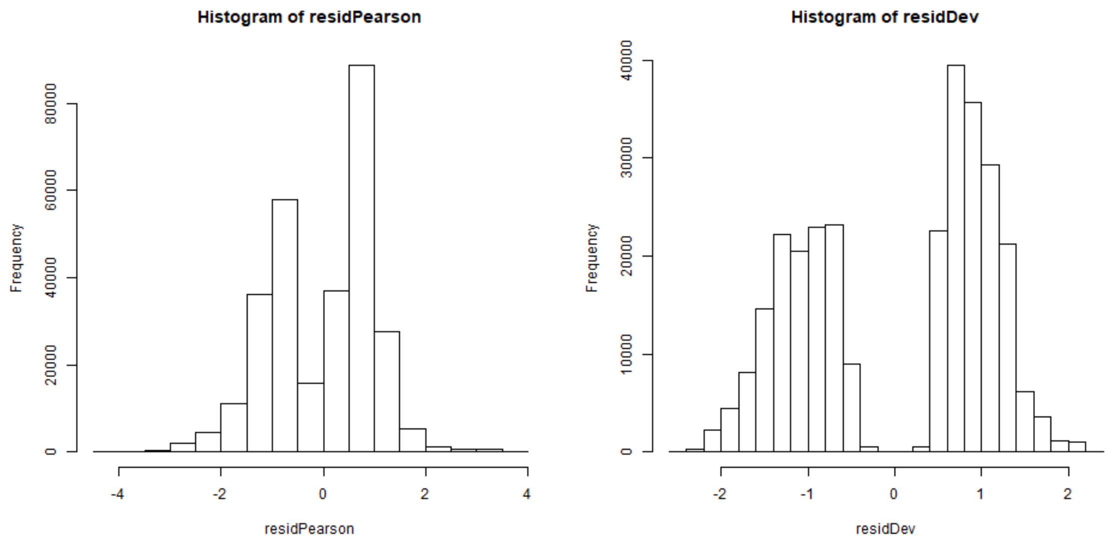

Figure 21. Residual Histograms (Pearson and Deviance Residuals Respectively) for Main Effects Model Using Lasso (1se)

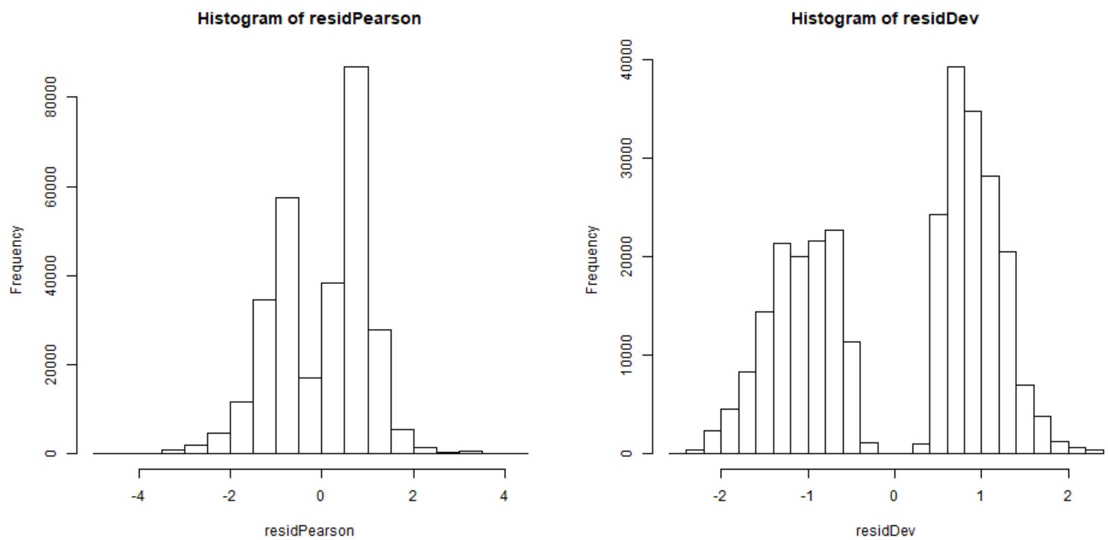

Figure 22. Residual Histograms (Pearson and Deviance Residuals Respectively) for Main Effects Model Using Lasso (min)

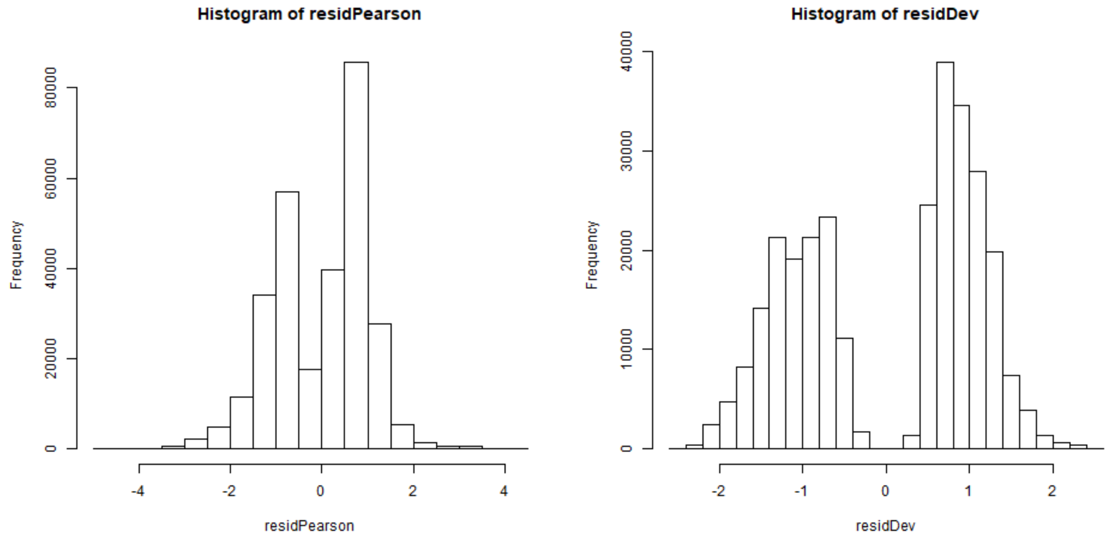

Figure 23. Residual Histograms (Pearson and Deviance Residuals Respectively) for Main Effects and Prey-Predator Two-way Interactions Model Using Traditional Method

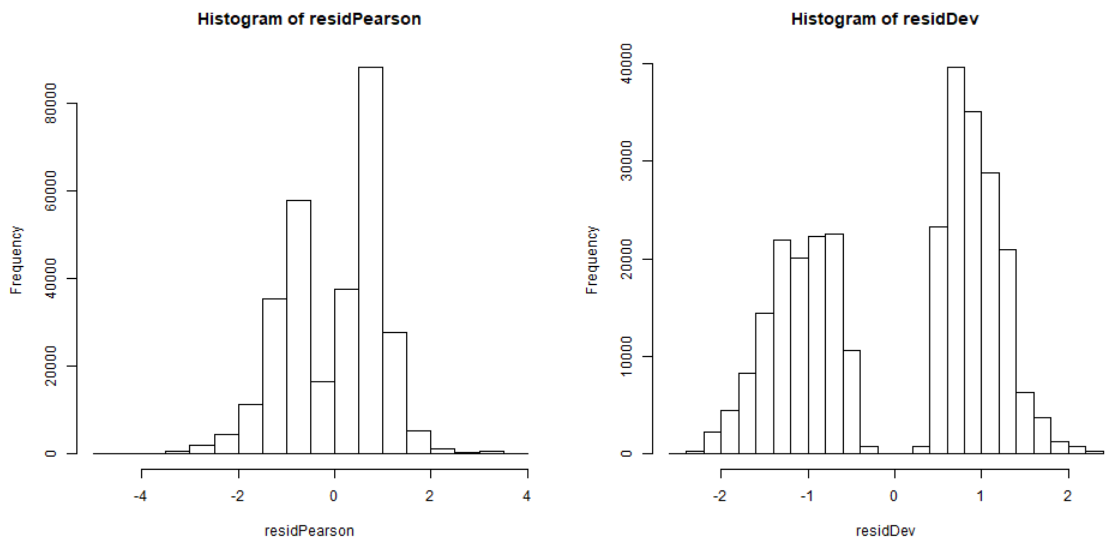

Figure 24. Residual Histograms (Pearson and Deviance Residuals Respectively) for Main Effects and Prey-Predator Two-way Interactions Model Using Lasso (1se)

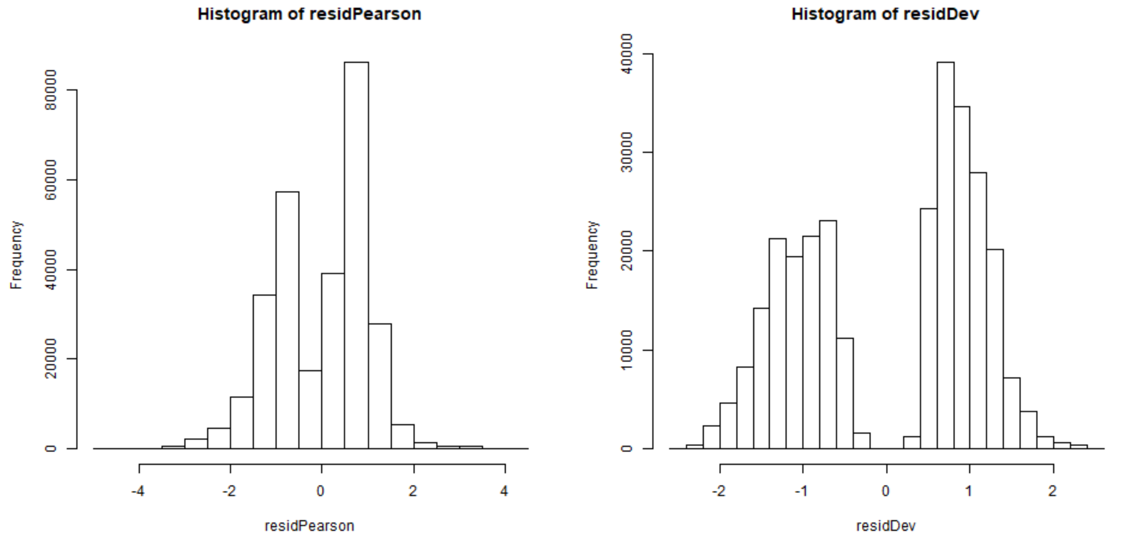

Figure 25. Residual Histograms (Pearson and Deviance Residuals Respectively) for Main Effects and Prey-Predator Two-way Interactions Model Using Lasso (min)

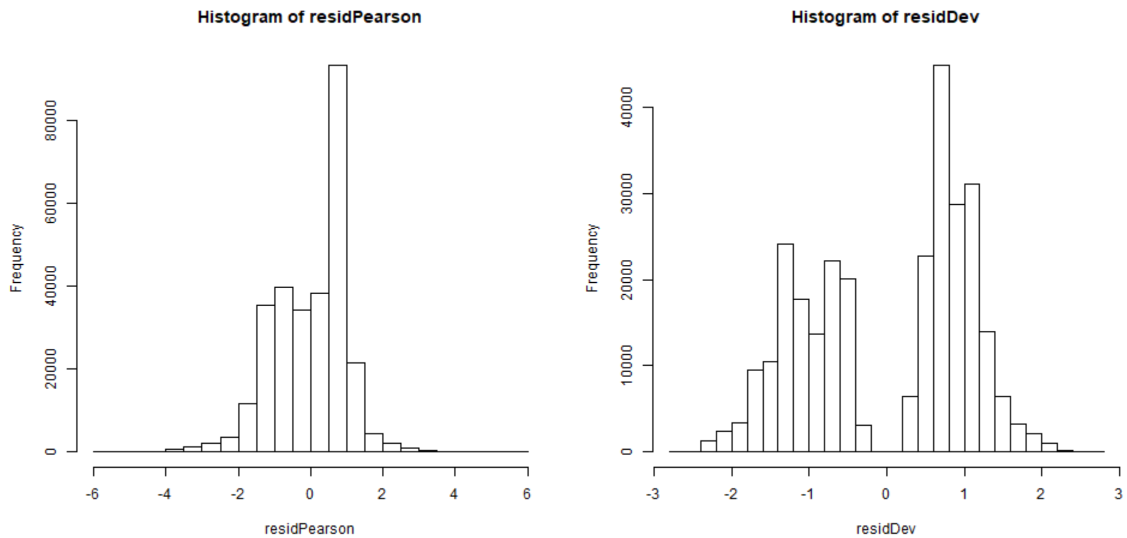

Figure 26. Residual Histograms (Pearson and Deviance Residuals Respectively) for Main Effects and Two-way Interactions Model Using Lasso (1se)

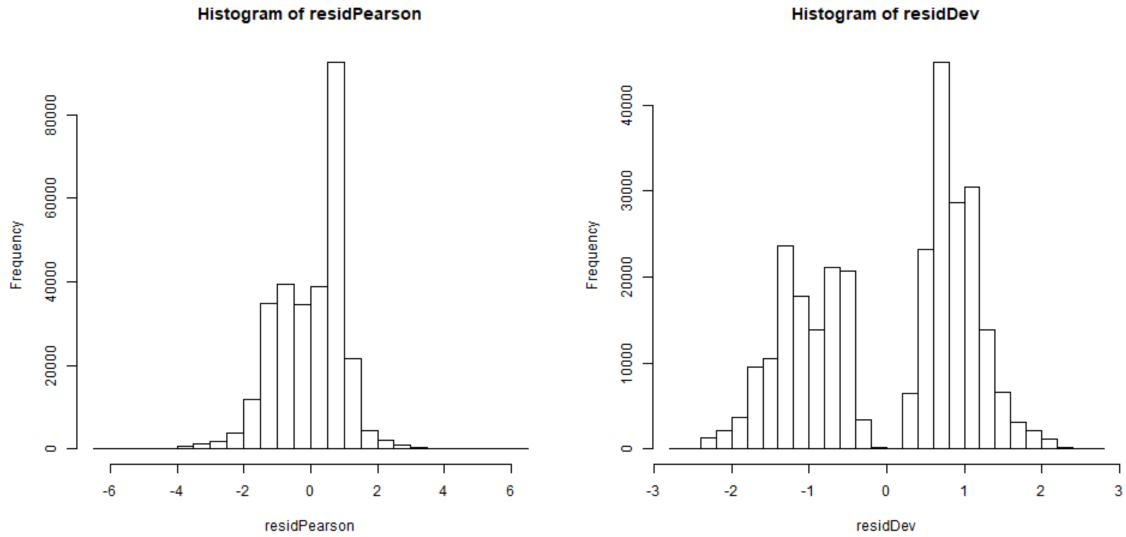

Figure 27. Residual Histograms (Pearson and Deviance Residuals Respectively) for Main Effects and Two-way Interactions Model Using Lasso (min)

APPENDIX D

ROC CURVES AND CROSS-VALIDATION PLOTS FOR LASSO

The following are the ROC curves for each model using the various variable selection methods as well as cross-validation plots for the LASSO.

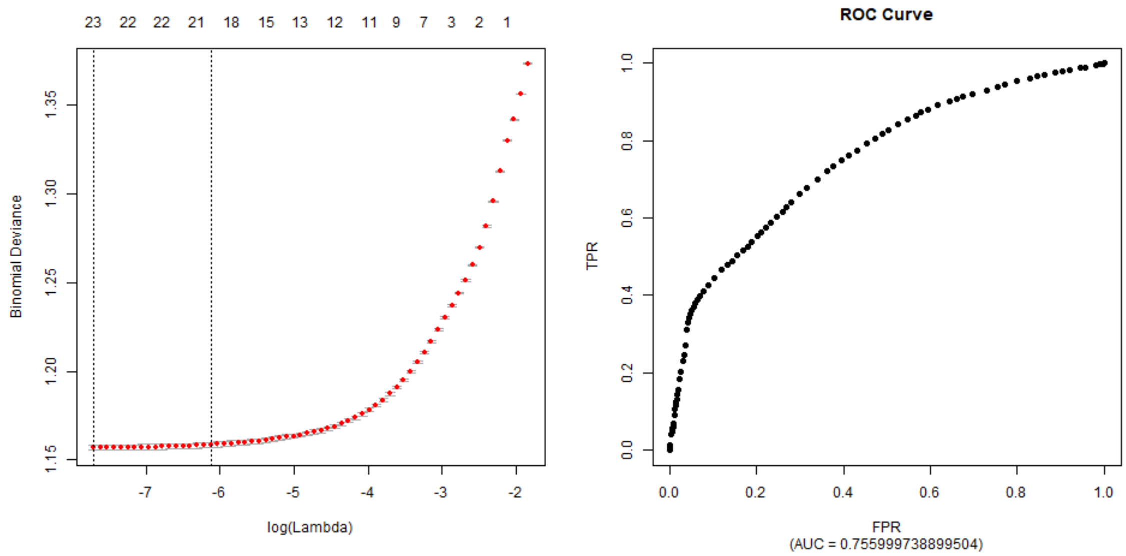

Figure 28. Cross-validation Plot for Lambda and ROC Curve for Main Effects Model Using Lasso (1se)

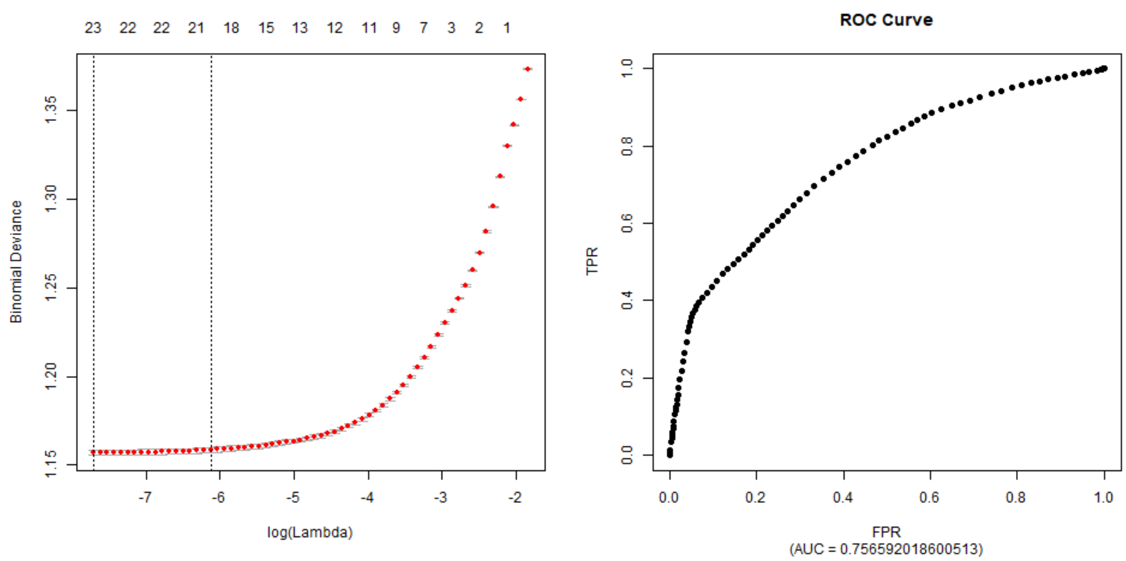

Figure 29. Cross-validation Plot for Lambda and ROC Curve for Main Effects Model Using Lasso (min)

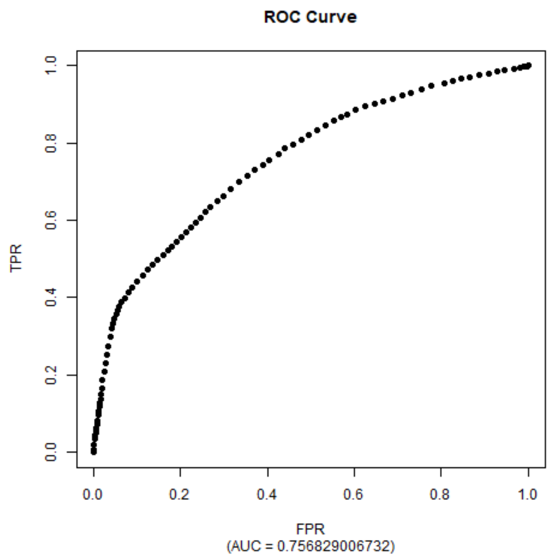

Figure 30. ROC Curve for Main Effects and Prey-Predator Two-way Interactions Model Using Traditional Method

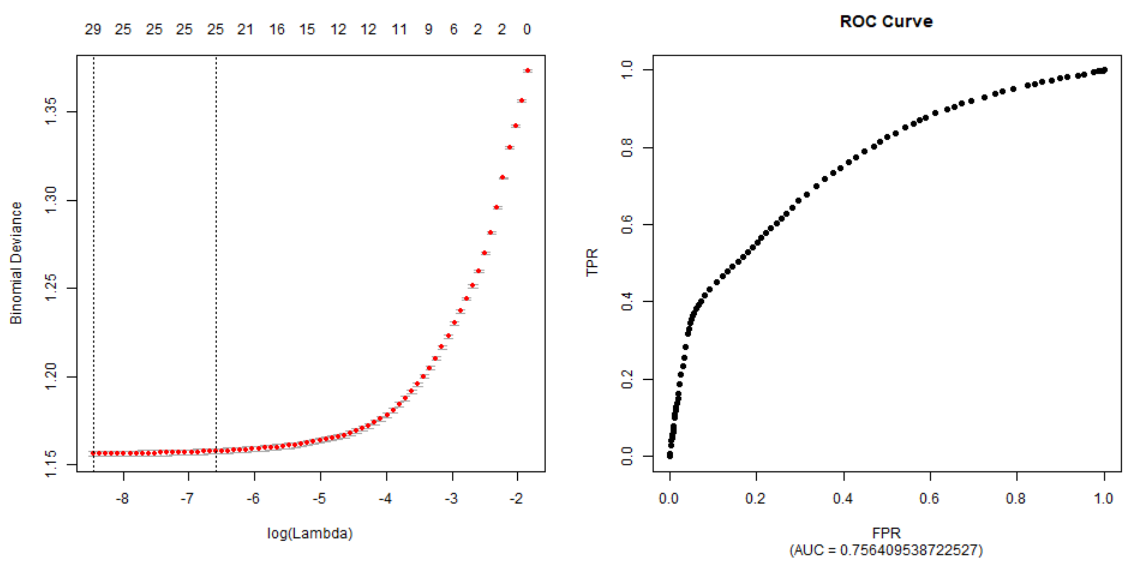

Figure 31. Cross-validation Plot for Lambda and ROC Curve for Main Effects and Prey-Predator Two-way Interactions Model Using Lasso (1se)

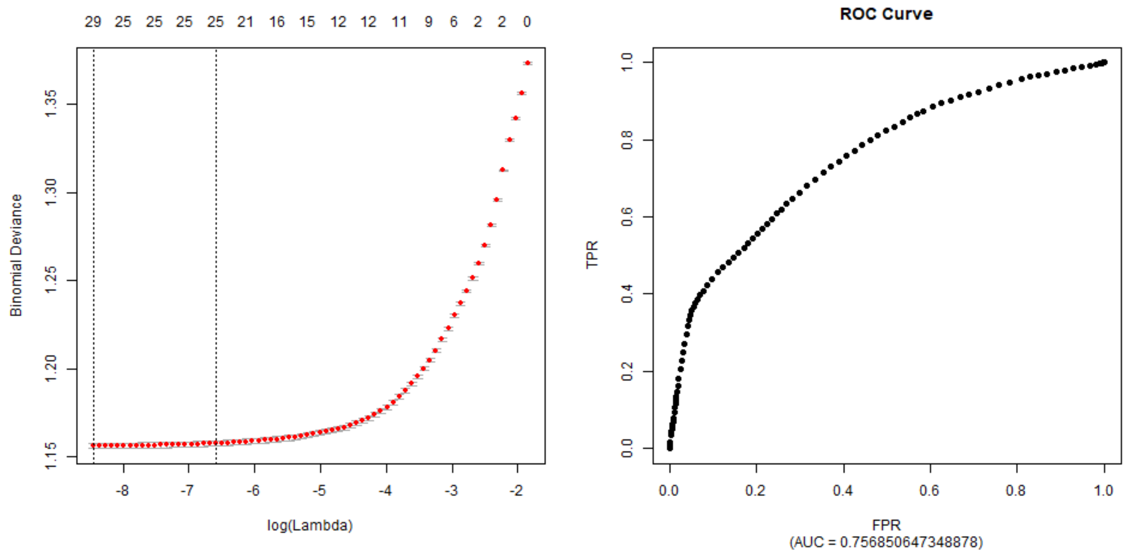

Figure 32. Cross-validation Plot for Lambda and ROC Curve for Main Effects and Prey-Predator Two-way Interactions Model Using Lasso (min)

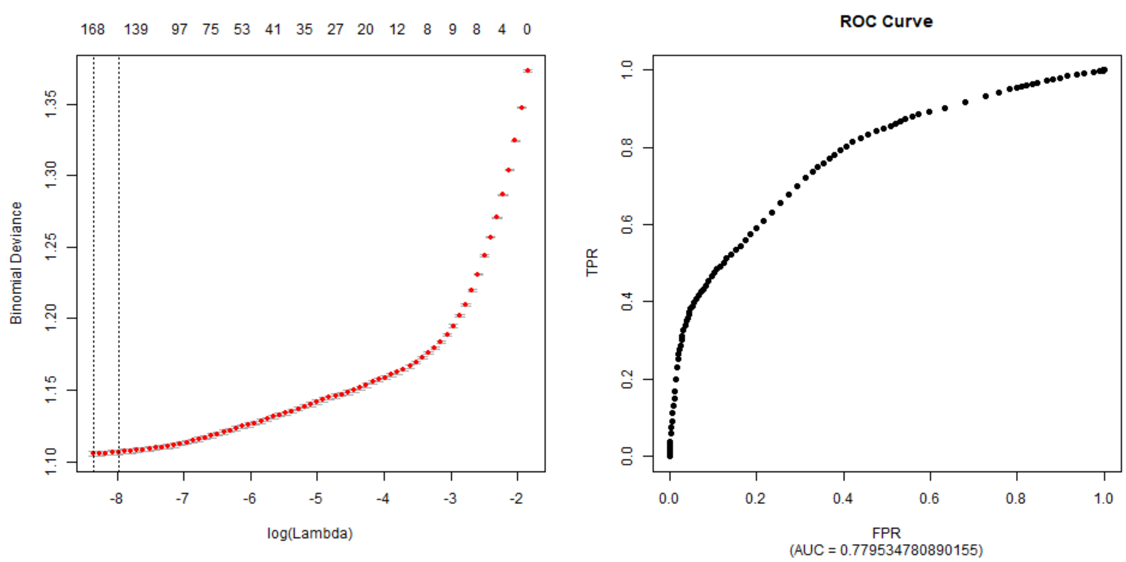

Figure 33. Cross-validation Plot for Lambda and ROC Curve for Main Effects and Full Two-way Interactions Model Using Lasso (1se)

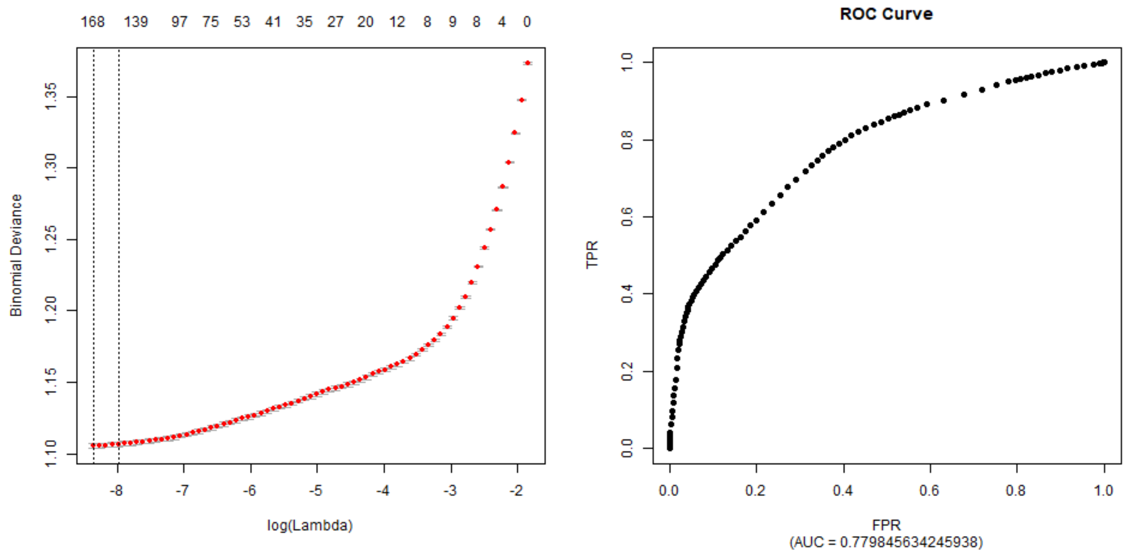

Figure 34. Cross-validation Plot for Lambda and ROC Curve for Main Effects and Full Two-way Interactions Model Using Lasso (min)

APPENDIX E

TABLE OF SIGNIFICANT COEFFICIENTS

The following are coefficients for model #1 (main effects model).

Table 6: Table of Coefficients for Main Effects Model for Traditional Methods

Traditional Method—Variables	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**	**Odds Ratio**
(Intercept)	0.0254	0.0795	0.3192	0.7496	1.0257
prey.max.velocity	0.0048	0.0006	7.6727	0.0000	1.0048
prey.agility	-0.0057	0.0026	-2.1993	0.0279	0.9943
prey.acceleration	0.0135	0.0006	21.9070	0.0000	1.0136
prey.vision.distance	0.0111	0.0016	7.0599	0.0000	1.0112
prey.vision.angle	-0.0012	0.0001	-8.9622	0.0000	0.9988
time.to.turn	-0.0017	0.0005	-3.4209	0.0006	0.9983
time.to.return.to.foraging	0.0003	0.0000	6.1600	0.0000	1.0003
time.spent.circling	0.0085	0.0030	2.8328	0.0046	1.0085
predator.max.velocity	-0.0130	0.0007	-18.5485	0.0000	0.9871
predator.agility	-0.1457	0.0026	-56.2032	0.0000	0.8644
predator.acceleration	0.0043	0.0005	9.0604	0.0000	1.0043
predator.deceleration	0.0114	0.0005	21.2716	0.0000	1.0115
predator.vision.distance	-0.0612	0.0015	-41.7962	0.0000	0.9407
predator.vision.angle	0.0014	0.0001	11.0042	0.0000	1.0014
time.to.give.up	-0.0014	0.0001	-27.8470	0.0000	0.9986
proportion.obstacles	-3.5177	0.0927	-37.9655	0.0000	0.0297
obstacle.radius	0.5068	0.0151	33.4794	0.0000	1.6600
obstacle.sensitivity.for.prey	-0.2512	0.0304	-8.2615	0.0000	0.7779
obstacle.sensitivity.for.predators	0.5541	0.0278	19.9270	0.0000	1.7403
safe.zone.attractiveness	0.0007	0.0000	24.4591	0.0000	1.0007
number.of.safe.zones	0.7129	0.0053	134.1480	0.0000	2.0398

Table 7: Table of Coefficients for Main Effects Model for Lasso Regression With Lambda (1se)

Lasso (1se)—Variables	**Estimate**	**Odds Ratio**
prey.max.velocity	0.0038	1.0038
prey.agility	0.0000	1.0000
prey.acceleration	0.0074	1.0074
prey.deceleration	-0.0025	0.9975
prey.vision.distance	0.0000	1.0000
prey.vision.angle	-0.0007	0.9993
time.to.turn	0.0000	1.0000
time.to.return.to.foraging	0.0001	1.0001
time.spent.circling	0.0030	1.0030
predator.max.velocity	-0.0110	0.9891

predator.agility	-0.1265	0.8812
predator.acceleration	0.0006	1.0006
predator.deceleration	0.0099	1.0100
predator.vision.distance	-0.0588	0.9429
predator.vision.angle	0.0004	1.0004
time.to.give.up	-0.0008	0.9992
proportion.obstacles	-2.9859	0.0505
obstacle.radius	0.4852	1.6244
obstacle.radius.range	-0.0097	0.9904
obstacle.sensitivity.for.prey	-0.0428	0.9581
obstacle.sensitivity.for.predators	0.3888	1.4752
safe.zone.attractiveness	0.00071	1.0007
number.of.safe.zones	0.68937	1.9925

Table 8: Table of Coefficients for Main Effects Model for Lasso (min)

Lasso (min)—Variables	**Estimate**	**Odds Ratio**
prey.max.velocity	0.0044	1.0044
prey.agility	-0.0014	0.9986
prey.acceleration	0.0124	1.0125
prey.deceleration	-0.0009	0.9991
prey.vision.distance	0.0079	1.0079
prey.vision.angle	-0.0012	0.9988
time.to.turn	-0.0012	0.9988
time.to.return.to.foraging	0.0002	1.0002
time.spent.circling	0.0054	1.0054
predator.max.velocity	-0.0125	0.9875
predator.agility	-0.1399	0.8695
predator.acceleration	0.0036	1.0036
predator.deceleration	0.0111	1.0111
predator.vision.distance	-0.0603	0.9415
predator.vision.angle	0.0012	1.0012
time.to.give.up	-0.0013	0.9987
proportion.obstacles	-3.3734	0.0343
obstacle.radius	0.5011	1.6505
obstacle.radius.range	-0.0059	0.9941
obstacle.sensitivity.for.prey	-0.2127	0.8084
obstacle.sensitivity.for.predators	0.5141	1.6722
safe.zone.attractiveness	0.0007	1.0007
number.of.safe.zones	0.7078	2.0295

The following are coefficients for model #2 (main effects and prey-predator two-way interactions):

Table 9: Table of Coefficients for Main Effects and Prey-Predator Two-way Interactions Model for Traditional Methods

Traditional Method—Variables	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**	**Odds Ratio**
(Intercept)	0.2137	0.1163	1.8367	0.0663	1.2382
prey.max.velocity	-0.0037	0.0013	-2.9478	0.0032	0.9963
predator.max.velocity	-0.0220	0.0013	-16.3623	0.0000	0.9783
prey.agility	-0.0479	0.0053	-9.0280	0.0000	0.9533
predator.agility	-0.2000	0.0054	-37.2898	0.0000	0.8187
prey.acceleration	0.0132	0.0006	21.2598	0.0000	1.0133
predator.acceleration	0.0060	0.0005	11.1890	0.0000	1.0060
prey.deceleration	-0.0013	0.0006	-2.0882	0.0368	0.9987
predator.deceleration	0.0105	0.0006	17.5896	0.0000	1.0106
prey.vision.distance	0.0096	0.0016	6.0389	0.0000	1.0097
predator.vision.distance	-0.0622	0.0015	-41.0504	0.0000	0.9397
prey.vision.angle	-0.0004	0.0004	-0.8532	0.3935	0.9996
predator.vision.angle	0.0021	0.0004	5.7905	0.0000	1.0021
obstacle.sensitivity.for.prey	0.2001	0.0529	3.7802	0.0002	1.2215
obstacle.sensitivity.for.predators	0.9794	0.0511	19.1672	0.0000	2.6629
time.to.turn	-0.0019	0.0005	-3.6746	0.0002	0.9981
time.to.return.to.foraging	0.0002	0.0000	4.8942	0.0000	1.0002
time.spent.circling	0.0156	0.0030	5.1504	0.0000	1.0157
time.to.give.up	-0.0014	0.0001	-26.7047	0.0000	0.9986
proportion.obstacles	-3.5763	0.0961	-37.2085	0.0000	0.0280
obstacle.radius	0.5294	0.0156	33.9568	0.0000	1.6979
safe.zone.attractiveness	0.0007	0.0000	22.1301	0.0000	1.0007
number.of.safe.zones	0.7147	0.0053	134.3607	0.0000	2.0436
prey.max.velocity:predator.max.velocity	0.0004	0.0001	7.6584	0.0000	1.0004
prey.agility:predator.agility	0.0090	0.0008	11.0134	0.0000	1.0090
prey.vision.angle:predator.vision.angle	0.0000	0.0000	-2.4237	0.0154	1.0000
obstacle.sensitivity.for.prey: obstacle.sensitivity.for.predators	-0.8980	0.0818	-10.9848	0.0000	0.4074

Table 10: Table of Coefficients for Main Effects and Prey-Predator Two-way Interactions Model for Lasso (1se)

Lasso (1se)—Variables	**Estimate**	**Odds Ratio**
prey.max.velocity	0.0009	1.0009

predator.max.velocity	-0.0148	0.9853
prey.max.velocity:predator.max.velocity	0.0001	1.0001
predator.agility	-0.1335	0.8750
prey.acceleration	0.0076	1.0076
predator.acceleration	0.0001	1.0001
prey.acceleration:predator.acceleration	0.0001	1.0001
prey.deceleration	-0.0019	0.9981
predator.deceleration	0.0102	1.0103
prey.vision.distance	0.0026	1.0026
predator.vision.distance	-0.0590	0.9427
prey.vision.angle	-0.0010	0.9990
predator.vision.angle	0.0007	1.0007
obstacle.sensitivity.for.prey	-0.0810	0.9222
obstacle.sensitivity.for.predators	0.4959	1.6419
obstacle.sensitivity.for.prey:obstacle.sensitivity.for.predators	-0.0646	0.9375
time.to.turn	-0.0006	0.9994
time.to.return.to.foraging	0.0002	1.0002
time.spent.circling	0.0042	1.0042
time.to.give.up	-0.0010	0.9990
proportion.obstacles	-3.17602	0.0418
obstacle.radius	0.48982	1.6320
obstacle.radius.range	-0.01993	0.9803
safe.zone.attractiveness	0.0007	1.0007
number.of.safe.zones	0.69777	2.0093

Table 11: Table of Coefficients for Main Effects and Prey-Predator Two-way Interactions Model for Lasso (min)

Lasso (min)—Variables	**Estimate**	**Odds Ratio**
prey.max.velocity	-0.0004	0.9996
predator.max.velocity	-0.0185	0.9816
prey.max.velocity:predator.max.velocity	0.0003	1.0003
prey.agility	-0.0322	0.9683
predator.agility	-0.1792	0.8360
prey.agility:predator.agility	0.0061	1.0061
prey.acceleration	0.0123	1.0124
predator.acceleration	0.0047	1.0047
prey.acceleration:predator.acceleration	0.0000	1.0000
prey.deceleration	-0.0011	0.9989
predator.deceleration	0.0109	1.0110

prey.deceleration:predator.deceleration	0.0000	1.0000
prey.vision.distance	0.0081	1.0081
predator.vision.distance	-0.0625	0.9394
prey.vision.distance:predator.vision.distance	0.0000	1.0000
prey.vision.angle	-0.0009	0.9991
predator.vision.angle	0.0015	1.0015
prey.vision.angle:predator.vision.angle	0.0000	1.0000
obstacle.sensitivity.for.prey	0.0305	1.0309
obstacle.sensitivity.for.predators	0.8058	2.2384
obstacle.sensitivity.for.prey:obstacle.sensitivity.for.predators	-0.5844	0.5575
time.to.turn	-0.0017	0.9983
time.to.return.to.foraging	0.0002	1.0002
time.spent.circling	0.0097	1.0098
time.to.give.up	-0.0013	0.9987
proportion.obstacles	-3.5022	0.0301
obstacle.radius	0.5164	1.6760
obstacle.radius.range	-0.00412	0.9959
safe.zone.attractiveness	0.00067	1.0007
number.of.safe.zones	0.71159	2.0372

The following are coefficients for model #3 (main effects and two-way interactions):

Table 12: Table of Coefficients for Main Effects and Two-way Interactions Model for Lasso (1se)

LASSO (1se)—Variables	**Estimate**	**Odds Ratio**
prey.deceleration	0.0053	1.0053
time.spent.circling	0.0331	1.0337
predator.agility	-0.2525	0.7768
predator.acceleration	-0.0180	0.9822
predator.vision.distance	-0.0663	0.9358
predator.vision.angle	0.0007	1.0007
proportion.obstacles	-6.1638	0.0021
safe.zone.attractiveness	0.0000	1.0000
prey.max.velocity:prey.acceleration	0.0003	1.0003
prey.max.velocity:prey.deceleration	0.0002	1.0002
prey.max.velocity:time.to.turn	0.0001	1.0001
prey.max.velocity:time.to.return.to.foraging	0.0000	1.0000
prey.max.velocity:predator.max.velocity	-0.0001	0.9999
prey.max.velocity:predator.deceleration	-0.0002	0.9998
prey.max.velocity:proportion.obstacles	0.0077	1.0077

prey.max.velocity:obstacle.radius	-0.0023	0.9977
prey.max.velocity:safe.zone.attractiveness	0.0000	1.0000
prey.max.velocity:number.of.safe.zones	-0.0025	0.9975
prey.agility:prey.deceleration	0.0002	1.0002
prey.agility:time.to.return.to.foraging	0.0000	1.0000
prey.agility:time.spent.circling	0.0004	1.0004
prey.agility:predator.max.velocity	-0.0001	0.9999
prey.agility:predator.agility	0.0147	1.0148
prey.agility:predator.acceleration	-0.0004	0.9996
prey.agility:predator.vision.distance	0.0070	1.0070
prey.agility:time.to.give.up	-0.0001	0.9999
prey.agility:proportion.obstacles	-0.0642	0.9379
prey.agility:obstacle.radius	-0.0041	0.9959
prey.agility:obstacle.radius.range	0.0060	1.0060
prey.agility:obstacle.sensitivity.for.predators	-0.0508	0.9504
prey.agility:number.of.safe.zones	-0.0024	0.9976
prey.agility:number.of.target.patches	0.0012	1.0012
prey.acceleration:prey.deceleration	0.0001	1.0001
prey.acceleration:prey.vision.distance	-0.0005	0.9995
prey.acceleration:prey.vision.angle	0.0000	1.0000
prey.acceleration:time.to.turn	0.0000	1.0000
prey.acceleration:predator.acceleration	0.0001	1.0001
prey.acceleration:predator.deceleration	0.0000	1.0000
prey.acceleration:predator.vision.angle	0.0001	1.0001
prey.acceleration:proportion.obstacles	0.0275	1.0278
prey.acceleration:obstacle.radius	-0.0121	0.9880
prey.acceleration:obstacle.sensitivity.for.predators	-0.0062	0.9938
prey.acceleration:safe.zone.attractiveness	0.0000	1.0000
prey.acceleration:number.of.safe.zones	0.0004	1.0004
prey.deceleration:predator.agility	-0.0010	0.9990
prey.deceleration:predator.deceleration	0.0000	1.0000
prey.deceleration:predator.vision.distance	-0.0003	0.9997
prey.deceleration:predator.vision.angle	0.0000	1.0000
prey.deceleration:time.to.give.up	0.0000	1.0000
prey.deceleration:obstacle.radius	0.0001	1.0001
prey.deceleration:safe.zone.attractiveness	0.0000	1.0000
prey.deceleration:number.of.safe.zones	-0.0045	0.9955
prey.vision.distance:prey.vision.angle	0.0000	1.0000
prey.vision.distance:time.to.turn	-0.0001	0.9999

prey.vision.distance:time.spent.circling	0.0027	1.0027
prey.vision.distance:predator.acceleration	0.0008	1.0008
prey.vision.distance:time.to.give.up	-0.0001	0.9999
prey.vision.distance:proportion.obstacles	0.0000	1.0000
prey.vision.distance:obstacle.radius.range	-0.0038	0.9962
prey.vision.distance:obstacle.sensitivity.for.prey	0.0131	1.0132
prey.vision.distance:obstacle.sensitivity.for.predators	-0.0143	0.9858
prey.vision.distance:safe.zone.attractiveness	0.0000	1.0000
prey.vision.distance:number.of.safe.zones	0.0281	1.0285
prey.vision.angle:predator.max.velocity	0.0000	1.0000
prey.vision.angle:predator.agility	0.0000	1.0000
prey.vision.angle:predator.deceleration	0.0000	1.0000
prey.vision.angle:predator.vision.distance	-0.0001	0.9999
prey.vision.angle:time.to.give.up	0.0000	1.0000
prey.vision.angle:obstacle.radius.range	0.0001	1.0001
time.to.turn:time.to.return.to.foraging	0.0000	1.0000
time.to.turn:predator.max.velocity	0.0001	1.0001
time.to.turn:predator.agility	-0.0001	0.9999
time.to.turn:predator.acceleration	0.0001	1.0001
time.to.turn:number.of.safe.zones	-0.0006	0.9994
time.to.return.to.foraging:time.spent.circling	0.0000	1.0000
time.to.return.to.foraging:predator.max.velocity	0.0000	1.0000
time.to.return.to.foraging:predator.vision.angle	0.0000	1.0000
time.to.return.to.foraging:time.to.give.up	0.0000	1.0000
time.to.return.to.foraging:proportion.obstacles	0.0024	1.0024
time.to.return.to.foraging:obstacle.radius	-0.0001	0.9999
time.to.return.to.foraging:obstacle.radius.range	0.0000	1.0000
time.to.return.to.foraging:obstacle.sensitivity.for.prey	0.0001	1.0001
time.to.return.to.foraging:number.of.safe.zones	0.0006	1.0006
time.spent.circling:predator.agility	-0.0024	0.9976
time.spent.circling:predator.vision.angle	0.0001	1.0001
time.spent.circling:proportion.obstacles	0.2585	1.2950
time.spent.circling:obstacle.radius	-0.0516	0.9497
time.spent.circling:obstacle.sensitivity.for.predators	0.0088	1.0089
time.spent.circling:safe.zone.attractiveness	0.0000	1.0000
time.spent.circling:number.of.safe.zones	-0.0152	0.9849
predator.max.velocity:predator.agility	0.0015	1.0015
predator.max.velocity:predator.vision.angle	0.0000	1.0000
predator.max.velocity:time.to.give.up	0.0000	1.0000

predator.max.velocity:proportion.obstacles	-0.0340	0.9665
predator.max.velocity:obstacle.radius.range	0.0014	1.0014
predator.max.velocity:obstacle.sensitivity.for.prey	-0.0022	0.9978
predator.max.velocity:obstacle.sensitivity.for.predators	0.0031	1.0031
predator.max.velocity:safe.zone.attractiveness	0.0000	1.0000
predator.max.velocity:number.of.safe.zones	-0.0014	0.9986
predator.agility:predator.acceleration	0.0032	1.0032
predator.agility:predator.vision.distance	-0.0128	0.9873
predator.agility:predator.vision.angle	-0.0002	0.9998
predator.agility:proportion.obstacles	0.6099	1.8402
predator.agility:obstacle.radius	-0.0332	0.9673
predator.agility:obstacle.radius.range	-0.0154	0.9847
predator.agility:obstacle.sensitivity.for.predators	0.0194	1.0196
predator.agility:safe.zone.attractiveness	0.0001	1.0001
predator.agility:number.of.safe.zones	0.0165	1.0167
predator.acceleration:predator.deceleration	0.0000	1.0000
predator.acceleration:predator.vision.distance	0.0003	1.0003
predator.acceleration:predator.vision.angle	0.0000	1.0000
predator.acceleration:time.to.give.up	0.0000	1.0000
predator.acceleration:proportion.obstacles	-0.0711	0.9314
predator.acceleration:obstacle.radius	0.0044	1.0044
predator.acceleration:obstacle.sensitivity.for.prey	-0.0049	0.9951
predator.acceleration:obstacle.sensitivity.for.predators	0.0034	1.0035
predator.acceleration:number.of.safe.zones	0.0064	1.0064
predator.acceleration:number.of.target.patches	-0.0002	0.9998
predator.deceleration:predator.vision.angle	0.0000	1.0000
predator.deceleration:obstacle.sensitivity.for.prey	0.0027	1.0027
predator.deceleration:obstacle.sensitivity.for.predators	0.0030	1.0030
predator.deceleration:number.of.safe.zones	0.0002	1.0002
predator.vision.distance:time.to.give.up	-0.0001	0.9999
predator.vision.distance:obstacle.radius	0.0197	1.0199
predator.vision.distance:obstacle.sensitivity.for.predators	0.0906	1.0948
predator.vision.distance:safe.zone.attractiveness	0.0000	1.0000
predator.vision.distance:number.of.safe.zones	-0.0089	0.9912
predator.vision.angle:time.to.give.up	0.0000	1.0000
predator.vision.angle:obstacle.radius	0.0001	1.0001
predator.vision.angle:obstacle.radius.range	0.0003	1.0003
predator.vision.angle:obstacle.sensitivity.for.predators	0.0004	1.0004
predator.vision.angle:number.of.safe.zones	-0.0012	0.9988

time.to.give.up:proportion.obstacles	-0.0048	0.9953
time.to.give.up:obstacle.radius	0.0009	1.0009
time.to.give.up:obstacle.radius.range	-0.0004	0.9996
time.to.give.up:obstacle.sensitivity.for.prey	-0.0004	0.9996
time.to.give.up:number.of.safe.zones	0.0013	1.0013
time.to.give.up:number.of.target.patches	0.0000	1.0000
proportion.obstacles:obstacle.radius	3.7708	43.4129
proportion.obstacles:obstacle.sensitivity.for.prey	-1.7284	0.1776
proportion.obstacles:obstacle.sensitivity.for.predators	1.7250	5.6123
proportion.obstacles:number.of.safe.zones	-1.2619	0.2831
obstacle.radius:obstacle.sensitivity.for.predators	-0.0041	0.9959
obstacle.radius:safe.zone.attractiveness	-0.0003	0.9997
obstacle.radius:number.of.safe.zones	0.3895	1.4762
obstacle.radius.range:obstacle.sensitivity.for.prey	0.0203	1.0205
obstacle.radius.range:obstacle.sensitivity.for.predators	0.1652	1.1797
obstacle.radius.range:safe.zone.attractiveness	0.0000	1.0000
obstacle.radius.range:number.of.safe.zones	0.0961	1.1009
obstacle.radius.range:number.of.target.patches	-0.0020	0.9980
obstacle.sensitivity.for.prey:safe.zone.attractiveness	0.0002	1.0002
obstacle.sensitivity.for.predators:safe.zone.attractiveness	-0.0003	0.9997
obstacle.sensitivity.for.predators:number.of.safe.zones	-0.0830	0.9203
obstacle.sensitivity.for.predators:number.of.target.patches	0.0003	1.0003
safe.zone.attractiveness:number.of.safe.zones	0.0000	1.0000
obstacle.sensitivity.for.predators:number.of.target.patches	0.0041	1.0041
safe.zone.attractiveness:number.of.safe.zones	0.0000	1.0000

Table 13: Table of Coefficients for Main Effects and Two-way Interactions Model for Lasso (min)

LASSO (min)—Variables	**Estimate**	**Odds Ratio**
prey.max.velocity	-0.0012	0.9988
prey.deceleration	0.0083	1.0083
time.spent.circling	0.0371	1.0378
predator.agility	-0.2937	0.7455
predator.acceleration	-0.0278	0.9726
predator.vision.distance	-0.0767	0.9262
predator.vision.angle	0.0011	1.0011
time.to.give.up	-0.0003	0.9997
proportion.obstacles	-7.0051	0.0009
prey.max.velocity:prey.acceleration	0.0004	1.0004

prey.max.velocity:prey.deceleration	0.0002	1.0002
prey.max.velocity:time.to.turn	0.0001	1.0001
prey.max.velocity:time.to.return.to.foraging	0.0000	1.0000
prey.max.velocity:predator.max.velocity	-0.0001	0.9999
prey.max.velocity:predator.deceleration	-0.0002	0.9998
prey.max.velocity:proportion.obstacles	0.0265	1.0268
prey.max.velocity:obstacle.radius	-0.0036	0.9964
prey.max.velocity:obstacle.radius.range	0.0000	1.0000
prey.max.velocity:safe.zone.attractiveness	0.0000	1.0000
prey.max.velocity:number.of.safe.zones	-0.0027	0.9973
prey.agility:prey.acceleration	0.0001	1.0001
prey.agility:prey.deceleration	0.0000	1.0000
prey.agility:time.to.return.to.foraging	0.0000	1.0000
prey.agility:time.spent.circling	0.0012	1.0012
prey.agility:predator.max.velocity	-0.0003	0.9997
prey.agility:predator.agility	0.0161	1.0162
prey.agility:predator.acceleration	-0.0005	0.9995
prey.agility:predator.vision.distance	0.0078	1.0078
prey.agility:time.to.give.up	-0.0001	0.9999
prey.agility:proportion.obstacles	-0.1003	0.9046
prey.agility:obstacle.radius	-0.0009	0.9991
prey.agility:obstacle.radius.range	0.0077	1.0077
prey.agility:obstacle.sensitivity.for.predators	-0.0569	0.9447
prey.agility:number.of.safe.zones	-0.0008	0.9992
prey.agility:number.of.target.patches	0.0013	1.0013
prey.acceleration:prey.deceleration	0.0001	1.0001
prey.acceleration:prey.vision.distance	-0.0005	0.9995
prey.acceleration:prey.vision.angle	0.0000	1.0000
prey.acceleration:time.to.turn	-0.0001	0.9999
prey.acceleration:time.to.return.to.foraging	0.0000	1.0000
prey.acceleration:predator.agility	0.0006	1.0006
prey.acceleration:predator.acceleration	0.0002	1.0002
prey.acceleration:predator.deceleration	0.0000	1.0000
prey.acceleration:predator.vision.angle	0.0001	1.0001
prey.acceleration:proportion.obstacles	0.0445	1.0455
prey.acceleration:obstacle.radius	-0.0146	0.9855
prey.acceleration:obstacle.sensitivity.for.predators	-0.0118	0.9883
prey.acceleration:safe.zone.attractiveness	0.0000	1.0000
prey.acceleration:number.of.safe.zones	0.0002	1.0002

prey.deceleration:prey.vision.distance	0.0000	1.0000
prey.deceleration:predator.agility	-0.0011	0.9989
prey.deceleration:predator.acceleration	0.0000	1.0000
prey.deceleration:predator.deceleration	0.0000	1.0000
prey.deceleration:predator.vision.distance	-0.0004	0.9996
prey.deceleration:predator.vision.angle	0.0000	1.0000
prey.deceleration:time.to.give.up	0.0000	1.0000
prey.deceleration:number.of.safe.zones	-0.0043	0.9957
prey.vision.distance:prey.vision.angle	0.0001	1.0001
prey.vision.distance:time.to.turn	-0.0002	0.9998
prey.vision.distance:time.spent.circling	0.0026	1.0026
prey.vision.distance:predator.acceleration	0.0009	1.0009
prey.vision.distance:predator.vision.angle	0.0000	1.0000
prey.vision.distance:time.to.give.up	-0.0001	0.9999
prey.vision.distance:obstacle.radius.range	-0.0055	0.9946
prey.vision.distance:obstacle.sensitivity.for.prey	0.0136	1.0137
prey.vision.distance:obstacle.sensitivity.for.predators	-0.0221	0.9781
prey.vision.distance:safe.zone.attractiveness	0.0000	1.0000
prey.vision.distance:number.of.safe.zones	0.0314	1.0318
prey.vision.angle:predator.max.velocity	0.0000	1.0000
prey.vision.angle:predator.agility	0.0000	1.0000
prey.vision.angle:predator.deceleration	0.0000	1.0000
prey.vision.angle:predator.vision.distance	-0.0001	0.9999
prey.vision.angle:time.to.give.up	0.0000	1.0000
prey.vision.angle:obstacle.radius.range	0.0001	1.0001
time.to.turn:time.to.return.to.foraging	0.0000	1.0000
time.to.turn:predator.max.velocity	0.0001	1.0001
time.to.turn:predator.agility	-0.0001	0.9999
time.to.turn:predator.acceleration	0.0001	1.0001
time.to.turn:obstacle.radius	0.0000	1.0000
time.to.turn:number.of.safe.zones	-0.0011	0.9989
time.to.turn:number.of.target.patches	0.0000	1.0000
time.to.return.to.foraging:time.spent.circling	-0.0001	0.9999
time.to.return.to.foraging:predator.max.velocity	0.0000	1.0000
time.to.return.to.foraging:predator.acceleration	0.0000	1.0000
time.to.return.to.foraging:predator.vision.angle	0.0000	1.0000
time.to.return.to.foraging:proportion.obstacles	0.0029	1.0029
time.to.return.to.foraging:obstacle.radius.range	-0.0002	0.9998
time.to.return.to.foraging:obstacle.sensitivity.for.prey	0.0001	1.0001

time.to.return.to.foraging:number.of.safe.zones	0.0006	1.0006
time.spent.circling:predator.agility	-0.0020	0.9980
time.spent.circling:predator.acceleration	0.0001	1.0001
time.spent.circling:predator.vision.angle	0.0000	1.0000
time.spent.circling:time.to.give.up	0.0000	1.0000
time.spent.circling:proportion.obstacles	0.2900	1.3364
time.spent.circling:obstacle.radius	-0.0629	0.9390
time.spent.circling:obstacle.sensitivity.for.predators	0.0222	1.0224
time.spent.circling:safe.zone.attractiveness	0.0000	1.0000
time.spent.circling:number.of.safe.zones	-0.0188	0.9814
predator.max.velocity:predator.agility	0.0014	1.0014
predator.max.velocity:predator.vision.angle	0.0000	1.0000
predator.max.velocity:proportion.obstacles	-0.0319	0.9686
predator.max.velocity:obstacle.radius.range	0.0022	1.0022
predator.max.velocity:obstacle.sensitivity.for.prey	-0.0039	0.9961
predator.max.velocity:obstacle.sensitivity.for.predators	0.0075	1.0075
predator.max.velocity:safe.zone.attractiveness	0.0000	1.0000
predator.max.velocity:number.of.safe.zones	-0.0020	0.9980
predator.agility:predator.acceleration	0.0034	1.0034
predator.agility:predator.deceleration	0.0000	1.0000
predator.agility:predator.vision.distance	-0.0124	0.9877
predator.agility:predator.vision.angle	-0.0003	0.9997
predator.agility:proportion.obstacles	0.6145	1.8487
predator.agility:obstacle.radius	-0.0292	0.9712
predator.agility:obstacle.radius.range	-0.0180	0.9822
predator.agility:obstacle.sensitivity.for.predators	0.0396	1.0404
predator.agility:safe.zone.attractiveness	0.0001	1.0001
predator.agility:number.of.safe.zones	0.0189	1.0191
predator.agility:number.of.target.patches	0.0002	1.0002
predator.acceleration:predator.deceleration	0.0000	1.0000
predator.acceleration:predator.vision.distance	0.0005	1.0005
predator.acceleration:predator.vision.angle	0.0000	1.0000
predator.acceleration:time.to.give.up	0.0000	1.0000
predator.acceleration:proportion.obstacles	-0.0764	0.9265
predator.acceleration:obstacle.radius	0.0071	1.0072
predator.acceleration:obstacle.radius.range	0.0001	1.0001
predator.acceleration:obstacle.sensitivity.for.prey	-0.0056	0.9945
predator.acceleration:obstacle.sensitivity.for.predators	0.0054	1.0054
predator.acceleration:number.of.safe.zones	0.0067	1.0067

predator.acceleration:number.of.target.patches	-0.0004	0.9996
predator.deceleration:predator.vision.angle	0.0000	1.0000
predator.deceleration:proportion.obstacles	0.0008	1.0008
predator.deceleration:obstacle.sensitivity.for.prey	0.0024	1.0024
predator.deceleration:obstacle.sensitivity.for.predators	0.0036	1.0036
predator.deceleration:number.of.safe.zones	0.0003	1.0003
predator.deceleration:number.of.target.patches	-0.0001	0.9999
predator.vision.distance:time.to.give.up	-0.0001	0.9999
predator.vision.distance:obstacle.radius	0.0161	1.0163
predator.vision.distance:obstacle.sensitivity.for.predators	0.0926	1.0970
predator.vision.distance:safe.zone.attractiveness	0.0000	1.0000
predator.vision.distance:number.of.safe.zones	-0.0093	0.9907
predator.vision.angle:time.to.give.up	0.0000	1.0000
predator.vision.angle:obstacle.radius	0.0003	1.0003
predator.vision.angle:obstacle.radius.range	0.0004	1.0004
predator.vision.angle:obstacle.sensitivity.for.predators	0.0006	1.0006
predator.vision.angle:number.of.safe.zones	-0.0012	0.9988
time.to.give.up:proportion.obstacles	-0.0050	0.9950
time.to.give.up:obstacle.radius	0.0012	1.0012
time.to.give.up:obstacle.radius.range	-0.0004	0.9996
time.to.give.up:obstacle.sensitivity.for.prey	-0.0003	0.9997
time.to.give.up:number.of.safe.zones	0.0014	1.0014
time.to.give.up:number.of.target.patches	0.0000	1.0000
proportion.obstacles:obstacle.radius	3.9172	50.2591
proportion.obstacles:obstacle.sensitivity.for.prey	-2.0849	0.1243
proportion.obstacles:obstacle.sensitivity.for.predators	2.2379	9.3734
proportion.obstacles:safe.zone.attractiveness	-0.0003	0.9997
proportion.obstacles:number.of.safe.zones	-1.4278	0.2398
obstacle.radius:obstacle.radius.range	-0.0209	0.9793
obstacle.radius:obstacle.sensitivity.for.prey	0.0560	1.0576
obstacle.radius:obstacle.sensitivity.for.predators	-0.1783	0.8367
obstacle.radius:safe.zone.attractiveness	-0.0003	0.9997
obstacle.radius:number.of.safe.zones	0.4148	1.5141
obstacle.radius.range:obstacle.sensitivity.for.prey	0.0509	1.0522
obstacle.radius.range:obstacle.sensitivity.for.predators	0.2122	1.2364
obstacle.radius.range:safe.zone.attractiveness	0.0000	1.0000
obstacle.radius.range:number.of.safe.zones	0.1038	1.1093
obstacle.radius.range:number.of.target.patches	-0.0043	0.9957
obstacle.sensitivity.for.prey:safe.zone.attractiveness	0.0002	1.0002

obstacle.sensitivity.for.predators:safe.zone.attractiveness	-0.0004	0.9996
obstacle.sensitivity.for.predators:number.of.safe.zones	-0.1327	0.8757
obstacle.sensitivity.for.predators:number.of.target.patches	0.0052	1.0052
safe.zone.attractiveness:number.of.safe.zones	-0.0001	0.9999
safe.zone.attractiveness:number.of.safe.zones	-0.0001	0.9999